agriculture

MDPI

Article

A Two-Stage G×E Modeling Framework Improves Crop Yield Prediction and Adaptive Selection

Qi Wang [1,2], Xiaohe Liang [1,2], Jiayu Zhuang [1,2], Jiajia Liu[1,2] and Ailian Zhou [1,2,*]

[1] Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China; wangqi03@caas.cn (Q.W.); liangxiaohe@caas.cn (X.L.); zhuangjiayu@caas.cn (J.Z.); liujiajia@caas.cn (J.L.)
[2] Key Laboratory of Agricultural Blockchain Application, Ministry of Agriculture and Rural Affairs, Beijing 100081, China
* Correspondence: zhouailian@caas.cn

Abstract

Accurate maize yield prediction across diverse environments is pivotal for modern breeding programs. While machine learning (ML) excels at capturing non-linear environmental effects, Genomic Best Linear Unbiased Prediction (GBLUP) remains a benchmark for modeling polygenic small-effect contributions. However, principled integration of these paradigms—while explicitly accounting for genotype-by-environment interaction (G×E)—remains a formidable challenge. We propose a two-step framework evaluated on the Genomes to Fields (G2F) 2022 dataset. In Step 1, ML models are employed to fit environmental main effects; in Step 2, genomic residuals are modeled via additive-dominance relationship matrices, augmented by an explicit low-rank G×E matrix. Candidate interaction markers were screened through plasticity-based genome-wide association studies (GWAS) across six phenotypic stability metrics and used to construct a low-rank candidate G×E representation, with a cross-validation-selected scaling parameter applied to control the contribution of the predicted G×E component. TwoStep_G×E_alpha0.33, achieved a within–environment Pearson correlation coefficient (PCC) of 0.376, outperformed both GBLUP and the competition-winning model (PCC = 0.357) in within-environment ranking. Furthermore, environment-adaptive selection yielded a genetic gain of 0.454 Mg ha^{-1}, representing a 34.7% improvement over GBLUP. Overall, the proposed framework provides a practical approach for environment-specific yield prediction and adaptive selection in maize breeding.

Keywords: crop yield prediction; genotype–by–environment interaction; plasticity-based genome-wide association studies; two-step prediction framework; machine learning

1. Introduction

Climate change poses a severe threat to global food production. The World Health Organization estimates that by 2030, climate change–induced heat stress, hunger, and disease will cause approximately 250,000 additional deaths annually [1]. Meanwhile, according to the United Nations World Population Prospects 2024, the global population is projected to reach approximately 9.7 billion by around 2050 [2], placing unprecedented pressure on global food security. As one of the most important sources of food, feed, and industrial starch worldwide, maize plays an indispensable role in ensuring food security. However, its yield is highly vulnerable to abiotic stresses such as drought, salinity, and flooding, which is particularly exacerbated under the context of climate change [3]. In this context, advances in genomic technologies, high-throughput environmental sensing [4], and statistical

Received: 2 April 2026
Revised: 7 May 2026
Accepted: 19 May 2026
Published: 2 June 2026

and machine-learning modeling approaches, including emerging deep-learning methods, are creating new opportunities to integrate genotype and environmental information for predicting cultivar performance across diverse environments, thereby supporting precision breeding and yield optimization under changing climatic conditions [5–7].

Grain yield in maize is jointly determined by a complex genetic architecture and environmental factors. Previous studies have shown that the heritability of yield exhibits substantial heterogeneity across different experimental contexts, and it decreases markedly under abiotic stress conditions such as drought and heat stress [8]. In 86 environments from the G2F project, the heritability of grain yield ranged from 0.04 to 0.35. Heritability was lowest in low-yield environments and increased with yield level, stabilizing at approximately 0.3 in most environments [9]. In terms of environmental response patterns, maize yield exhibits pronounced nonlinear responses to key environmental factors. Studies have indicated a critical temperature range of approximately 29–35 °C, beyond which yield declines sharply. This threshold varies significantly across genotypes, geographic regions, and developmental stages [10]. In addition, the effects of environmental stress on yield are strongly stage-specific: plants are more sensitive to heat stress during early grain filling than during late grain filling [11], while drought stress occurring at flowering often leads to pollen sterility and kernel abortion, resulting in irreversible yield losses [12]. From a genetic perspective, grain yield is a typical complex quantitative trait governed by numerous small-effect quantitative trait loci (QTLs) [13], and dominance effects contribute substantially to heterosis in hybrid maize [14]. More importantly, G×E are pervasive in yield-related traits, causing substantial re-ranking of genotypes across environments. For instance, under drought and heat combined stress, the genetic correlations of grain yield across different treatments (drought, heat stress, and combined drought–heat stress) are extremely low (−0.01 to 0.29), suggesting near-independent genetic control under different stress conditions [8]. Similarly, inbred lines Mo17 and B73 exhibit contrasting response strategies under the same combined stress, with yield losses differing by as much as 37% (49% vs. 86%), highlighting strong genotype-specific G×E [15]. In summary, the nonlinear nature of environmental responses, the polygenic architecture with predominantly small-effect loci, the widespread dominance effects, and the pervasive G×E collectively pose significant challenges for accurate maize yield prediction models.

Genomic best linear unbiased prediction (GBLUP) is one of the most widely used approaches for multi-environment yield prediction. By constructing a genome-wide relationship matrix, GBLUP models additive genetic effects and has demonstrated strong advantages in capturing polygenic architectures composed of numerous small-effect loci, while also requiring relatively modest training population sizes [16]. However, GBLUP assumes equal contribution of all single nucleotide polymorphisms (SNPs) to genetic variance and is essentially a linear model, which makes it difficult to capture nonlinear structures arising from epistasis (G×G) and G×E [17]. In recent years, machine learning (ML) and deep learning (DL) methods have attracted increasing attention due to their ability to model complex nonlinear patterns in large-scale, multi-source datasets [18]. Nevertheless, existing studies have shown that for traits predominantly controlled by a large number of small-effect additive loci, complex machine learning models do not consistently outperform GBLUP in predictive accuracy [19]. Moreover, deep learning approaches typically require training datasets far larger than those commonly available in plant breeding, entail substantial computational costs, and remain limited in their ability to explicitly model polygenic additive genetic architectures.

G×E is one of the most important sources of variation in maize grain yield [20]. Recent studies have attempted to improve multi-environment maize genomic prediction by incorporating machine learning and environmental covariates. Barreto et al. compared

statistical genomic prediction models with machine learning methods for predicting maize single-cross hybrids across multi-environment trials and showed that both approaches can be effective, but the best-performing strategy was case-dependent and varied with prediction scenarios and data imbalance [16]. Fernandes et al. further integrated genetic and feature-engineered environmental information using gradient boosting models in the G2F maize dataset and showed that environmental covariates can improve prediction accuracy and indirectly account for G×E patterns [21]. However, these studies also highlight remaining limitations: machine learning models can flexibly combine genetic and environmental predictors, but they often model G×E implicitly and do not explicitly separate environmental main effects, genomic main effects, and genotype-dependent environmental responses. Conversely, directly constructing SNP-by-environment features can provide a more explicit representation of G×E, but may substantially increase feature dimensionality, memory use, and computational cost in high-dimensional multi-environment datasets. Plasticity genome-wide association studies (plasticity GWAS) provide a complementary strategy by quantifying phenotypic plasticity—defined as the variation of a genotype's performance across environments—and using it as a trait for association mapping. Previous studies have shown that the genetic architectures underlying trait means and phenotypic plasticity can be partly distinct [22]. For instance, in flowering time, GWAS based on plasticity indices has been used to identify markers associated with photoperiod-related variation [23]. These findings suggest that plasticity GWAS can be useful for prioritizing candidate markers associated with cross-environment response variability. However, because plasticity metrics aggregate responses across multiple environments into genotype-level scalar traits, they do not directly identify environment-specific molecular mechanisms or validated causal G×E loci. Therefore, in this study, plasticity-associated SNPs were used as candidate markers for constructing a computationally tractable G×E feature representation, rather than as confirmed mechanistic G×E loci.

The G2F 2022 Maize G×E Prediction Competition provides a multidimensional dataset comprising phenotypic records, high-density genotypic data, soil properties, meteorological measurements, environmental covariates, and metadata. Compared with the 2024 competition dataset, the 2022 dataset contains denser genomic marker information and is accompanied by published official evaluation results, thereby providing a valuable public benchmark for multi-environment maize yield prediction [24]. The winning CLAC submission used an ensemble strategy in which the final prediction was obtained by averaging two complementary model outputs. The first component was a GBLUP-based genetic main-effect model, whereas the second component decomposed yield prediction into an environmental-mean component and a genotype-specific deviation component. The environmental mean was modeled using an ensemble of random forest, ridge regression, and ordinary least squares, whereas genotype-specific deviations were inferred using a selection-index approach derived from a multivariate GBLUP model with an unstructured G×E covariance structure [25]. This type of strategy, focusing on location means, could bring predicted yields onto a scale closer to the observed values, thereby reducing absolute prediction errors [25]. Although global RMSE is important for evaluating the absolute accuracy of yield prediction, within-environment genotype ranking is particularly relevant for breeding and environment-specific cultivar recommendation. This is because farmers and breeders are often interested in identifying the genotypes that perform best locally rather than those with the highest average performance across all environments. Therefore, model performance in the competition was evaluated from two complementary perspectives: average within-environment predictive ability and global predictive accuracy. Pearson correlation coefficient (PCC) and root mean square error (RMSE) were used as the corresponding evaluation metrics, respectively.

However, existing approaches still face clear limitations in simultaneously achieving high predictive accuracy, computational efficiency, and biological interpretability. To address these challenges, this study proposes a two-step predictive framework that explicitly separates environmental main effects from genotype-dependent residual variation. The framework combines machine learning for nonlinear environmental modeling with genomic relationship matrices for polygenic small-effect prediction, with the aim of improving global predictive accuracy and, more importantly, within-environment genotype ranking under strong genotype-by-environment interaction. Furthermore, to better model genotype-dependent residual variation after accounting for environmental mean effects, we combine an additive-dominance genomic relationship matrix (arc-kernel matrix) with an explicit G×E matrix. To construct a biologically motivated and computationally efficient G×E representation, we use six complementary phenotypic plasticity metrics—coefficient of variation (CV), standard deviation (SD), median absolute deviation (MAD), 10th percentile (Q10), and the slope and root mean square error (RMSE) derived from Finlay–Wilkinson reaction norm regression—to select candidate interaction SNPs, thereby reducing noise, lowering matrix rank and computational burden, and providing a candidate basis for modeling genotype-dependent environmental responses. To ensure fair comparison with existing studies, we use the G2F 2022 maize G×E prediction competition dataset and follow the official evaluation metrics, allowing our framework to be assessed against both the benchmark GBLUP model and the competition-winning approach. In addition, from a practical breeding perspective, we quantify model utility by estimating both overall selection gain and environment-specific selection gain [26], thereby linking predictive performance to breeding-oriented decision-making. Finally, we construct a genotype–environment interaction network as a post hoc exploratory analysis to provide biological context for representative statistical interaction patterns and to generate candidate hypotheses for future validation.

2. Materials and Methods

2.1. Dataset and Study Material

This study utilized the publicly available Genomes to Fields (G2F) 2022 maize multi-environment genomic prediction dataset. The training set spans the years 2014–2021 and includes 217 environments (defined as location × year) and 4683 hybrid lines, while the test set corresponds to the year 2022 and comprises 26 environments and 548 hybrid lines. The response variable is grain yield standardized to a uniform moisture content. In addition to field phenotypic observations, the dataset includes management and site metadata, soil physicochemical properties, daily weather records, and environmental covariates (ECs) derived from a simplified APSIM pipeline. A genotype VCF file shared across both training and test sets is also provided, which is used for constructing genomic relationship matrices and extracting candidate loci for interaction modeling.

2.2. Data Preprocessing

For metadata processing, we followed the publicly available CLAC workflow from the G2F 2022 maize G×E prediction competition [27]. Specifically, irrigation information was extracted from the original management field (treatment, Trt) and encoded as a binary variable (irrigated). Based on agronomic rules, management categories were consolidated into three groups: drought (Dry), late planting (Late), and standard management (Standard). The first two characters of the environment code were parsed to represent the state, whereas the first four characters were used to define the station. In addition, the previous crop (PC) variable was recoded into three broad categories: Legume, Wheat, and Other.

Genotypic data were processed using PLINK2 for quality control, including filtering for minor allele frequency (MAF > 0.1) and marker missing rate (GENO < 0.1). The quality-controlled marker set was used for construction of the K2X-transformed additive–dominance genomic feature matrix and for plasticity GWAS. For candidate SNP selection after GWAS, LD clumping was subsequently applied to remove locally redundant association signals.

2.3. Single-Stage GBLUP Baseline Model

To provide a conventional genomic prediction baseline, we implemented a single-stage genomic best linear unbiased prediction (GBLUP) model. In this model, hybrid lines were treated as random effects, with K2X used as the reduced additive–dominance genomic feature representation of hybrids. Specifically, K2X denotes the reduced additive–dominance genomic feature matrix obtained using the K2X procedure provided in the CLAC source code [27]. Fixed effects included the processed metadata variables, namely state, treatment (Trt), previous crop (PC), and irrigation status (irrigated). Given the large number of levels and sparsity associated with the station factor, it was not directly included as a high-dimensional categorical variable. Instead, it was transformed into a station frequency variable based on summary statistics from the training set, thereby improving estimation stability. The R package "bWGR" [28] was used for fitting the GBLUP model.

2.4. Two-Step Prediction Framework

The core methodology of this study is an environment mean–residual two-step prediction framework. In the first step, environmental main effects are modeled at the environment level using processed metadata and environmental covariates derived from the raw datasets, with the mean yield of each training environment as the response variable, to predict the mean yield of each target environment. The optimal stage-1 model is selected by five-fold cross-validation within the training set, and once fixed, all stage-2 models share the same set of predicted environmental means, thereby ensuring a fair comparison among methods. In the second step, residuals, defined in the training set as the difference between observed yield and the corresponding environmental mean, are used as the response variable to fit and compare models based on genomic information only, genomic information plus an explicit G×E matrix, and genomic information plus a scaled G×E component. To avoid information leakage, candidate SNP selection, matrix construction, and scaling parameter estimation are all performed using the training data only before being applied to the test set.

Let $\mu(e)$ denote the environmental mean for environment e in the training set, and let $y(h, e)$ represent the observed yield of hybrid h in environment e. The training residual is defined as

$$r(h, e) = y(h, e) - \mu(e).$$

In the test set, the final prediction is obtained by combining the predicted environmental mean and the predicted residual:

$$\hat{y}(h, e) = \hat{\mu}(e) + \hat{r}(h, e).$$

Figure 1 summarizes the overall workflow of the proposed two-step prediction framework.

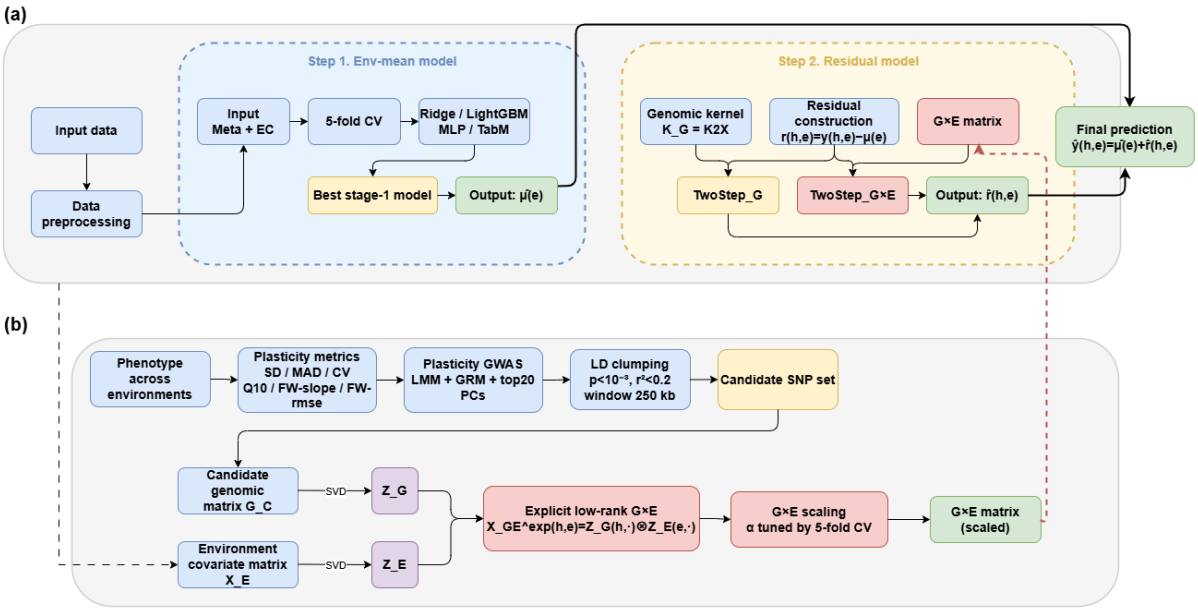

Figure 1. Overview of the proposed two-step prediction framework. (**a**) Overall workflow of the framework. Stage 1 predicts environment-specific mean yield using processed metadata and environmental covariates. Stage 2 models the remaining genotype-dependent residual variation using the K2X-transformed genomic representation and, where applicable, an explicit G×E component. The final yield prediction is obtained by adding the predicted environmental mean and the predicted residual. (**b**) Construction of the explicit G×E representation. Candidate interaction markers are selected using plasticity-based GWAS, reduced genetic and environmental representations are obtained by SVD, and their Kronecker product is used to form the explicit G×E feature matrix. The predicted G×E component is then scaled by the cross-validation-selected parameter α before final residual prediction.

2.5. Stage 1: Environmental Mean Prediction Model

The inputs for stage 1 include the processed metadata variables (state, Trt, PC, and irrigated) as well as the environmental covariate matrix. Although environmental mean prediction was also used in the CLAC workflow, our stage-1 procedure was implemented as a cross-validation-based model-selection step rather than as a fixed ensemble-based procedure [27]. We evaluated a range of candidate models, including a linear model (ridge regression), a tree-based model (LightGBM; Light Gradient Boosting Machine), and neural network-based models, namely a multilayer perceptron (MLP) and TabM (Tabular Deep Learning with Parameter-Efficient Ensembling) [29]. Five-fold cross-validation was performed using the training set only. The optimal model for environmental mean prediction was selected using the average cross-validated RMSE as the primary criterion, because errors in environment-level mean prediction directly affect the global prediction error of the final two-step model. PCC was retained as a secondary diagnostic metric to assess whether the predicted environmental means preserved the relative ordering of environments.

For each environment e, the environmental mean yield was modeled as

$$\mu_e = f_m(\mathbf{x}_e) + \eta_e,$$

where μ_e denotes the observed mean yield of environment e in the training set, $\mathbf{x}_e$ denotes the processed metadata and environmental covariates for that environment, $f_m(\cdot)$ represents a candidate environmental mean prediction model, and η_e is the residual error.

The optimal stage-1 model was selected from the candidate model set

$$\mathcal{M} = \{\text{ridge}, \text{LightGBM}, \text{MLP}, \text{TabM}\}$$

using the average cross-validated RMSE:

$$m^* = \arg\min_{m \in \mathcal{M}} \mathrm{RMSE}_{CV}(m).$$

For an unseen target environment $e \in \mathcal{E}_{\text{test}}$, the predicted environmental mean was obtained as

$$\hat{\mu}_e = f_{m^*}(\mathbf{x}_e),$$

where $\mathbf{x}_e$ denotes the metadata and environmental covariates of the test environment, and m^* is the optimal model selected by cross-validation using the training set only.

2.6. Stage 2: Residual Model

First, we constructed a second-stage baseline model without explicit G×E terms (TwoStep_G). This model was fitted to the stage-1 residuals using only additive–dominance genomic feature matrix [27]. Because no environmental interaction terms were included, it served as a direct baseline for evaluating models augmented with explicit G×E structures.

For the TwoStep_G model, the stage-1 residual of hybrid h in environment e was modeled as

$$r(h,e) = \mu_r + \mathbf{x}_G(h)^\top \boldsymbol{\beta}_G + \varepsilon(h,e),$$

where $r(h,e)$ denotes the stage-1 residual, μ_r is the residual intercept, $\mathbf{x}_G(h)$ is the K2X-transformed genomic feature vector of hybrid h, and $\boldsymbol{\beta}_G$ is the vector of random regression coefficients for the genomic components.

To construct candidate G×E features, we first summarized cross-environment phenotypic plasticity at the hybrid level. For each hybrid, grain-yield observations across available training environments, where each environment was defined as a location × year combination, were used to calculate six genotype-level plasticity metrics: standard deviation (SD), coefficient of variation (CV), median absolute deviation (MAD), the 10th percentile (Q10), and the slope and root mean square error (RMSE) derived from Finlay–Wilkinson reaction norm regression. Because G×E is expressed as differential responses of genotypes to environmental variation, SNPs associated with these plasticity traits were considered candidate loci related to environmental sensitivity or stability, rather than formally validated causal G×E loci. Each plasticity metric was then used as a phenotype in a genome-wide association study (GWAS) under a linear mixed-model framework [30]. Population structure was controlled using a genomic relationship matrix together with the first 20 principal components. For each plasticity trait, linkage disequilibrium (LD) clumping was performed in PLINK2 [31], retaining the lead SNP with $p < 0.001$ within a 250 kb window and removing neighboring variants in LD with the lead signal ($r^2 > 0.2$). The union of lead SNPs across all plasticity traits was then used to define candidate plasticity-associated loci for subsequent explicit G×E feature construction. In contrast to multi-trait mixed-model approaches that jointly model phenotypic records from different environments as correlated traits [32], the plasticity-GWAS strategy used here reduced the candidate-screening step to six GWAS scans, one for each plasticity metric. This strategy provides a computationally tractable approach for highly unbalanced genotype-by-environment datasets, because each hybrid's cross-environment response is summarized into genotype-level plasticity traits before GWAS. To avoid information leakage, all plasticity calculations, GWAS analyses, and SNP selection procedures were conducted using the training data only.

Rather than explicitly constructing pairwise products between hundreds of candidate SNPs and all environmental variables, we introduced an explicit low-rank interaction representation. Specifically, the candidate genotype matrix G_C was extracted from the selected loci, centered, and standardized. Singular value decomposition (SVD) was then applied to G_C to obtain a low-dimensional genetic latent representation Z_G. In this study, the rank

of Z_G was 16. Because the candidate SNPs were selected based on hybrid-level plasticity across environments rather than on interactions with specific environmental factors, no environmental covariate screening was performed before constructing X_E. Similarly, the full environmental covariate matrix X_E, including all available environmental covariates, was centered, standardized, and decomposed by SVD to obtain a low-dimensional environmental latent representation Z_E. The retained environment latent representation Z_E had rank 8. The explicit G×E design for hybrid h in environment e was then defined as

$$X_{GE}^{\exp}(h,e) = Z_G(h,\cdot) \otimes Z_E(e,\cdot),$$

where $\otimes$ denotes the Kronecker product. This formulation preserves biological relevance by leveraging loci derived from plasticity GWAS, while the low-rank latent representations reduce noise, alleviate the dimensional expansion associated with full genotype-by-environment interactions, and improve computational efficiency.

The resulting explicit G×E feature matrix was denoted as $X_{GE}^{\exp}$, with rows corresponding to hybrid–environment observations. In the G×E-augmented residual model, this matrix was fitted together with the K2X-transformed genomic feature matrix. Specifically, for hybrid h in environment e, the residual model was written as

$$r(h,e) = \mu_r + \mathbf{x}_G(h)^\top \boldsymbol{\beta}_G + \mathbf{x}_{GE}(h,e)^\top \boldsymbol{\beta}_{GE} + \varepsilon(h,e),$$

where $r(h,e)$ denotes the stage-1 residual, μ_r is the residual intercept, $\mathbf{x}_G(h)$ is the K2X-transformed genomic feature vector of hybrid h, and $\mathbf{x}_{GE}(h,e)$ is the explicit G×E feature vector for the hybrid–environment combination (h,e). The coefficient vectors $\boldsymbol{\beta}_G$ and $\boldsymbol{\beta}_{GE}$ were fitted as random regression effects:

$$\boldsymbol{\beta}_G \sim N(0,\sigma_G^2 I), \qquad \boldsymbol{\beta}_{GE} \sim N(0,\sigma_{GE}^2 I), \qquad \varepsilon(h,e) \sim N(0,\sigma_e^2).$$

The predicted residual was therefore obtained as

$$\hat{r}(h,e) = \hat{\mu}_r + \mathbf{x}_G(h)^\top \hat{\boldsymbol{\beta}}_G + \mathbf{x}_{GE}(h,e)^\top \hat{\boldsymbol{\beta}}_{GE}.$$

2.7. Scaling of the G×E Design Matrix

Previous studies have shown that G×E can be incorporated into genomic prediction models as an additional kernel or random-effect component, and that such models can improve multi-environment prediction by capturing genotype-specific environmental responses [33]. However, directly adding an explicit G×E component does not guarantee improved prediction, because the interaction component may contain both useful signal and noise, especially when it is constructed from high-dimensional genotype–environment combinations. In multi-environment genomic prediction, weighted-kernel approaches have been proposed to regulate the contribution of different genomic or environment-related similarity structures and have been shown to improve prediction performance [34]. Motivated by this idea, we introduced a cross-validation-based scaling parameter α to control the contribution of the predicted G×E component in the residual model.

For the scaled G×E model, the predicted residual was calculated as

$$\hat{r}^{(\alpha)}(h,e) = \hat{\mu}_r + \mathbf{x}_G(h)^\top \hat{\boldsymbol{\beta}}_G + \alpha\, \mathbf{x}_{GE}(h,e)^\top \hat{\boldsymbol{\beta}}_{GE}.$$

where α was applied as a weight on the predicted G×E component rather than as a parameter used to construct the G×E design matrix.

A grid of candidate values,

$$\alpha \in \{-0.50, -0.25, -0.10, 0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.33, 0.50, 0.75, 1.00\},$$

was evaluated using five-fold cross-validation within the training set. For each candidate α, the residual model was fitted on the training folds, and predictions were evaluated on the held-out validation fold after reconstructing final yield as

$$\hat{y}(h,e) = \hat{\mu}(e) + \hat{r}^{(\alpha)}(h,e).$$

Because within-environment genotype ranking was the primary objective of the G×E component, the final α was selected according to the average within-environment PCC from five-fold cross-validation within the training set. After selecting α, the final model was refitted using the full training set and then applied to the 2022 test environments.

2.8. Evaluation Metrics

To comprehensively evaluate model performance from both global and within-environment perspectives, we adopted two sets of evaluation metrics. Global performance was assessed using root mean square error (global RMSE) and Pearson correlation coefficient (global PCC), both calculated across all observations in the test set. Within-environment performance was assessed by calculating RMSE and PCC separately within each test environment and then averaging these values across all test environments to obtain the final within-environment RMSE and within-environment PCC.

2.9. Quantification of Selection Gain

In addition to standard predictive accuracy metrics, we further evaluated the prediction-guided selection utility of each model from the perspective of breeding decision-making. Following the selection-evaluation logic used for enviromically adapted varieties [26], this analysis mimicked the practical use of prediction models, in which predicted yields are used for ranking and selection, while observed yield phenotypes are used to calculate the realized selection gain. A fixed selection intensity of the top 10% was applied. Three scenarios were considered:

(1) No selection: For each test environment, the mean true grain yield across all candidate hybrids was calculated, and the mean and standard deviation were then summarized across test environments.
(2) Global selection: Hybrids were ranked according to their predicted mean performance across environments. The top 10% were selected as globally superior materials, and their observed grain yield mean and standard deviation were then calculated across test environments.
(3) Environment-specific selection: Within each test environment, hybrids were ranked according to their predicted performance in that environment. The top 10% were selected as environment-adapted superior materials, and their observed grain yield mean and standard deviation were then calculated for each environment and summarized across test environments.

2.10. Genotype–Environment Interaction Network and Functional Contextualization

As a post hoc exploratory analysis, we conducted SNP–environmental covariate interaction tests to examine whether candidate SNPs used in the explicit G×E feature construction showed statistical interactions with representative environmental covariates. To identify representative environmental factors, the original 765 environmental covariates were first filtered. This filtering step was intended to reduce redundancy among highly correlated environmental covariates and to obtain a manageable set of variables for exploratory inter-

action testing. Specifically, the top 50 environmental variables ranked by importance in the stage-1 environmental mean model were selected, after which highly correlated variables ($|r| > 0.5$) were removed, resulting in a set of representative environmental covariates.

For each candidate SNP and each representative environmental factor, a multivariate linear regression model including an interaction term was fitted to screen for statistical SNP × environment-factor interaction signals. Candidate SNPs with stronger interaction signals were prioritized for network visualization and downstream annotation. Prioritized candidate SNPs were annotated using the maize B73 reference genome annotation file `Zm-B73-REFERENCE-NAM-5.0_Zm00001eb.1.gff3`. For each prioritized SNP, we first determined whether the SNP was located within an annotated gene region. If the SNP did not overlap any annotated gene, the nearest annotated gene was assigned as the candidate neighboring gene. Gene Ontology (GO) annotations were then summarized for these annotated or nearest genes to provide functional contextualization of the prioritized candidate regions. These annotations were used to describe possible biological context associated with the prioritized statistical interaction signals, but were not treated as functional validation. All SNP–environment interaction tests were conducted using the training data only. Because these tests were performed after candidate SNP selection, the resulting p values were used for prioritization and visualization rather than for confirmatory inference.

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i \mathrm{PC}_i + \beta_{\mathrm{SNP}} \cdot \mathrm{SNP} + \beta_{\mathrm{Env}} \cdot \mathrm{Env} + \beta_{\mathrm{Int}} \cdot (\mathrm{SNP} \times \mathrm{Env}) + \varepsilon \quad (1)$$

where PC_i denotes the i-th principal component used to control for population structure.

3. Results

3.1. Selection of the Optimal Environmental Mean Model Based on Machine Learning and Deep Learning

To capture the nonlinear effects of environmental factors on mean yield, we compared several candidate models based on different learning principles, including ridge regression, LightGBM, MLP, and TabM. Five-fold cross-validation was performed using the training set to evaluate model performance under different hyperparameter settings (Figure 2).

Overall, MLP and TabM showed the weakest performance in predicting environmental mean yield, with relatively high RMSE and low PCC. Ridge regression exhibited unstable performance across different hyperparameter settings. In contrast, LightGBM consistently achieved the best overall trade-off between RMSE and PCC. The optimal model, highlighted in orange with a red box in Figure 2, was a LightGBM model with the following hyperparameters: learning rate = 0.03, number of leaves = 31, minimum child samples = 10, subsample ratio = 0.8, colsample by tree = 1.0, and L_2 regularization (λ) = 1.0. Therefore, this LightGBM model was selected as the stage-1 environmental mean prediction model and was subsequently used in all two-step models analyzed in this study. The full hyperparameter tuning results are provided in Table S1.

3.2. Improved Yield Prediction Performance of the Two-Step Framework

We compared five yield prediction models (Table 1). In this study, all two-step models used LightGBM as the fixed stage-1 environmental mean model. The TwoStep_G model incorporated only additive and dominance genomic information. It achieved a within-environment PCC of 0.371 and a global PCC of 0.650, outperforming both GBLUP (0.357 and 0.543, respectively) and CLAC (0.357 and 0.631) in terms of PCC.

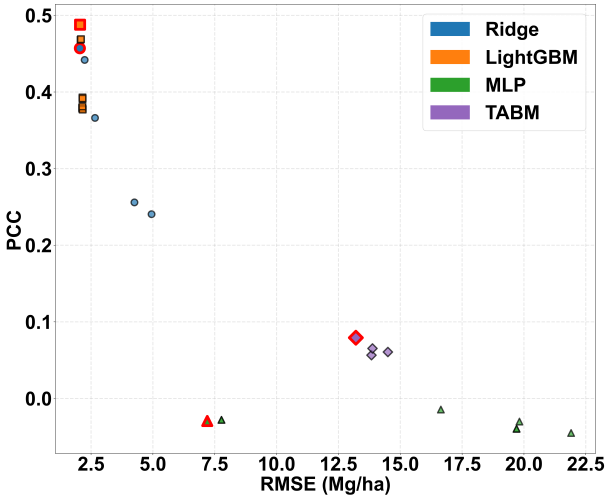

Figure 2. Performance comparison of candidate models for predicting environment-level mean yield. Each point represents one hyperparameter setting of a candidate model evaluated by cross-validation. Colors and marker shapes indicate different model classes: Ridge regression, LightGBM, multilayer perceptron (MLP), and TabM. The red-outlined point highlights the selected hyperparameter setting for each model class. The x-axis shows prediction error measured by root mean square error (RMSE; Mg/ha), and the y-axis shows prediction agreement measured by Pearson correlation coefficient (PCC). Models closer to the upper-left corner have lower prediction error and higher correlation, indicating better overall performance for environment-mean prediction.

After introducing explicit candidate G×E interactions, the TwoStep_G×E model showed slightly improved global performance relative to TwoStep_G, but its within-environment performance decreased (PCC = 0.359 vs. 0.371). This result suggests that directly incorporating the raw candidate interaction matrix may introduce additional noise.

Based on TwoStep_G×E, we further applied a cross-validation-selected scaling parameter to the predicted G×E component. The scaling parameter α was tuned by five-fold cross-validation using the training set, and the detailed tuning results are shown in Figure S2. Because within-environment ranking was treated as the primary objective, $\alpha = 0.33$ was selected as the final value. The resulting model, TwoStep_G×E_alpha0.33, achieved the highest within-environment PCC (0.376), together with a within-environment RMSE of 2.362, a global RMSE of 2.506, and a global PCC of 0.654. Compared with the conventional GBLUP model, TwoStep_G×E_alpha0.33 reduced within-environment RMSE by 13.0%, increased within-environment PCC by 5.5%, reduced global RMSE by 14.7%, and improved global PCC by 20.4%. Compared with CLAC, it achieved higher within-environment PCC and global PCC, although CLAC retained a lower global RMSE.

To further evaluate whether the proposed framework was robust beyond the original 2022 test set, we performed an additional environment–level five–fold cross–validation analysis using the G2F data. Because the two–step framework requires environmental covariates to predict environment-level mean yield, the data were split by environments rather than by individual plot records. After applying the same preprocessing and retaining environments with available phenotype and environmental covariate information, 186 environments from 2014–2022 were randomly divided into five folds. In each fold, approximately 80% of the environments were used for training and the remaining environments were used for validation. The same fold partition was used for GBLUP, TwoStep_G, TwoStep_G×E, and TwoStep_G×E_alpha to ensure a fair comparison.

Table 1. Evaluation and comparison of maize yield prediction models.

Model	Within-Environment		Global	
	RMSE	PCC	RMSE	PCC
GBLUP	2.713	0.357	2.937	0.543
CLAC	2.329	0.357	2.458	0.631
TwoStep_G	2.365	0.371	2.513	0.650
TwoStep_G×E	2.367	0.359	2.505	0.653
TwoStep_G×E_alpha0.33	2.362	0.376	2.506	0.654

Note: Within-environment metrics represent the average RMSE and PCC calculated separately within each test environment. Global metrics were calculated across all test observations. The unit of RMSE is Mg ha^{-1}.

For TwoStep_G×E and TwoStep_G×E_alpha, plasticity traits, plasticity GWAS, candidate SNP selection, and low-rank G×E matrix construction were repeated within each fold using only the fold-specific training environments. The held-out environments were used only for evaluation. Across the five folds, the average numbers of training and validation environments were 148.8 and 37.2, respectively. GBLUP achieved the lowest mean RMSE values, with a global RMSE of 2.474 ± 0.131 and a within-environment RMSE of 2.346 ± 0.154. However, the two-step models showed higher average correlation-based performance. TwoStep_G achieved a global PCC of 0.539 ± 0.069 and a within-environment PCC of 0.349 ± 0.035, compared with 0.515 ± 0.079 and 0.336 ± 0.033 for GBLUP. Incorporating the explicit G×E component further improved within-environment ranking, with TwoStep_G×E and TwoStep_G×E_alpha achieving mean within-environment PCC values of 0.351 ± 0.035 and 0.352 ± 0.035, respectively. These results indicate that the proposed two-step framework was not specific to the original 2022 test set and maintained improved correlation-based ranking performance across environment-level validation folds. Detailed fold-specific results, selected environment-mean models, fold-specific candidate SNPs, and five-fold mean ± SD summaries are provided in Table S6.

3.3. Model Performance Across Different Environments

Model performance varied substantially across test environments (Figure 3). To facilitate comparison across environments, test environments were ordered according to their observed mean yield, from low to high. We evaluated model performance using both RMSE and PCC, which reflect different aspects of prediction quality. Within a given environment, RMSE is affected by both stage-1 environmental mean prediction and stage-2 residual prediction, whereas PCC is invariant to the environment-specific mean offset and therefore mainly reflects the ability of the residual model to rank genotypes within that environment.

Overall, GBLUP showed the largest RMSE gap relative to the two-step models. Although GBLUP achieved lower RMSE than the proposed two-step models in a few environments, such as NCH1_2022 and NYH2_2022, its RMSE was higher than that of the two-step approaches in most environments. This pattern suggests that the nonlinear stage-1 environmental mean model generally provided better environmental fit than the linear formulation used in GBLUP.

Differences in PCC across environments were comparatively smaller. Because both GBLUP and the two-step residual models used linear structures for residual fitting, the differences among models in within-environment PCC remained modest in many environments. After introducing the explicit G×E matrix, performance changes were limited in most environments and even declined in some cases, such as WIH3_2022, suggesting that directly incorporating the raw G×E matrix may introduce additional noise. After G×E component scaling, within-environment PCC generally improved relative to the un-

scaled TwoStep_G×E model, indicating that scaling enhanced the stability of environment-specific ranking.

Nevertheless, the G×E-component-scaled model TwoStep_G×E_alpha0.33 generally outperformed the unscaled TwoStep_G×E model in most environments, indicating that G×E-component-scaled provided a more stable improvement in within-environment ranking ability. However, in certain environments, such as IAH2_2022, the unscaled explicit interaction model still performed better, suggesting that environment-specific interaction structures remain complex and may not be fully captured by a uniform scaling strategy.

To further explore the source of environment-specific variation in ranking performance, we examined the relationship between within-environment PCC and two potential explanatory factors: the observed mean yield of each test environment and the number of hybrids in each test environment that overlapped with the training set (Figure S3). Within-environment PCC tended to be higher in environments with higher mean yield, suggesting that genotype ranking was more predictable in high-yield environments. In contrast, lower-yield environments may involve stronger environmental constraints or stress-related effects, which could reduce the stability of genotype ranking and increase the difficulty of prediction. By contrast, the number of overlapping hybrids between the training and test sets showed no clear association with within-environment PCC, suggesting that the environment-specific variation in prediction performance was unlikely to be mainly driven by unequal representation of shared hybrids.

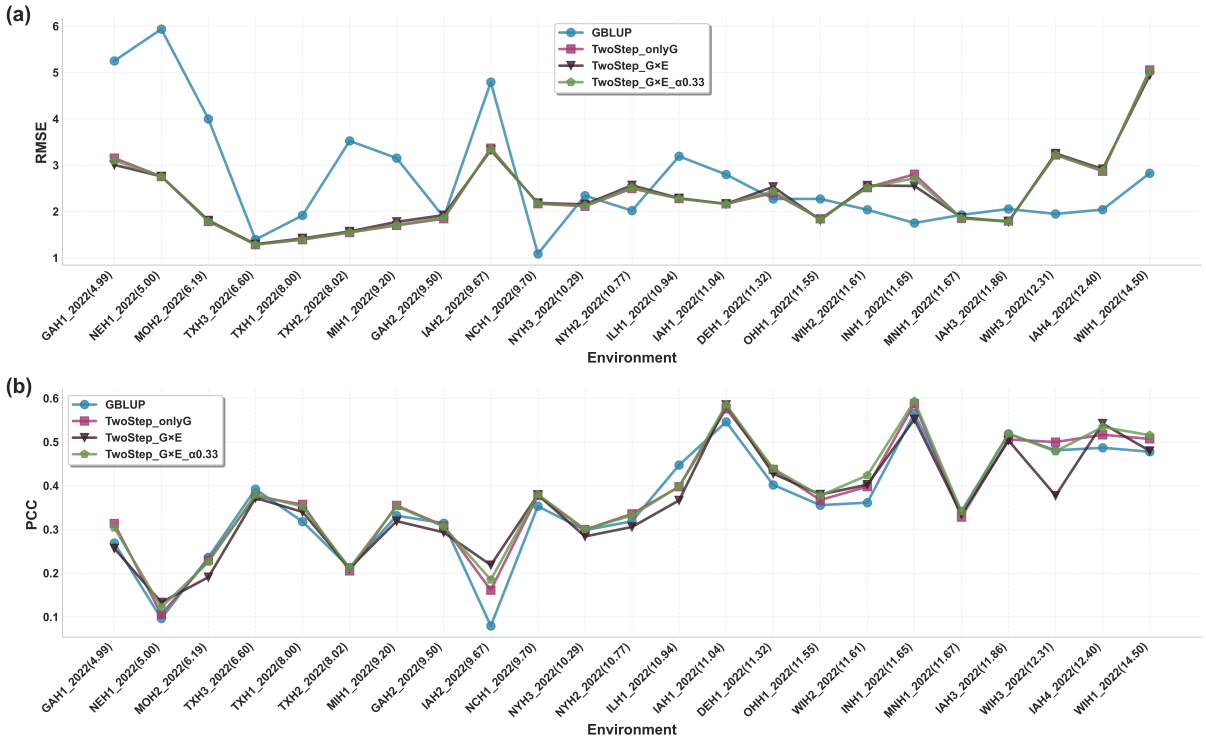

Figure 3. Environment-specific prediction performance of selected models in the test set. Each point represents the performance of one model in one test environment, where each environment is defined as a location × year combination and is labeled on the x-axis. The number in parentheses after each environment name indicates the observed mean grain yield of that environment (Mg/ha). Four models are compared: the benchmark GBLUP model, the two-step model using only genomic main effects after environment-mean correction (TwoStep_onlyG), the two-step model with an unscaled explicit G×E component (TwoStep_G×E), and the two-step model with a scaled G×E component

using $\alpha = 0.33$ (TwoStep_G×E_α0.33). (**a**) Root mean square error (RMSE; Mg/ha) across test environments, where lower values indicate smaller prediction errors. (**b**) Pearson correlation coefficient (PCC) between observed and predicted yield within each test environment, where higher values indicate better within-environment genotype ranking.

3.4. Selection of Candidate G×E Loci Based on Phenotypic Plasticity

To construct an explicit G×E matrix, we first developed complementary phenotypic plasticity indices across different training environments and performed genome-wide association studies (GWAS) for each index (Figure 4). These plasticity indices were designed to capture different aspects of genotype-level environmental response, including dispersion, robustness, lower-tail performance, and Finlay–Wilkinson regression-based sensitivity. The GWAS results obtained from different indices exhibited both complementarity and consistency. By integrating association signals across multiple plasticity measures, we identified candidate loci associated with G×E effects for downstream G×E feature construction.

Because the plasticity GWAS was used for candidate marker screening rather than confirmatory locus discovery, we further evaluated the calibration of GWAS *p*-values. QQ plots were generated for all six plasticity traits (Figure S4). The genomic control inflation factors (λ_{GC}) ranged from 0.883 to 0.961 across the six plasticity GWAS analyses, indicating no evidence of systematic *p*-value inflation. The QQ plots also showed that the observed *p*-value distributions were generally well calibrated, with no global upward deviation from the null expectation.

We also evaluated the results under strict multiple-testing correction. After genotype quality control using MAF > 0.1 and GENO < 0.1, a total of 264,720 SNPs were retained for GWAS. No SNPs passed Bonferroni correction or BH-FDR < 0.05 for any of the six plasticity traits. Therefore, the loci retained from the plasticity GWAS were not interpreted as genome-wide significant or validated causal G×E loci. Instead, they were used as suggestive candidate markers for downstream G×E feature construction.

To justify the use of $p < 10^{-3}$ as the screening threshold, we performed a threshold sensitivity analysis using three candidate thresholds: $p < 10^{-4}$, $p < 10^{-3}$, and $p < 10^{-2}$ (Table S5). For each threshold, SNPs passing the threshold in each plasticity GWAS were further LD-pruned using PLINK clumping with $r^2 = 0.2$ within a 250-kb window, and the clumped lead SNPs were then unioned across the six plasticity traits. The more stringent threshold of $p < 10^{-4}$ retained only 62 LD-clumped union lead SNPs, which may discard potentially useful plasticity-related signals. In contrast, the more permissive threshold of $p < 10^{-2}$ retained 3,206 LD-clumped union lead SNPs, substantially increasing feature dimensionality and the potential for noise. The intermediate threshold of $p < 10^{-3}$ retained 398 LD-clumped union lead SNPs, providing a practical balance between signal retention, LD redundancy reduction, and feature-set size (Table S5).

To remove local redundancy due to linkage disequilibrium, LD clumping was applied ($p_1 = 10^{-3}$, $r^2 = 0.2$, window size = 250 kb). The results showed that each plasticity metric identified a distinct number of LD-clumped lead SNPs, including SD (98), MAD (79), CV (82), Q10 (75), FW_slope (55), and FW_rmse (68). After integration across all indices, a total of 398 LD-pruned candidate SNPs were retained, providing the basis for subsequent explicit G×E modeling. Detailed information on these 398 SNPs is provided in Table S2, and their genomic distribution is shown in Figure S1.

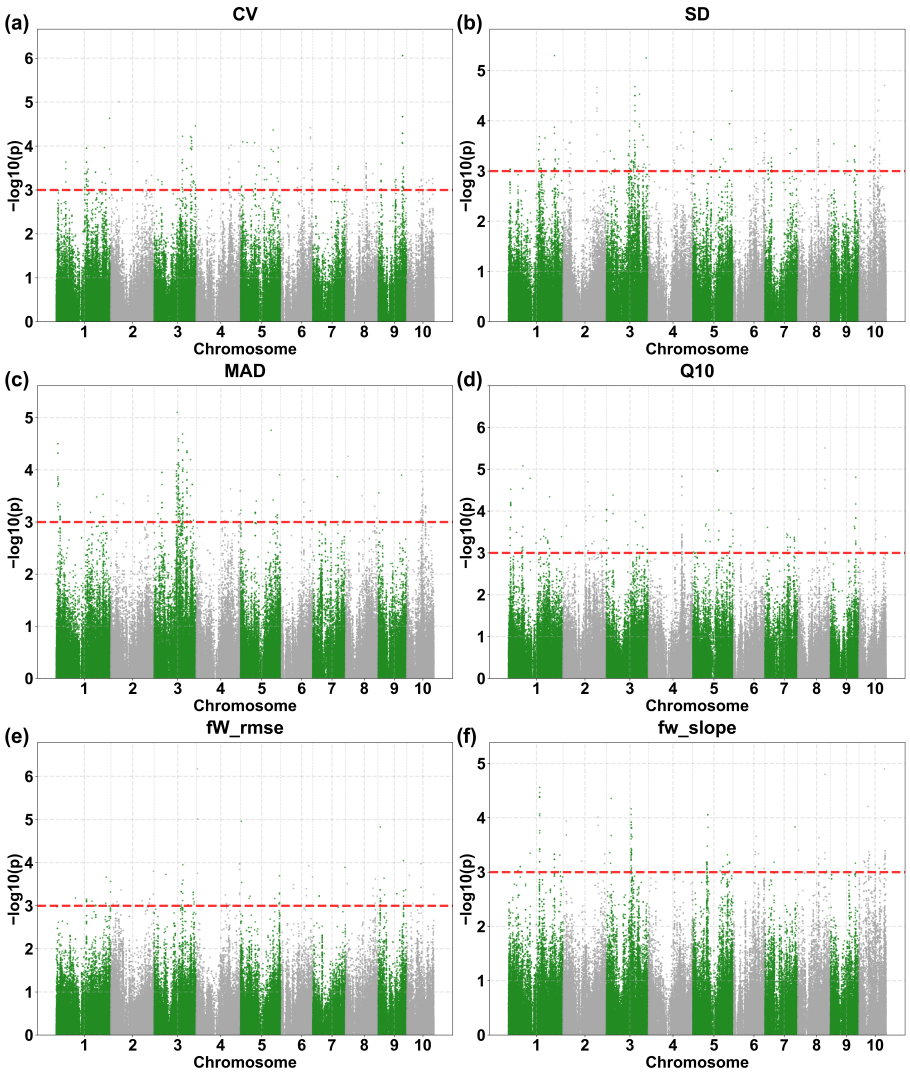

Figure 4. Plasticity-based GWAS screening for candidate markers used in G×E feature construction. Six genotype-level plasticity metrics were calculated from cross-environment yield performance in the training set: (**a**) coefficient of variation (CV), (**b**) standard deviation (SD), (**c**) median absolute deviation (MAD), (**d**) 10th percentile (Q10), (**e**) root mean square error from Finlay–Wilkinson regression (FW_rmse), and (**f**) slope from Finlay–Wilkinson regression (FW_slope). In each Manhattan plot, the x-axis represents genomic position ordered by chromosome, and the y-axis represents $-\log_{10}(p)$ from the GWAS test. Alternating green and gray points indicate SNPs on adjacent chromosomes. The red dashed line indicates the suggestive screening threshold of $p = 10^{-3}$ applied before linkage disequilibrium (LD) clumping. This threshold was used for candidate G×E feature construction rather than for declaring genome-wide significant loci. The retained lead SNPs were treated as candidate plasticity-associated markers rather than validated causal G×E loci.

3.5. Environmental Adaptation-Based Selection Improves Breeding Gains

To quantify the practical breeding value of different yield prediction models, we compared the mean selection gains and their variability across environments under two strategies: overall selection and environment-adaptive selection (Table 2). Under the no-selection scenario, the mean yield across the test set was 9.947 Mg ha^{-1}. Across all models, environment-adaptive selection consistently outperformed overall selection, with the largest difference observed for GBLUP, for which adaptive selection provided an additional gain of 0.273 Mg ha^{-1}.

The TwoStep_G×E_alpha0.33 model achieved the highest gains under both selection strategies. Under overall selection, it yielded a gain of 0.379 Mg ha^{-1}, representing an approximately sixfold increase over GBLUP and a 41.4% improvement over TwoStep_G. Under environment-adaptive selection, it achieved a gain of 0.454 Mg ha^{-1}, corresponding to a 34.7% improvement over GBLUP and a 4.8% improvement over TwoStep_G. These results indicate that incorporating candidate G×E effects, particularly after G×E component scaling, improves not only predictive performance but also the practical utility of the model for environment-specific breeding decisions.

Table 2. Quantitative assessment of selection gain under overall and environment-adaptive selection strategies.

Model	Overall Selection		Adaptive Selection	
	Mean ± SD	Gain	Mean ± SD	Gain
GBLUP	10.011 ± 2.770	0.064	10.285 ± 2.573	0.337
TwoStep_G	10.215 ± 2.553	0.268	10.381 ± 2.609	0.433
TwoStep_G×E	10.307 ± 2.594	0.359	10.349 ± 2.606	0.402
TwoStep_G×E_alpha0.33	10.326 ± 2.576	0.379	10.401 ± 2.617	0.454

Note: Units are Mg ha^{-1}. Mean and standard deviation (SD) were calculated within each test environment before aggregation. Gain was defined relative to the no-selection scenario.

3.6. Genotype–Environment Interaction Network and Biological Contextualization

Because the 398 candidate interaction SNPs identified in the previous analysis were not directly linked to specific environmental factors, we further conducted SNP–environmental covariate interaction tests as an exploratory, hypothesis-generating analysis to provide biological context for the explicit G×E model. Based on the ranking of environmental covariates in the stage-1 environmental mean model, the top 50 environmental factors were first selected. After applying a PCC threshold of 0.5 to remove highly correlated variables, a set of 22 representative and relatively independent environmental factors was obtained. These factors were then used for combinatorial interaction analysis with the 398 candidate SNPs. Detailed information on the 22 representative environmental variables is provided in Table S3.

The heatmap of environmental variable importance (Figure 5) shows that the 22 environmental factors are distributed across seven of the nine phenological stages. Among them, three stages exhibited relatively higher average importance: from flowering to start of grain filling (pFlwStG), start of grain filling to end of grain filling (pStGEnG), and flag leaf to flowering (pFlaFlw), with mean importance values of 584.72, 401.44, and 338.43, respectively. Previous studies have shown that developmental stages closer to pollination and fertilization can have increasingly irreversible effects on final yield [12]. Our results are broadly consistent with these observations and therefore provide supportive, although not confirmatory, evidence for the biological relevance of the identified environmental windows.

We then performed interaction tests between the 22 environmental factors and the 398 candidate SNPs (Table S4). The top three statistically significant G×E interactions were all associated with the duration of soil evaporation during the second stage of grain filling (`T_pStGEnG`) (Figure 6a–c). Although these interaction signals were highly ranked statistically, their effect sizes were modest, with partial R^2 values ranging from 0.0130 to 0.0144 for the top three interactions, indicating that each interaction explained approximately 1.3–1.4% of the phenotypic variance. The interaction plots suggest that, under environments with prolonged evaporation duration during this stage, yield differences among genotypes tend to converge. This pattern is consistent with the possibility that

sustained soil water loss strengthens environmental main effects while partially attenuating the expression of genetic variation. In contrast, under short or moderate evaporation duration, genotypes `3:186982406_AA`, `7:18193876_TT`, and `2:1538358_AA` tended to show higher yield performance. Meanwhile, `2:1538358_GG` showed relatively high yield under Q3 (1.63), suggesting a potentially nonlinear response of this locus to soil water-loss gradients, in which certain genotypes may exhibit locally adaptive advantages within specific environmental ranges.

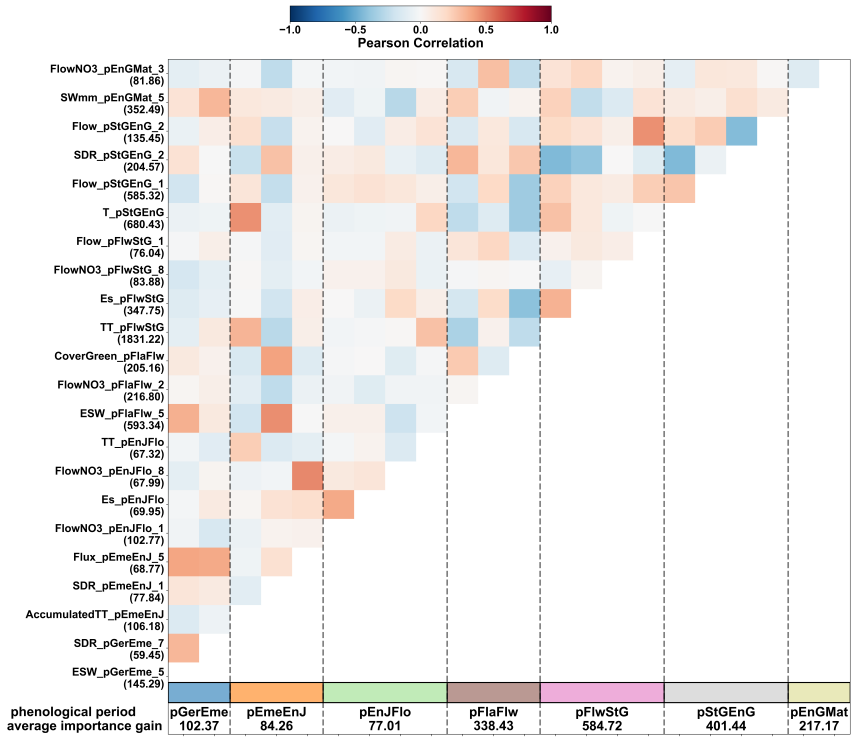

Figure 5. Correlation structure and importance of the 22 representative environmental factors used for exploratory SNP–environmental covariate interaction analysis. Rows represent environmental covariates retained after filtering the top 50 variables from the stage-1 environmental mean model and removing highly correlated variables. Columns represent the 22 retained environmental covariates, grouped by phenological stage. The heatmap color indicates the Pearson correlation coefficient between pairs of environmental covariates, with blue representing negative correlation, white representing weak or no correlation, and red representing positive correlation. Dashed vertical lines separate different phenological stages. The colored bar at the bottom indicates the corresponding phenological period, and the number below each period denotes the average importance score of environmental variables within that period in the stage-1 environmental mean model.

A highly ranked statistical interaction signal was also observed between ESW_pGerEme_5 and the variant 8:177190067:T:C, with a partial R^2 of 0.0123, indicating that this interaction explained approximately 1.2% of the phenotypic variance. This locus is located approximately 196 bp upstream of the transcription factor gene *wox13b*. Given the reported roles of WOX family genes in plant developmental regulation and abiotic stress responses, together with the importance of soil moisture during germination and emergence for seed germination, seedling uniformity, and early root establishment [35], this locus may represent a plausible candidate associated with maize adaptation to early-stage soil moisture conditions. The interaction plot (Figure 6d) shows that the CC genotype achieved higher yield under elevated ESW_pGerEme_5 conditions, whereas the TC genotype performed better under lower ESW_pGerEme_5 conditions, indicating that the favorable allele at this locus may be environment-dependent.

Importantly, these interaction results should be interpreted with caution. Although several signals reached extremely strong statistical significance, they were derived from a post hoc screening framework and have not been validated by independent biological experiments. Therefore, the biological interpretations presented here should be regarded as exploratory and hypothesis-generating rather than confirmatory evidence of causal mechanisms.

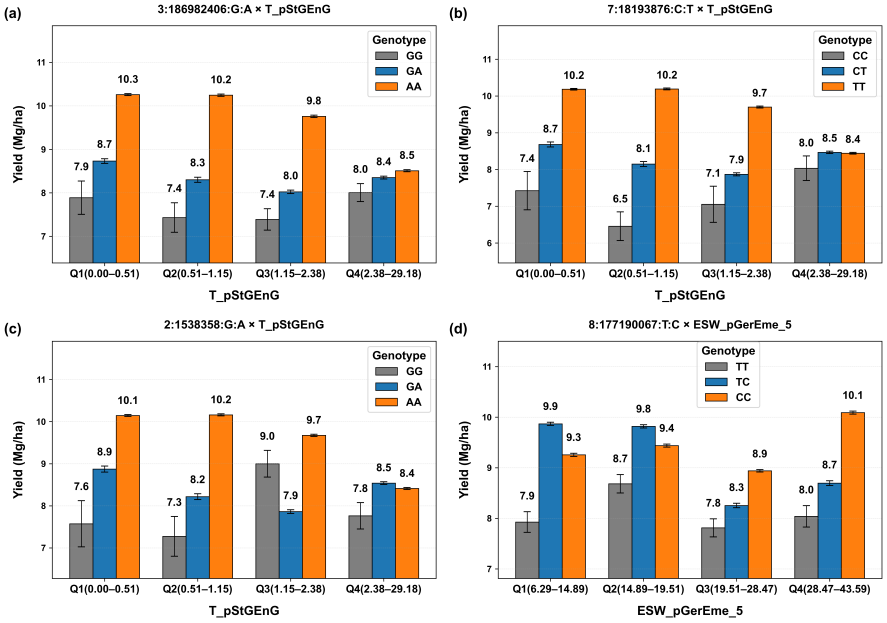

Figure 6. Representative visualizations of prioritized statistical SNP × environmental covariate interaction signals. Each panel shows one SNP–environmental covariate combination: (**a**) 3:186982406:G:A × T_pStGEnG, (**b**) 7:18193876:C:T × T_pStGEnG, (**c**) 2:1538358:G:A × T_pStGEnG, and (**d**) 8:177190067:T:C × ESW_pGerEme_5. Continuous environmental covariates were divided into four quantile groups (Q1–Q4), and the value range of each group is shown in parentheses on the x-axis. Bars indicate the mean observed grain yield (Mg/ha) for each genotype class within each environmental quantile group, error bars indicate standard errors, and the numbers above the bars show the corresponding mean yields. These plots were used to summarize genotype-specific response patterns across environmental gradients for selected interaction signals identified in the post hoc analysis.

4. Discussion

4.1. Rationale and Limitations of the Two-Step Framework

The framework separates components with different effect-size distributions and data structures, rather than strictly decomposing sources of variation. Compared with existing multi-environment prediction methods, the core idea of the proposed framework is to explicitly decompose yield variation into environment-level mean effects and genotype-dependent residual variation. Environment-level mean yield is often driven by observable environmental factors, such as weather, soil, management, and site-year conditions, which can exert relatively large effects shared by all hybrids within the same environment [10]. In contrast, genotype-dependent deviations from the environmental mean are more closely associated with genetic main effects and G×E interactions. For complex traits such as yield, these genetic effects are typically distributed across many loci with small effects [13], making genomic relationship or kernel-based models suitable for residual prediction. This difference in effect-size distributions and data structures provides the main rationale for the two-step framework.

Notably, the framework employs a dominance-aware genomic relationship matrix (arccosine kernel) to model genetic main effects, which enables the simultaneous capture of additive and dominance effects. The importance of dominance effects in contributing to heterosis in hybrid maize has been well documented [14]. However, this formulation does not explicitly account for epistatic interactions among loci, which may introduce modeling bias for traits controlled by a small number of major-effect loci with strong epistasis. Because grain yield is a complex trait, epistatic effects may also contribute to genotype performance, especially under stress conditions. Future work could explore extensions of the proposed framework that incorporate epistatic kernels, computationally efficient marker-pair interaction screening, RKHS or deep-kernel models, and multi-kernel formulations. These approaches may help capture higher-order genetic interactions and jointly model additive, dominance, epistatic, and G×E effects, although their computational cost and interpretability would need to be carefully evaluated.

Compared with the competition-winning model CLAC, the proposed framework achieved a higher within-environment PCC (0.376 vs. 0.357), indicating better ranking performance within environments, which is a key objective in breeding applications. It is worth noting that CLAC attained a slightly lower global RMSE (2.458 vs. 2.506), likely due to its reliance on weighting and ensemble strategies based on training–testing environmental similarity [25]. While such strategies may improve overall mean prediction accuracy, they may also be more dependent on the similarity structure between observed and target environments. In contrast, the proposed framework does not rely on such weighting assumptions and may therefore offer a more generalizable alternative for scenarios involving novel or underrepresented environments, which are common in practical breeding programs.

Beyond the competition-winning CLAC model, recent multi-environment maize prediction studies have used different strategies to incorporate environmental information and G×E. Lopez-Cruz and de los Campos proposed a multi-trait/environment sparse genomic prediction framework (MT-SGP) that combines sparse selection indices with sparse genomic prediction to borrow information from selected correlated traits or environments and genetically related training genotypes [36]. This strategy can avoid forcing all environments or traits into the prediction equation and may improve prediction when useful correlation structures exist among environments. However, similar to environment-similarity-based weighting strategies, its effectiveness depends on the availability of informative related environments or traits and sufficient genetic correlation among them. In addition, because the framework treats multiple environments as correlated traits and constructs candidate-specific sparse prediction equations, computational cost may increase when the number of environments, traits, or training genotypes becomes large. In contrast, our framework focuses on separating environment-level mean prediction from genotype-dependent residual prediction and introducing an explicit but scaled G×E feature representation, providing a complementary strategy for environment-specific genomic prediction. Hu et al. extended MegaLMM for genomic prediction in new environments by learning regressions of latent factor loadings on environmental covariates within a large-scale multivariate linear mixed-model framework [37]. This strategy effectively borrows information across correlated trials and uses environmental covariates to extrapolate genetic values to new environments. However, the G×E structure is represented through latent factors and environment-level covariance rather than an explicit candidate SNP-by-environment feature matrix. In contrast, our framework first separates environment-level mean prediction from genotype-dependent residual prediction and then introduces a scaled explicit G×E feature representation, providing a complementary strategy for environment-specific genomic prediction. Li et al. recently evaluated hybrid deep learning architectures, including CNN-LSTM, CNN-ResNet, LSTM-ResNet, and CNN-ResNet-LSTM, for crop genomic

prediction across wheat, rice, and maize datasets [38]. Their results showed that LSTM-ResNet achieved the best prediction accuracy for 10 of 18 traits, highlighting the potential of hybrid deep learning models for capturing nonlinear genotype–phenotype relationships. However, their framework mainly focused on SNP-based phenotype prediction and did not explicitly model multi-environment G×E structure or separate environmental main effects from genotype-dependent residual variation. In contrast, our framework is designed for multi-environment yield prediction by separating environment-level mean prediction, genomic residual prediction, and explicit scaled G×E feature construction.

4.2. Advantages of LightGBM for High-Dimensional Environmental Covariates

In the first-stage environmental mean prediction, LightGBM consistently outperformed ridge regression, MLP, and TabM in both predictive accuracy and robustness across hyperparameter settings. This finding is consistent with previous reports highlighting the strong performance of gradient boosting frameworks in crop phenotypic prediction tasks [18]. From a methodological perspective, the superiority of LightGBM may be attributed to two main factors. First, its histogram-based splitting strategy efficiently captures sparse and complex interactions among high-dimensional environmental covariates (765 features in this study) without requiring explicit parametric assumptions about feature relationships. Second, its ensemble learning structure provides inherent robustness to outlier environments, enabling the model to learn dominant trends without overfitting extreme environmental samples.

In contrast, deep learning-based models such as MLP and TabM performed suboptimally in this study, exhibiting higher RMSE and lower PCC. This observation is consistent with the findings of Kelly & McLaughlin (2024) [19], who reported that for traits governed by numerous small-effect additive components, complex neural network architectures do not systematically outperform classical methods and may suffer from increased overfitting risk in small-sample, high-dimensional settings. Ridge regression showed strong sensitivity to the choice of regularization parameters, indicating limited capacity to capture the nonlinear response structure of yield to environmental covariates.

4.3. Noise in G×E Integration and the Role of G×E Component Scaling

After directly incorporating the candidate G×E matrix into the prediction model, no substantial improvement in overall predictive performance was observed, and a decrease in within-environment PCC was detected (TwoStep_G×E: 0.359 vs. TwoStep_G: 0.371). This result highlights a common challenge in high-dimensional G×E modeling: explicit interaction terms may contain both useful genotype-specific environmental signals and additional noise. When the number of candidate interaction features is large relative to the effective number of training environments, directly fitting the full G×E component may reduce ranking stability within environments. Similar overfitting issues have been reported in other G×E integration studies, where common mitigation strategies include kernel-based methods such as RKHS [39] and shrinkage-based modeling approaches.

In this study, we introduced a cross-validation-based scaling parameter α to control the contribution of the predicted G×E component in the residual model. Rather than modifying the construction of the G×E design matrix itself, α was applied after model fitting as a weight on the predicted G×E residual component:

$$\hat{r}^{(\alpha)}(h,e) = \hat{\mu}_r + \mathbf{x}_G(h)^\top \hat{\boldsymbol{\beta}}_G + \alpha\, \mathbf{x}_{GE}(h,e)^\top \hat{\boldsymbol{\beta}}_{GE}.$$

Thus, α regulates how strongly the explicit G×E component contributes to the final residual prediction. When $\alpha = 1$, the full predicted G×E component is retained; when $0 < \alpha < 1$, its contribution is shrunk; and when $\alpha = 0$, the model reduces to the genomic residual component without the explicit G×E contribution.

The scaling parameter α was selected by five-fold cross-validation within the training set. Because within-environment genotype ranking was the primary objective of the G×E component, the final value was selected according to the average within-environment PCC. The selected value, $\alpha = 0.33$, improved within-environment PCC relative to the unscaled TwoStep_G×E model and slightly outperformed TwoStep_G (0.376 vs. 0.371). These results suggest that reducing the contribution of the explicit G×E component can help limit noise from candidate interaction features while preserving useful interaction signals for environment-specific ranking.

In addition, the strategy used to select candidate interaction SNPs has a direct impact on noise levels. This study employs a phenotypic plasticity GWAS approach to identify candidate interaction markers, which avoids performing association tests for each specific environmental factor and thus substantially reduces computational complexity. However, this strategy comes at the cost that the selected SNPs primarily reflect the genetic basis of cross-environment plasticity rather than direct genotype-by-environment interactions, which may limit interpretability and introduce selection bias, including winner's curse.

4.4. Breeding Value of Environmental Adaptation-Based Selection

From the perspective of practical breeding decisions, we translated model performance into a quantifiable selection gain metric. The results show that, across all models, adaptive selection consistently outperforms overall selection; this gap is particularly pronounced in the proposed framework. Specifically, TwoStep_G×E_alpha0.33 achieves a gain of 0.454 Mg ha^{-1} under adaptive selection, which is 19.8% higher than that under overall selection (0.379 Mg ha^{-1}).

These findings highlight the core breeding value of incorporating G×E information: its primary advantage does not lie in improving average predictive accuracy, but in addressing the central question of multi-environment breeding, namely "which genotype performs best in which environment." In practical breeding programs, resource constraints prevent breeders from evaluating all candidate genotypes across all target environments. The proposed framework requires only environmental covariates from the target environment, without the need for field trials in that environment, enabling the prediction of environment-specific genotype performance. This facilitates early-stage targeted selection and reduces experimental costs.

These observations are consistent with the enviromics-assisted selection framework proposed by Gogna et al. (2026) [26], further supporting the feasibility of systematically integrating environmental omics information into breeding decision-making pipelines.

4.5. Biological Contextualization of G×E Interactions

The biological contextualization presented here should be regarded as an exploratory, hypothesis-generating analysis rather than a direct validation component of the predictive framework. In the G×E network analysis, environmental factors during the grain-filling stage (pStGEnG, Start to End of Grain Fill) exhibited the highest average importance (401.44), which is broadly consistent with the findings of Çakir (2004) [12] that water stress during grain filling can lead to irreversible yield losses. This agreement suggests that the prioritized environmental window is consistent with previously reported agronomic sensitivity during grain filling, although it should not be interpreted as mechanistic validation.

In the genotype–environment interaction tests, the top three statistically significant interactions were all associated with the duration of soil evaporation during the second stage of grain filling (T_pStGEnG), suggesting that soil moisture dynamics during grain filling may represent an important component of G×E structure in this dataset. In addition, the interaction between locus 8:177190067:T:C and soil available water during germination and emergence (ESW_pGerEme_5) reached Bonferroni-corrected significance. This locus is located approximately 196 bp upstream of the *wox13b* gene, a member of the WOX transcription factor family. Given the reported roles of WOX genes in plant responses to abiotic stress [35], this nearby gene annotation provides possible biological context for this statistical interaction signal, although this interpretation remains speculative and requires independent validation.

It is important to note that the biological contextualization presented here remains limited. Although the framework uses plasticity-based GWAS to prioritize candidate SNPs for explicit G×E feature construction, the plasticity metrics summarize genotype-level responses across environments rather than responses to individual environmental covariates. Moreover, the subsequent dimensionality reduction and Kronecker-product construction further transform SNP and environmental information into compressed G×E features. As a result, the predictive model itself does not directly output SNP-specific interaction significance, variance explained, or mechanistic effects for individual environmental covariates. The SNP–environmental covariate tests and gene annotations presented in this section were therefore conducted only as post hoc exploratory analyses to provide biological context for prioritized statistical interaction patterns. Some interaction terms in the SNP–environmental covariate analysis exhibited extremely strong statistical significance (e.g., Bonferroni-adjusted $p < 10^{-100}$). However, these signals were derived from a post hoc screening framework in which both candidate SNPs and representative environmental factors were independently pre-screened, thereby introducing selection bias into the interaction tests. As a result, the reported p-values should not be interpreted as valid controls of the family-wise error rate, nor should they be taken as evidence of causal biological mechanisms. Therefore, the biological contextualization presented here should be viewed as exploratory and hypothesis-generating, and will require validation through independent experiments, expression analyses, physiological measurements, or functional assays.

4.6. Limitations and Future Perspectives

This study has several limitations. First, the test set comprises only 26 environments, resulting in limited statistical power. Therefore, some conclusions regarding the structure of G×E require further validation using larger multi-environment datasets. Second, the proposed framework relies heavily on high-quality and systematically collected environmental covariates. In regions or crop systems with incomplete environmental monitoring infrastructure, the performance of the first-stage environmental mean model may deteriorate substantially. Third, the selection of candidate interaction SNPs based on plasticity GWAS is subject to winner's curse bias. Future studies could consider more conservative selection thresholds or permutation-based procedures to mitigate this issue. Fourth, because the plasticity-based GWAS, dimensionality reduction, and Kronecker-product construction transform SNP and environmental information into compressed G×E features, the resulting biological interpretation should be viewed as post hoc exploratory contextualization rather than direct evidence of SNP-specific mechanisms or functional validation.

Looking forward, with the rapid development of multi-omics technologies (e.g., transcriptomics and metabolomics) and high-throughput phenotyping platforms, integrating gene expression and metabolic information into G×E modeling frameworks may further elucidate the molecular mechanisms underlying environmental responses. In

addition, extending the proposed framework to other crop species—particularly inbred populations, where dominance modeling would need to be adjusted—and evaluating cultivar adaptation under future climate change scenarios represent important directions for future research.

5. Conclusions

This study developed a two-step framework for maize yield prediction that explicitly incorporates G×E information. The results show that the proposed framework improves prediction performance by combining machine-learning-based modeling of environmental main effects with genomic relationship matrices for polygenic small-effect prediction. Furthermore, candidate interaction markers were identified through plasticity-based genome-wide association analysis and used to construct an explicit G×E matrix. Although the direct inclusion of candidate G×E features may introduce additional noise, our results indicate that appropriate G×E scaling can mitigate noise amplification and improve model stability.

From a practical breeding perspective, the proposed framework showed clear value in environment-specific selection, with adaptive selection strategies achieving greater genetic gain than overall selection. These findings suggest that the main advantage of the framework lies not only in improving prediction accuracy, but also in supporting environment-specific genotype ranking and breeding decisions. In addition, post hoc interaction network analysis and gene annotation provided exploratory biological context for prioritized statistical G×E patterns, but these results should be interpreted as hypothesis-generating rather than as functional validation. Overall, the proposed framework provides a useful and scalable computational approach for environment-specific yield prediction and adaptive selection in maize breeding.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture16111233/s1, Table S1: Hyperparameter search results for environment-mean prediction models evaluated by five-fold cross-validation on the training set. Table S2: Complete list of 398 candidate interaction SNPs identified through plasticity-based GWAS across six phenotypic stability metrics, with chromosomal positions and contributing metrics annotated. Table S3: Summary of 22 representative environmental factors selected for genotype–environment interaction testing, including variable descriptions, phenological stage assignments, and LightGBM importance scores. Table S4: Complete results of genotype–environmental covariate interaction tests for all combinations of 398 candidate SNPs and 22 representative environmental factors, including regression coefficients, standard errors, partial R^2 values, nearest genes, and GO annotations. Table S5: Threshold sensitivity analysis evaluating the robustness of candidate SNP selection under different plasticity-GWAS significance thresholds. Table S6: Environment-level five-fold cross-validation results used to evaluate the robustness of GBLUP, TwoStep_G, TwoStep_G×E, and TwoStep_G×E_alpha across G2F environments. Environments were defined as location–year combinations and were split at the environment level rather than at the plot-record level. The table contains four sheets: (i) the selected environment-mean prediction model and its parameters for each fold; (ii) the fold-specific candidate SNPs selected from plasticity GWAS using only the training environments within each fold; (iii) per-fold prediction performance of GBLUP, TwoStep_G, TwoStep_G×E, and TwoStep_G×E_alpha, including global RMSE, global PCC, within-environment RMSE, and within-environment PCC; and (iv) five-fold mean ± SD summaries of all evaluation metrics. TwoStep_G×E denotes the unscaled G×E model with $\alpha = 1$, whereas TwoStep_G×E_alpha denotes the scaled G×E model with fold-specific α selected according to the highest within-environment PCC. Figure S1: Genomic distribution of 398 candidate interaction SNPs across the ten maize chromosomes. Figure S2: Cross-validation performance of the G×E-scaled model across different values of the scaling parameter (α) applied to the G×E feature matrix in the training set. The x-axis represents α; the left y-axis denotes Pearson correlation coefficient (PCC), and the right y-axis denotes root mean square error (RMSE). Although some metrics showed better

values at larger α, $\alpha = 0.33$ was selected because average within-environment PCC was used as the primary tuning criterion and reached its maximum at this value. Figure S3: Exploratory analysis of environment-specific prediction performance. Panel (a) shows the relationship between the observed mean yield of each test environment and within-environment PCC. Panel (b) shows the relationship between the number of hybrids in each test environment that overlapped with the training set and within-environment PCC. Each point represents one test environment, labeled by location and year. Within-environment PCC tended to be higher in environments with higher mean yield, whereas the number of overlapping hybrids between the training and test sets showed no clear association with within-environment PCC. Figure S4: Quantile–quantile (QQ) plots and genomic inflation factor values (λ_{GC}) for the plasticity-based GWAS analyses across the six phenotypic stability metrics.

Author Contributions: Conceptualization, Q.W. and X.L.; methodology, Q.W.; software, Q.W.; validation, X.L., J.Z. and J.L.; formal analysis, Q.W.; investigation, Q.W.; resources, Q.W.; data curation, Q.W.; writing—original draft preparation, Q.W.; writing—review and editing, X.L., J.Z. and J.L.; visualization, Q.W.; supervision, A.Z.; project administration, J.Z. and A.Z.; funding acquisition, A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Central Public-interest Scientific Institution Basal Research Fund (No. Y2026ZZ33), the Agricultural Science and Technology Innovation Program (No. CAAS-ZDRW202503), the Public Welfare Scientific Research Institutes Fundamental Research Fund of Agricultural Information Institute, Chinese Academy of Agricultural Sciences (No. JBYW-AII-2026-15).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in the Genomes to Fields 2022 Maize Genotype by Environment Prediction Competition at https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/GenomesToFields_GenotypeByEnvironment_PredictionCompetition_2023 (accessed on 2 April 2026).

Acknowledgments: The authors would like to express their sincere gratitude to all co-authors and anonymous reviewers for their constructive suggestions and valuable feedback, which have significantly improved the quality and clarity of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CV	Coefficient of Variation
DL	Deep Learning
DNN	Deep Neural Network
EC	Environmental Covariate
FW	Finlay–Wilkinson Regression
G×E	Genotype-by-Environment Interaction
G2F	Genomes to Fields
GBLUP	Genomic Best Linear Unbiased Prediction
GO	Gene Ontology
GWAS	Genome-Wide Association Study
LD	Linkage Disequilibrium
LightGBM	Light Gradient Boosting Machine
MAD	Median Absolute Deviation
ML	Machine Learning
MLP	Multi-Layer Perceptron
PC	Principal Component
PCC	Pearson Correlation Coefficient

QTL	Quantitative Trait Locus
RMSE	Root Mean Square Error
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
TabM	Tabular Model (neural network architecture)
VCF	Variant Call Format
WOX	WUSCHEL Homeobox

References

1. World Health Organization. *Climate Change and Health*; Fact Sheet; World Health Organization: Geneva, Switzerland, 2023.
2. United Nations. *World Population Prospects 2024*; Online Edition; United Nations, Department of Economic and Social Affairs, Population Division: New York, NY, USA, 2024.
3. Lobell, D.B.; Gourdji, S.M. The Influence of Climate Change on Global Crop Productivity. *Plant Physiol.* **2012**, *160*, 1686–1697. [CrossRef]
4. Xu, Y. Envirotyping for deciphering environmental impacts on crop plants. *Theor. Appl. Genet.* **2016**, *129*, 653–673. [CrossRef]
5. Alemu, A.; Åstrand, J.; Montesinos-López, O.A.; Isidro y Sánchez, J.; Fernández-Gónzalez, J.; Tadesse, W.; Vetukuri, R.R.; Carlsson, A.S.; Ceplitis, A.; Crossa, J.; et al. Genomic selection in plant breeding: Key factors shaping two decades of progress. *Mol. Plant* **2024**, *17*, 552–578. [CrossRef]
6. Crossa, J.; Martini, J.W.R.; Vitale, P.; Pérez-Rodríguez, P.; Costa-Neto, G.; Fritsche-Neto, R.; Runcie, D.; Cuevas, J.; Toledo, F.; Li, H.; et al. Expanding genomic prediction in plant breeding: Harnessing big data, machine learning, and advanced software. *Trends Plant Sci.* **2025**, *30*, 756–774. [CrossRef] [PubMed]
7. Montesinos-López, A.; Montesinos-López, O.A.; Ramos-Pulido, S.; Mosqueda-González, B.A.; Guerrero-Arroyo, E.A.; Crossa, J.; Ortiz, R. Artificial intelligence meets genomic selection: Comparing deep learning and GBLUP across diverse plant datasets. *Front. Genet.* **2025**, *16*, 1568705. [CrossRef]
8. Nelimor, C.; Badu-Apraku, B.; Tetteh, A.Y.; N'guetta, A.S.P. Assessment of Genetic Diversity for Drought, Heat and Combined Drought and Heat Stress Tolerance in Early Maturing Maize Landraces. *Plants* **2019**, *8*, 518. [CrossRef]
9. Tolley, S.A.; Brito, L.F.; Wang, D.R.; Tuinstra, M.R. Genomic prediction and association mapping of maize grain yield in multi-environment trials based on reaction norm models. *Front. Genet.* **2023**, *14*, 1221751. [CrossRef]
10. Zhao, Q.; Wang, C.; Wang, X.; Rezaei, E.E.; Müller, C.; Wang, E.; Webber, H.; Zhang, L.; Li, X.; Sang, Y.; et al. Temperature thresholds of extreme heat-induced yield loss in maize and soybean reveal geographic heterogeneity across the Northern Hemisphere. *Nat. Food* **2026**, *7*, 194–205. [CrossRef] [PubMed]
11. Niu, S.; Yu, L.; Li, J.; Qu, L.; Wang, Z.; Li, G.; Guo, J.; Lu, D. Effect of high temperature on maize yield and grain components: A meta-analysis. *Sci. Total Environ.* **2024**, *952*, 175898. [CrossRef]
12. Çakir, R. Effect of water stress at different development stages on vegetative and reproductive growth of corn. *Field Crops Res.* **2004**, *89*, 1–16. [CrossRef]
13. Li, X.; Zhao, X.; Sun, S.; He, M.; Wang, J.; Xiang, X.; Niu, Y. Genetic Architecture and Meta-QTL Identification of Yield Traits in Maize (*Zea mays* L.). *Plants* **2025**, *14*, 3067. [CrossRef]
14. Li, D.; Zhou, Z.; Lu, X.; Jiang, Y.; Li, G.; Li, J.; Wang, H.; Chen, S.; Li, X.; Würschum, T.; et al. Genetic Dissection of Hybrid Performance and Heterosis for Yield-Related Traits in Maize. *Front. Plant Sci.* **2021**, *12*, 774478. [CrossRef]
15. Bheemanahalli, R.; Ramamoorthy, P.; Poudel, S.; Samiappan, S.; Wijewardane, N.; Reddy, K.R. Effects of drought and heat stresses during reproductive stage on pollen germination, yield, and leaf reflectance properties in maize (*Zea mays* L.). *Plant Direct* **2022**, *6*, e434. [CrossRef] [PubMed]
16. Barreto, C.A.V.; Das Graças Dias, K.O.; De Sousa, I.C.; Azevedo, C.F.; Nascimento, A.C.C.; Guimarães, L.J.M.; Guimarães, C.T.; Pastina, M.M.; Nascimento, M. Genomic prediction in multi-environment trials in maize using statistical and machine learning methods. *Sci. Rep.* **2024**, *14*, 1062. [CrossRef] [PubMed]
17. Zhang, Q.X.; Zhu, T.; Lin, F.; Fang, D.; Chen, X.; Lou, X.; Tong, Z.; Xiao, B.; Xu, H.M. mmGEBLUP: An advanced genomic prediction scheme for genetic improvement of complex traits in crops through integrative analysis of major genes, polygenes, and genotype–environment interactions. *Brief. Bioinform.* **2024**, *26*, bbaf058. [CrossRef]
18. Westhues, C.C.; Mahone, G.S.; Da Silva, S.; Thorwarth, P.; Schmidt, M.; Richter, J.C.; Simianer, H.; Beissinger, T.M. Prediction of Maize Phenotypic Traits with Genomic and Environmental Predictors Using Gradient Boosting Frameworks. *Front. Plant Sci.* **2021**, *12*, 699589. [CrossRef] [PubMed]
19. Kelly, C.M.; McLaughlin, R.L. Comparison of machine learning methods for genomic prediction of selected *Arabidopsis thaliana* traits. *PLoS ONE* **2024**, *19*, e0308962. [CrossRef]

20. Rogers, A.R.; Dunne, J.C.; Romay, C.; Bohn, M.; Buckler, E.S.; Ciampitti, I.A.; Edwards, J.; Ertl, D.; Flint-Garcia, S.; Gore, M.A.; et al. The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. *G3 Genes | Genomes | Genet.* **2021**, *11*, jkaa050. [CrossRef]
21. Fernandes, I.K.; Vieira, C.C.; Dias, K.O.G.; Fernandes, S.B. Using machine learning to combine genetic and environmental data for maize grain yield predictions across multi-environment trials. *Theor. Appl. Genet.* **2024**, *137*, 189. [CrossRef]
22. Liu, N.; Du, Y.; Warburton, M.L.; Xiao, Y.; Yan, J. Phenotypic Plasticity Contributes to Maize Adaptation and Heterosis. *Mol. Biol. Evol.* **2021**, *38*, 1262–1275. [CrossRef]
23. Fu, R.; Wang, X. Modeling the influence of phenotypic plasticity on maize hybrid performance. *Plant Commun.* **2023**, *4*, 100548. [CrossRef]
24. Lima, D.C.; Washburn, J.D.; Varela, J.I.; Chen, Q.; Gage, J.L.; Romay, M.C.; Holland, J.; Ertl, D.; Lopez-Cruz, M.; Aguate, F.M.; et al. Genomes to Fields 2022 Maize Genotype by Environment Prediction Competition. *BMC Res. Notes* **2023**, *16*, 148. [CrossRef]
25. Washburn, J.D.; Varela, J.I.; Xavier, A.; Chen, Q.; Ertl, D.; Gage, J.L.; Holland, J.B.; Lima, D.C.; Romay, M.C.; Lopez-Cruz, M.; et al. Global genotype by environment prediction competition reveals that diverse modeling strategies can deliver satisfactory maize yield estimates. *Genetics* **2025**, *229*, iyae195. [CrossRef]
26. Gogna, A.; Kamali, B.; Wimmer, V.; Schmidt, R.H.; Rezaei, E.E.; Eckhoff, W.M.; Reif, J.C.; Zhao, Y. Predicting enviromically adapted varieties with big data. *Genome Biol.* **2026**, *27*, 3. [CrossRef]
27. Xavier, A. Lectures/MGC_2023: CLAC Model Source Code for the G2F Maize G×E Prediction Competition. Available online: https://github.com/alenxav/Lectures/tree/master/MGC_2023 (accessed on 2 April 2026).
28. Xavier, A.; Muir, W.M.; Rainey, K.M. bWGR: Bayesian whole-genome regression. *Bioinformatics* **2020**, *36*, 1957–1959. [CrossRef] [PubMed]
29. Gorishniy, Y.; Kotelnikov, A.; Babenko, A. TabM: Advancing Tabular Deep Learning with Parameter-Efficient Ensembling. *arXiv* **2025**, arXiv:2410.24210. [CrossRef]
30. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [CrossRef]
31. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **2015**, *4*, s13742-015-0047-8. [CrossRef] [PubMed]
32. Korte, A.; Vilhjálmsson, B.J.; Segura, V.; Platt, A.; Long, Q.; Nordborg, M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* **2012**, *44*, 1066–1071. [CrossRef]
33. Cuevas, J.; Crossa, J.; Montesinos-López, O.A.; Burgueño, J.; Pérez-Rodríguez, P.; de los Campos, G. Bayesian Genomic Prediction with Genotype × Environment Interaction Kernel Models. *G3 Genes | Genomes | Genet.* **2017**, *7*, 41–53. [CrossRef]
34. Hu, X.; Carver, B.F.; El-Kassaby, Y.A.; Zhu, L.; Chen, C. Weighted kernels improve multi-environment genomic prediction. *Heredity* **2023**, *130*, 82–91. [CrossRef]
35. Chen, X.; Hou, Y.; Cao, Y.; Wei, B.; Gu, L. A Comprehensive Identification and Expression Analysis of the WUSCHEL Homeobox-Containing Protein Family Reveals Their Special Role in Development and Abiotic Stress Response in *Zea mays* L. *Int. J. Mol. Sci.* **2023**, *25*, 441. [CrossRef] [PubMed]
36. Lopez-Cruz, M.; de Los Campos, G. Multi-trait/environment sparse genomic prediction using the SFSI R-package. *Plant Genome* **2025**, *18*, e70050. [CrossRef]
37. Hu, H.; Rincent, R.; Runcie, D.E. MegaLMM improves genomic predictions in new environments using environmental covariates. *Genetics* **2025**, *229*, iyae171. [CrossRef] [PubMed]
38. Li, R.; Zhang, D.; Han, Y.; Liu, Z.; Zhang, Q.; Zhang, Q.; Wang, X.; Pan, S.; Sun, J.; Wang, K. Hybrid Deep Learning Approaches for Improved Genomic Prediction in Crop Breeding. *Agriculture* **2025**, *15*, 1171. [CrossRef]
39. Jarquín, D.; Crossa, J.; Lacaze, X.; Du Cheyron, P.; Daucourt, J.; Lorgeou, J.; Piraux, F.; Guerreiro, L.; Pérez, P.; Calus, M.; et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* **2014**, *127*, 595–607. [CrossRef]