

Article

Effects of Phenotypic Adjustment Using Genetic and Spatial Models on Genomic Prediction and Top-N Recall in Early-Stage Maize Hybrid Breeding

Hao Liu ^{1,2}, Weijie Pan ^{1,2}, Shuqin Jiang ³, Xiaoguang Zhai ⁴, Tao Yu ¹, Gengbin Yang ¹, Wenyue Li ¹, Haochen Wang ¹, Yanjie Sun ⁵, Yuheng Wu ⁵, Zhonglin Niu ⁵ and Shiliang Cao ^{1,*}

¹ Maize Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China; anonymous401@163.com (H.L.); panweijie1998@163.com (W.P.); yutaoweiwei@163.com (T.Y.); kshmaize@163.com (G.Y.); liwenyue199211@163.com (W.L.); whcnky@163.com (H.W.)

² State Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University, Taian 271018, China

³ Qingdao KingAgroot Seed Science Co., Ltd., Beijing Biotechnology Branch, Beijing 100071, China; wanshi0066@126.com

⁴ Molbreeding Biotechnology Co., Ltd., Shijiazhuang 050000, China; zhaixiaoguang@molbreeding.com

⁵ Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China; sunyanjie1980@163.com (Y.S.); wuyuheng1997yw@163.com (Y.W.); niuzhonglin_163@163.com (Z.N.)

* Correspondence: caoshiliang@haas.cn

Abstract

Single-replicate field trials are common in early-stage maize hybrid screening, but raw plot observations contain field noise that can weaken genomic prediction. We evaluated whether pedigree and spatial adjustment improve the predictability of phenotypic targets and Top-N recovery, whether joint pedigree-spatial adjustment provides complementary gains, and whether target construction has a larger influence than baseline model choice. Across four sites, five agronomic traits, four phenotypic inputs, and six genomic prediction models, pedigree-corrected phenotypes (Ped) increased mean predictive ability from 0.465 ± 0.157 for raw observations (Raw) to 0.573 ± 0.131 and increased upper-tail Top-20 recall from 0.558 ± 0.146 to 0.656 ± 0.146 under a 20% predicted retention window. Spatial correction alone (Spa) was less predictable than Raw, whereas joint pedigree-spatial correction (Ped+Spa) improved over Raw but did not exceed Ped. Phenotypic input explained substantially more variation in predictive ability than model choice, whose main effect was not significant ($F_{5428} = 1.59$, $p = 0.161$). These results support pedigree correction as a useful target-construction strategy in the evaluated single-replicate trials while showing that spatial adjustment is conditional on model adequacy and trial structure. Top-N recall directly represents favorable direction recovery for grain yield; for grain moisture, plant height, ear height, and silking date, the archived upper-tail metric is interpreted as rank consistency because the favorable direction or ideotype is breeding-program-dependent.

Keywords: maize hybrids; genomic selection; phenotypic adjustment; Top-N candidate recall



Academic Editor: HongWei Cai

Received: 26 June 2026

Revised: 22 July 2026

Accepted: 25 July 2026

Published: 28 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Genomic selection (GS) can shorten maize breeding cycles by predicting candidate performance from genome-wide markers before all hybrids undergo extensive multi-environment evaluation [1–3]. Its operational value is greatest in early screening, where many experimental crosses must be reduced to a manageable set under limited seed,

land, and labor. Unreplicated trials are therefore common, but their plot-level observations combine genetic differences with soil gradients, management heterogeneity, stand variation, and measurement error [4]. A recent study of 1436 unreplicated maize testcrosses further showed that prediction depended strongly on training-set composition across environments and years, emphasizing that data design and target quality constrain genomic prediction as much as the algorithm itself [5].

Most genomic prediction optimization has focused on marker representation, training population design, genotype-by-environment modeling, or prediction algorithms [6–10]. In practice, however, predictive ability is usually calculated against an observed or statistically adjusted phenotype because true breeding values are unavailable [11]. The target is therefore part of the model definition: replacing a raw observation with a pedigree-adjusted, spatially adjusted, or jointly adjusted value changes the signal that is learned and evaluated. Recent maize multi-environment analyses likewise show that heritability, population history, and environmental structure alter attainable prediction performance [12,13].

Pedigree and spatial adjustment address different sources of uncertainty. Pedigree BLUP borrows information among related hybrids and can stabilize entry estimates when replication is limited [14,15]. Spatial models instead aim to separate local field heterogeneity from entry performance [16–18]. Their benefit is not automatic. In single-replicate trials, entry, family, and spatial effects may be difficult to identify separately, and an overly rigid or overly smooth surface may retain an environmental trend or attenuate a genetic signal. A 2024 hybrid-maize study showed that spatiotemporal information derived from longitudinal imagery could improve model fit and genotypic effect estimation, but it also demonstrated that the usefulness of spatial information depended on how the field process was represented [19].

Raw, pedigree-corrected (Ped), spatially corrected (Spa), and jointly pedigree-spatial-corrected (Ped+Spa) phenotypes were therefore compared as distinct prediction targets. Raw provides the operational single-plot baseline. Ped tests whether relatedness can recover a stable genetic signal. Spa tests whether the row-column field structure can be removed without an excessive loss of entry variation. Ped+Spa tests whether pedigree and field information contribute complementary improvement. This comparison addresses a gap not resolved by studies that change the prediction model while holding the phenotypic target fixed.

Pearson correlation summarizes agreement across the complete test set, whereas breeding decisions act on a selected tail of the candidate distribution [20–22]. Top-N recall asks whether a predicted retention set contains the candidates ranked within the observed top N and therefore complements average predictive ability [23,24]. Here, a 20% predicted retention window was used. High values are directly favorable for grain yield (GY). A lower grain moisture content (GMC), earlier silking date (SD), and, in some target environments, lower plant height (PH) or ear height (EH) may be preferred; nevertheless, optimal directions or ranges depend on the maturity group, lodging risk, harvest system, and breeding ideotype. Because the archived analysis ranked all traits in descending order, non-GY Top-N values are reported as upper-tail rank consistency rather than universal breeding-success measures.

Accordingly, we expected pedigree correction to improve the predictability of phenotypic targets and the recovery of top-ranked candidates relative to raw plot observations. Spatial correction was expected to provide additional gains when row-column field heterogeneity was adequately captured, while joint pedigree-spatial correction was expected to benefit from the complementary information contained in relationship and field-position data. We further expected phenotypic target construction to have a larger influence on prediction performance than the choice among commonly used baseline genomic predic-

tion models, and we expected improvements in predictive ability to be accompanied by improved Top-N recovery. These expectations were evaluated across four maize hybrid trial sites, five traits, four phenotypic inputs, six prediction models, and matched random data partitions.

2. Materials and Methods

2.1. Plant Materials and Field Design

Field trials of maize hybrids were conducted in 2025 at four testing sites in Heilongjiang Province, China: Boli (45.75279° N, 130.57211° E), Hongxinglong (46.72344° N, 131.58790° E), Harbin (45.76009° N, 126.61089° E), and Suihua (46.63954° N, 126.99508° E). Plot-level field information included row length (5.0 m), row spacing (65 cm), plant spacing (25 cm), and two rows per plot. Each hybrid was grown as a single plot without replication at each testing site, and field row and column coordinates were recorded for spatial modeling. A total of four hybrid evaluation populations were tested across the four environments. The Boli and Hongxinglong trials evaluated the same set of 465 maize hybrids, allowing for a direct comparison of the environmental effects between the two locations. The Harbin and Suihua trials consisted of 847 and 662 hybrids, respectively, representing separate hybrid sets evaluated independently at these locations. Plot-level metadata included hybrid identity, parental information, and field row and column coordinates. After genotype–phenotype matching, the effective sample sizes were 465, 465, 847, and 662 hybrids for Boli, Hongxinglong, Harbin, and Suihua, respectively (Table 1). The evaluated traits included grain yield (GY), grain moisture content (GMC), plant height (PH), ear height (EH), and silking date (SD). Grain yield was measured from harvested ears in each plot and adjusted to standard moisture content. Grain moisture content was determined at harvest using a grain moisture analyzer (PM-8188-A, Kett Electric Laboratory Co., Ltd., Tokyo, Japan). Plant height was measured from the soil surface to the tassel tip, and ear height was measured from the soil surface to the node bearing the uppermost ear. The silking date was recorded as the number of days from sowing to the date when approximately 50% of plants in a plot showed silk emergence.

Table 1. Phenotypic inputs, analyzed traits, and effective sample sizes after genotype–phenotype matching.

Trial Site	Phenotypic Inputs	Traits	Samples	Training Set	Test Set
Boli	Raw; Ped; Spa; Ped+Spa	GY, GMC, PH, EH, SD	465	140	325
Hongxinglong	Raw; Ped; Spa; Ped+Spa	GY, GMC, PH, EH, SD	465	140	325
Harbin	Raw; Ped; Spa; Ped+Spa	GY, GMC, PH, EH, SD	847	255	592
Suihua	Raw; Ped; Spa; Ped+Spa	GY, GMC, PH, EH, SD	662	199	463

Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes; Spa, spatially corrected phenotypes using AR(1) $\times$ AR(1), except documented fallbacks; Ped+Spa, jointly pedigree- and spatially corrected phenotypes. GY, grain yield; GMC, grain moisture content; PH, plant height; EH, ear height; SD, silking date.

2.2. Genotype Data Processing

The four testing sites were represented by three genotype datasets: Boli and Hongxinglong shared the same genotype dataset, Harbin used an independent dataset, and Suihua used an independent 10K marker dataset. Parental lines of the maize hybrids were genotyped using a liquid-phase SNP chip. Hybrid genotypes were then inferred from parental genotypes using the “Create Hybrid Genotypes” function in TASSEL 5 [25], which reconstructs the expected diploid genotype of each hybrid at every marker locus from

its known parental base pair. Genotype quality control was performed with SNPReady (0.9.7). IUPAC ambiguous codes were converted to standard biallelic genotypes; SNPs with call rate below 80% or minor allele frequency below 0.05 were removed; samples with a genotype call rate below 80% were excluded; and remaining missing genotypes were imputed using the Wright method. After quality control, the Boli/Hongxinglong dataset contained 465 samples and 3053 SNPs, the Harbin dataset contained 847 samples and 3069 SNPs, and the Suihua dataset contained 662 samples and 10,939 SNPs.

2.3. Phenotypic Correction Strategies

Four phenotypic inputs were generated separately for each location and trait. The strategies were chosen to correspond to the information available in early-stage trials and to the planned comparisons. Raw was the unadjusted single-plot observation and served as the operational baseline. Ped used parental pedigree to borrow information among related hybrids and stabilize the target under limited replication. Spa used recorded row and column coordinates to model field heterogeneity. Ped+Spa included both relationship and spatial components to test whether they provided complementary information. $AR(1) \times AR(1)$ was selected as the primary spatial structure because plots were arranged on a regular row–column grid and correlation was expected to decay with field distance. A two-dimensional generalized additive model (GAM) was used as an alternative spatial trend model when the autoregressive formulation was unsuitable. These methods span no adjustment, genetic adjustment, environmental adjustment, and joint adjustment while retaining a common downstream genomic-prediction design.

$$y = 1\mu + u + e, u \sim N(0, A\sigma_u^2), e \sim N(0, I\sigma_e^2) \quad (1)$$

For Ped, y is the vector of phenotypic observations, 1 is a vector of ones, μ is the overall intercept, u is the additive pedigree-based random effect of hybrid entry, e is the residual, A is the additive pedigree relationship matrix constructed from parental information, and σ_u^2 and σ_e^2 are the additive and residual variance components. The corrected phenotype was defined as the intercept plus the predicted additive pedigree value.

$$y = 1\mu + g + s + e, g \sim N(0, I\sigma_g^2), s \sim N(0, \sigma_s^2(R_c \otimes R_r)), e \sim N(0, I\sigma_e^2) \quad (2)$$

For Spa, g represents the random hybrid effect and follows $g \sim N(0, I\sigma_g^2)$. The spatial field effect s was modeled using a first-order autoregressive $\times$ first-order autoregressive [$AR(1) \times AR(1)$] covariance structure, where $s \sim N(0, \sigma_s^2(R_c \otimes R_r))$. Here, R_r and R_c denote the $AR(1)$ correlation matrices along field rows and columns, respectively. The residual term was assumed to follow $e \sim N(0, I\sigma_e^2)$.

$$y = 1\mu + u + s + e, u \sim N(0, A\sigma_u^2), s \sim N(0, \sigma_s^2(R_c \otimes R_r)), e \sim N(0, I\sigma_e^2) \quad (3)$$

For Ped+Spa, u represents the additive pedigree-based genetic effect and was modeled as $u \sim N(0, A\sigma_u^2)$, where A is the additive pedigree relationship matrix. The spatial component was modeled as $s \sim N(0, \sigma_s^2(R_c \otimes R_r))$ and the residual error was assumed as $e \sim N(0, I\sigma_e^2)$.

2.4. Genomic Prediction Models

Six fixed-configuration prediction models were compared: ridge-regression BLUP (rrBLUP), genomic BLUP (GBLUP), additive-dominance GBLUP (AD-GBLUP), random forest (RF), support vector regression (SVR), and reproducing kernel Hilbert space (RKHS) regression. Genotypes were represented as $-1/0/1$ for additive analyses; missing values,

if present, were imputed with the EM option of rrBLUP::A.mat. For GBLUP, the additive relationship matrix was $K_A = \frac{ZZ'}{2\sum_j p_j(1-p_j)}$, where Z is the allele-frequency-centered marker matrix. rrBLUP fitted marker effects directly with mixed.solve(y, Z = M) [26], whereas GBLUP fitted the dual form with mixed.solve(y, K = K_A) [27]. For dominance, 0/1/2 genotypes were transformed as $D_{ij} = 1 - |M_{ij} - 1|$, the columns were centered, and $K_D = \frac{D_c D_c'}{p}$. AD-GBLUP fitted K_A and K_D jointly in BGLR with 1500 iterations and 500 burn-in iterations [28]. For RKHS, centered marker rows were used to calculate squared Euclidean distances d²; the bandwidth b was the median positive pairwise d², and $K_{\text{RKHS}}(i, i') = \exp\left[-\frac{d^2(i, i')}{b}\right]$ [29]. RF used 500 trees, mtry = $\lfloor \sqrt{p} \rfloor$, and terminal node size 5 [30]. SVR used epsilon-regression with an RBF kernel, C = 1, $\epsilon = 0.1$, $\gamma = \frac{1}{p}$, and predictor scaling [31]. These were common benchmark configurations: no hyperparameter optimization or sensitivity analysis was performed, and no test-set information was used to select settings. Accordingly, model comparisons apply only to these implementations and do not represent each algorithm under optimal tuning. Analyses used R 4.5.2 with rrBLUP 4.6.3, BGLR 1.1.4, random Forest 4.7.1.2, and e1071 1.7.17.

2.5. Data Partitioning and Evaluation Metrics

For each location–trait–input–model combination, 30% of candidates were assigned to the training set and 70% were assigned to the test set. The large test fraction reflects the early-stage use case in which a relatively small evaluated set supports prediction for a larger candidate pool [32]. Fifty random partitions were generated and the same hybrid assignments to training and test sets were maintained across all phenotypic inputs and prediction models within each location–trait combination. This matched partition strategy ensured that predictive performance differences reflected phenotypic correction or model effects rather than variation in candidate sampling. Predictive ability was the Pearson correlation between genomic predictions and the corresponding Raw, Ped, Spa, or Ped+Spa target in the test set. Because these targets are observed or statistically adjusted phenotypes rather than true breeding values, the metric is termed predictive ability (PA), not prediction accuracy relative to true genetic value.

$$\text{Accuracy} = \text{corr}(\hat{y}_{\text{test}}, y_{\text{target, test}}) \quad (4)$$

Within each test set, candidates were sorted in descending order of genomic predicted value. The predicted retention set contained the first floor ($0.20 \times n_{\text{test}}$) candidates. Observed reference sets contained the top 10 or top 20 candidates ranked by the same target, and recall was the number of reference candidates recovered in the predicted retention set divided by 10 or 20, respectively. The expected recall under random selection is approximately 0.20. Because recall is normalized by N, it can be compared across sites with different test-set sizes. For GY, descending order corresponds to the favorable breeding direction. For GMC, PH, EH, and SD, the archived descending-order results quantify upper-tail rank consistency; they are not interpreted as the recovery of a universally favorable ideotype because lower values or intermediate targets may be preferred in a specific breeding program.

2.6. Statistical Analysis

Performance values from the 50 random splits were averaged within each location–trait–input–model combination before inferential summaries, so repeated splits were not treated as independent biological replicates. Exploratory ANOVA decomposed predictive ability (PA) into phenotypic input, location, trait, model, and selected input interaction effects. Overall paired comparisons used matched location–trait–model combination means.

Site-specific Ped-minus-Raw inference used the 30 paired trait $\times$ model means at each site, and trait-specific inference used the 24 paired location $\times$ model means for each trait. Two-sided paired t -tests and t -based 95% confidence intervals were calculated; site- and trait-specific p values were Holm-adjusted within each metric family.

Target-specific genomic variance ratios (h_g^2) were estimated separately for each phenotypic target using a GBLUP variance component model fitted via restricted maximum likelihood (REML): $y = 1\mu + u + e$, $u \sim N(0, G\sigma_g^2)$ and $e \sim N(0, I\sigma_e^2)$, where $u \sim N(0, G\sigma_g^2)$ represents the marker-based additive genomic effect derived from the genomic relationship matrix G , and $e \sim N(0, I\sigma_e^2)$ represents the residual effect. The genomic variance ratio was calculated as $h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$, where σ_g^2 and σ_e^2 denote the marker-based additive genomic variance and residual variance components, respectively. Model-based uncertainty for each of the 80 site–trait–input estimates was obtained after refitting the same REML model in sommer 4.4.4. Delta-method standard errors used the full estimated 2×2 covariance matrix of the genomic and residual variance components; 95% confidence intervals were calculated on the logit(h^2g) scale and back-transformed to $[0, 1]$. These estimate-specific intervals are reported in Supplementary Table S1. The four refitted site estimates within each trait–input combination were additionally summarized by their mean and a t -based 95% confidence interval, $\text{mean} \pm t_{0.975, 3(s/\sqrt{4})}$, to describe between-site variability.

2.7. GAM-Based Spatial Trend Correction

To examine whether the limited performance of the AR1 $\times$ AR1 correction was related to model rigidity or the misspecification of spatial field variation, we generated an alternative spatially adjusted phenotype using a two-dimensional generalized additive model (GAM) [33–35].

The spatial surface was fitted using a tensor-product smooth implemented in the mgcv package in R [36–38]. The model was specified as $y_i = \mu + f(r_i, c_i) + e_i$, where $f(r_i, c_i)$ was modeled using a tensor-product smooth: $\text{gam}(\text{trait} \sim \text{te}(\text{row}, \text{column}, k = c(k_x, k_y)), \text{method} = \text{"REML"})$ with thin-plate regression spline bases ($bs = c(\text{"tp"}, \text{"tp"})$) for row and column dimensions. The basis dimensions were determined according to the available field row and column levels for each site. Smoothing parameters were estimated by REML. If the tensor-product GAM failed to converge, an additive GAM with separate row and column smooth terms was fitted; when both models failed, the raw phenotype was retained as a documented fallback. The analysis used mgcv 1.9-4 with Gaussian errors and an identity link: $y \sim \text{te}(X, Y, k = c(k_x, k_y), bs = c(\text{"tp"}, \text{"tp"}))$. Basis dimensions were selected a priori as $k_x = \max[3, \min(6, nX - 1)]$ and $k_y = \max[5, \min(12, nY - 1)]$, giving $k = (5, 12)$ at Boli and $k = (6, 12)$ at Hongxinglong, Harbin, and Suihua. The adjusted target was $y - [\text{fitted spatial trend} - \text{mean}(\text{fitted spatial trend})]$. Combination-level basis dimensions, effective degrees of freedom, deviance explained, adjusted R^2 , REML score, sample size, raw-adjusted correlation, and fallback status are reported in Supplementary Table S2. The GAM and AR(1) $\times$ AR(1) comparison used identical candidates, missing-data handling, phenotype direction, six genomic models, 50 training/test indices, and evaluation criteria; only spatial target construction differed, producing 6000 matched split records and 120 location $\times$ trait $\times$ model means.

3. Results

3.1. Data Coverage and Analysis Scale

The analysis included four trial sites, five traits, four phenotypic inputs, six prediction models, and 50 repeated random splits. This produced 24,000 predictive ability records and 48,000 Top-N records. After averaging repeated splits within each location–trait–

input–model combination, 480 combination means were available for predictive ability and 960 combination means were available for Top-N metrics across the 20% retention windows.

3.2. Pedigree Correction Is the Strongest Phenotypic Input

As expected, pedigree correction improved both evaluation metrics. Across 120 location–trait–model combinations, mean predictive ability increased from 0.465 ± 0.157 under Raw to 0.573 ± 0.131 under Ped, and upper-tail Top-20 recall increased from 0.558 ± 0.146 to 0.656 ± 0.146 (Table 2; Figure 1). The matched Ped-minus-Raw differences were 0.108 for predictive ability ($t_{119} = 8.37$, $p = 1.26 \times 10^{-13}$) and 0.098 for Top-20 recall ($t_{119} = 9.95$, $p = 2.41 \times 10^{-17}$). Ped+Spa also exceeded Raw (predictive ability 0.549 ± 0.147 ; Top-20 recall 0.620 ± 0.171), whereas Spa alone was the lowest (0.313 ± 0.162 ; 0.432 ± 0.158).

Table 2. Overall performance of four phenotypic inputs across all four locations. Values are means $\pm$ SD across location–trait–model combination means.

Input	Units	PA	Top-10 Hits	Top-20 Hits
Raw	120	0.465 ± 0.157	5.882 ± 1.602 (0.588 ± 0.160)	11.157 ± 2.914 (0.558 ± 0.146)
Ped	120	0.573 ± 0.131	6.924 ± 1.492 (0.692 ± 0.149)	13.114 ± 2.921 (0.656 ± 0.146)
Spa	120	0.313 ± 0.162	4.514 ± 1.673 (0.451 ± 0.167)	8.632 ± 3.152 (0.432 ± 0.158)
Ped+Spa	120	0.549 ± 0.147	6.434 ± 1.809 (0.643 ± 0.181)	12.395 ± 3.420 (0.620 ± 0.171)

Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes; Spa, spatially corrected phenotypes; Ped+Spa, jointly pedigree- and spatially corrected phenotypes. PA: Predictive ability is Pearson’s r between genomic predictions and the corresponding phenotypic target. Top-10 and Top-20 values are mean hit counts, with recall in parentheses, under a 20% predicted retention window. Dispersion is reported as the standard deviation across location–trait–model combination means.

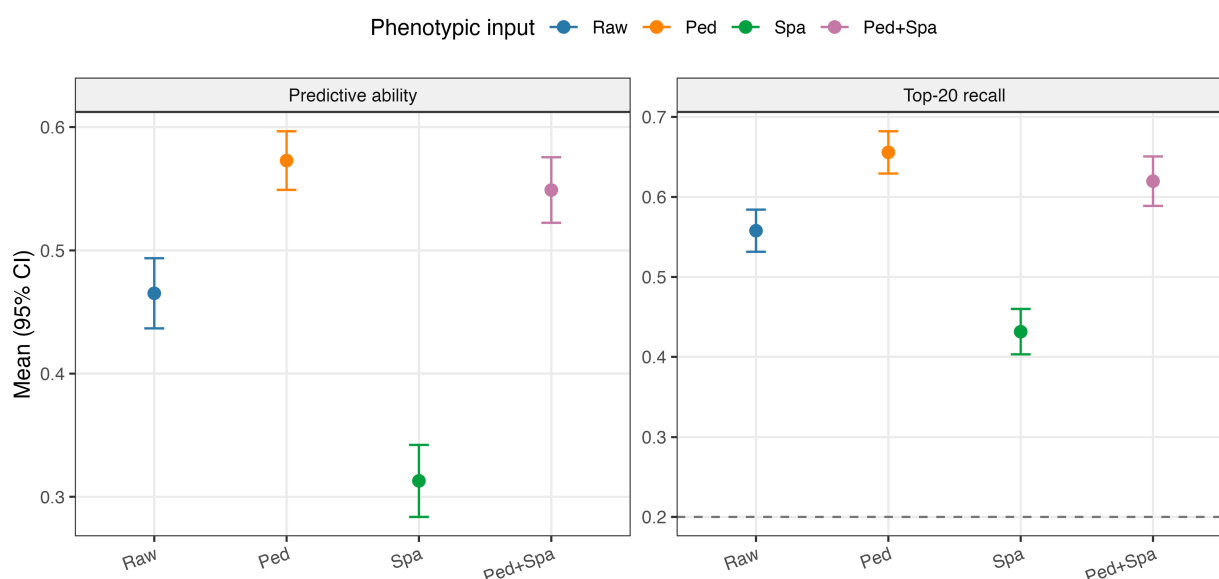


Figure 1. Overall predictive ability (PA) and Top-20 recall for the four phenotypic inputs. Points are means and error bars are t-based 95% confidence intervals across 120 location–trait–model combination means after the 50 random splits were averaged within each combination. The dashed line indicates random Top-20 recall (0.20). Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes; Spa, spatially corrected phenotypes; Ped+Spa, jointly pedigree- and spatially corrected phenotypes.

3.3. The Benefit of Pedigree Correction Varies Among Trial Sites

Ped increased PA at Boli by 0.065 (95% CI 0.059–0.071), Hongxinglong by 0.095 (0.057–0.133), Harbin by 0.020 (0.008–0.032), and Suihua by 0.250 (0.176–0.324), all Holm-adjusted $p \leq 0.0019$ (Table 3; Figure 2). Corresponding Top-20 recall gains were 0.051 (0.045–0.057), 0.081 (0.042–0.119), 0.056 (0.041–0.072), and 0.204 (0.152–0.255), respectively, all Holm-adjusted $p < 0.001$. These contrasts coincide with differences in sample size, marker density, hybrid composition, pedigree connectivity, and field behavior, and the current design cannot isolate any single contributor causally. In particular, Suihua had the highest marker density and an independent hybrid set, but those features were not experimentally separated from location.

Table 3. Ped-over-Raw gains by trial site under the 20% Top-N prediction window.

Site	Raw PA	Ped PA	Δ PA	Δ PA 95% CI	PA Holm P	Raw Top-20	Ped Top-20	Δ Recall	Δ Recall 95% CI
Boli	0.406	0.471	+0.065	0.059–0.071	<0.001	0.508	0.559	+0.051	0.045–0.057
Hongxinglong	0.423	0.518	+0.095	0.057–0.133	<0.001	0.478	0.558	+0.081	0.042–0.119
Harbin	0.549	0.569	+0.020	0.008–0.032	0.0019	0.635	0.692	+0.056	0.041–0.072
Suihua	0.483	0.733	+0.250	0.176–0.324	<0.001	0.610	0.814	+0.204	0.152–0.255

Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes. PA is Pearson's r with the corresponding target. Differences are Ped minus Raw. Site-specific inference used 30 paired trait $\times$ model combination means after averaging the 50 splits. Two-sided paired t -tests were used; p values were Holm-adjusted across four sites within each metric. All Top-20 comparisons had Holm-adjusted $p < 0.001$.

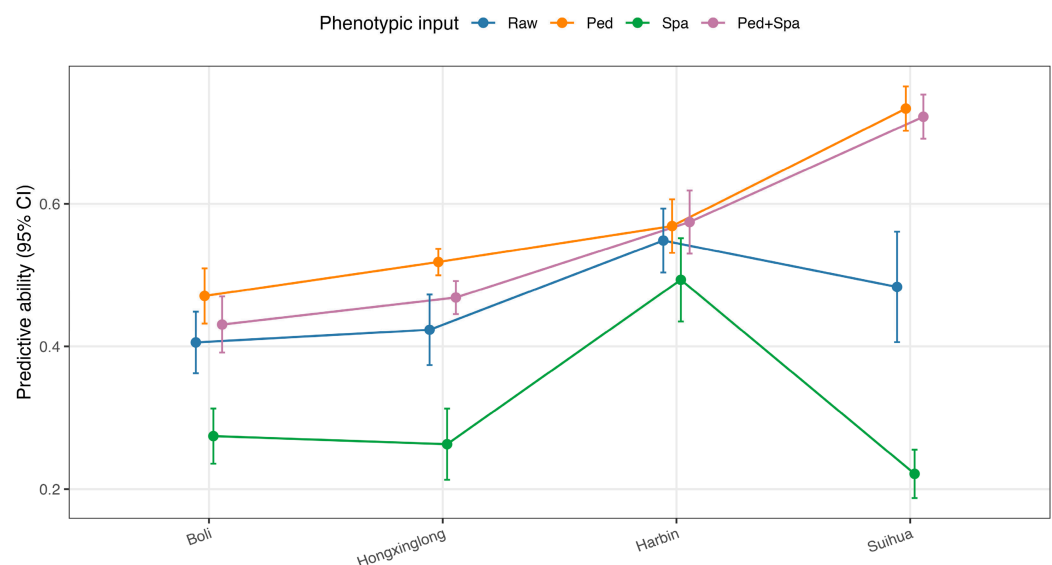


Figure 2. PA by trial site and phenotypic input. Points are means and error bars are t -based 95% confidence intervals across 30 trait–model combination means at each site; the 50 random splits were averaged within each combination. Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes; Spa, spatially corrected phenotypes; Ped+Spa, jointly pedigree- and spatially corrected phenotypes.

3.4. Pedigree-Correction Gains Are Strongest for Grain Yield and Ear Height

The largest Ped-minus-Raw predictive-ability gain occurred for GY (+0.193), followed by EH (+0.184; Table 4 and Figure 3). GY also showed the largest Top-20 recall gain (+0.195), providing direct evidence that Ped improved the recovery of high-yield candidates. For GMC, PH, EH, and SD, the reported Top-N statistics describe upper-tail rank consistency. They should be converted to a breeding-direction or target-range metric before being used to recommend lower moisture, earlier flowering, reduced stature, or another trait-specific ideotype.

Table 4. Ped-over-Raw gains by trait under the 20% Top-N prediction window.

Trait	Raw PA	Ped PA	Δ PA	Δ PA 95% CI	Raw Top-20	Ped Top-20	Δ Recall	Δ Recall 95% CI
GY	0.290	0.483	+0.193	0.145–0.241	0.389	0.584	+0.195	0.142–0.248
GMC	0.543	0.624	+0.080	0.045–0.115	0.686	0.745	+0.058	0.032–0.084
PH	0.531	0.579	+0.048	0.034–0.062	0.609	0.659	+0.050	0.041–0.059
EH	0.377	0.561	+0.184	0.083–0.285	0.465	0.602	+0.138	0.077–0.198
SD	0.585	0.618	+0.033	0.021–0.044	0.640	0.689	+0.048	0.034–0.063

Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes. Differences are Ped minus Raw. Trait-specific inference used 24 paired location $\times$ model combination means after averaging the 50 splits. All two-sided paired comparisons had Holm-adjusted $p < 0.001$; complete tests are reported in Supplementary Table S3. Top-20 is favorable-direction recall for GY and upper-tail rank consistency for GMC, PH, EH, and SD. GY, grain yield; GMC, grain moisture content; PH, plant height; EH, ear height; SD, silking date.

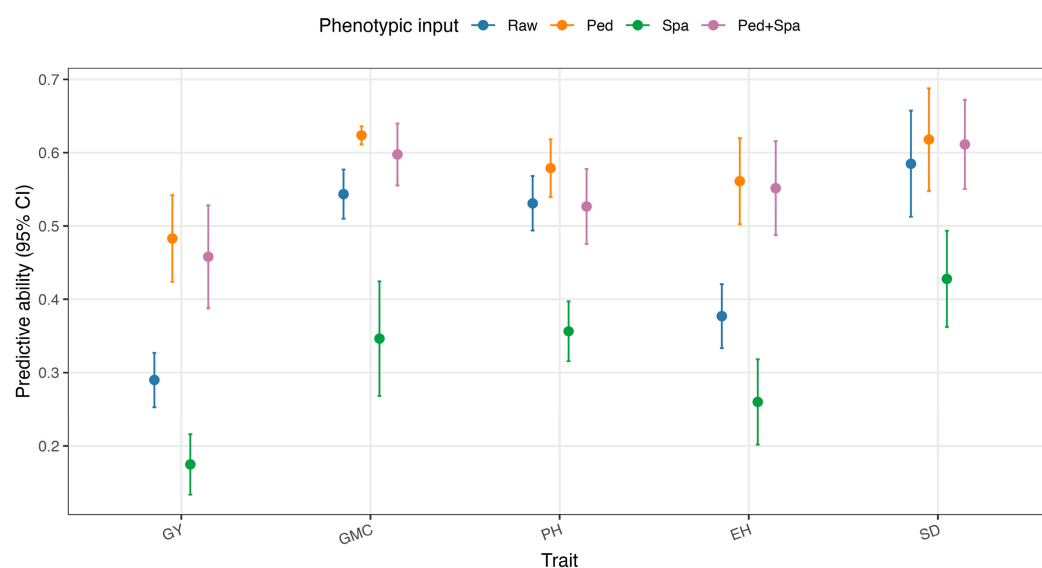


Figure 3. PA by trait and phenotypic input. Points are means and error bars are t-based 95% confidence intervals across 24 location–model combination means for each trait; the 50 random splits were averaged within each combination. Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes; Spa, spatially corrected phenotypes; Ped+Spa, jointly pedigree- and spatially corrected phenotypes. GY, grain yield; GMC, grain moisture content; PH, plant height; EH, ear height; SD, silking date.

3.5. Phenotypic Input Has a Larger Influence than Fixed Model Choice

Phenotypic input was the largest modeled source of variation in predictive ability ($F_{3428} = 229.20$, $p < 0.001$; Table 5). Location, trait, input $\times$ location, and input $\times$ trait were also significant (all $p < 0.001$). In contrast, the main effect of the fixed-configured prediction model was not significant ($F_{5428} = 1.59$, $p = 0.161$), and input $\times$ model was negligible ($F_{15,428} = 0.15$, $p = 0.9999$). Ped exceeded Ped+Spa by 0.024 in matched predictive ability ($t_{119} = 5.57$, $p = 1.58 \times 10^{-7}$), showing that adding the spatial component did not improve the overall target. Across matched combinations, the Ped-minus-Raw change in predictive ability was strongly correlated with the change in Top-20 recall ($r = 0.907$), consistent with the expectation that improvements in average predictive agreement would be accompanied by the improved recovery of top-ranked candidates while not implying that correlation and tail recovery are interchangeable.

3.6. Model Rankings Under the Fixed Benchmark Configurations

Under the fixed benchmark configurations, RKHS had the highest global mean predictive ability (0.488 ± 0.183), followed by AD-GBLUP (0.481 ± 0.186), SVR (0.477 ± 0.168), GBLUP (0.475 ± 0.192), rrBLUP (0.475 ± 0.192), and RF (0.453 ± 0.166). Under Ped, RKHS remained highest (0.589), whereas GBLUP and rrBLUP had the highest Top-20 recall (0.674),

closely followed by RKHS (0.671) and AD-GBLUP (0.663; Table 6). These recall values correspond to mean hit counts of approximately 13.48, 13.42, and 13.26 among the 20 reference candidates. The between-model differences were modest relative to the differences among phenotypic targets. Model rankings were broadly stable across inputs and sites: RKHS ranked first or second under each input, AD-GBLUP ranked first at Suihua, and RF ranked last in the evaluated configurations. However, RF and SVR were not tuned, and RKHS used one prespecified bandwidth rule. The observed ranking therefore describes a controlled baseline comparison rather than the best attainable performance of each algorithm. No model reversed the overall input ordering, reinforcing the conclusion that target construction had a larger influence than fixed model choice.

Table 5. Expanded ANOVA for PA based on location–trait–input–model combination means.

Factor	df	Sum Sq	Mean Sq	F	p
Location	3	2.266	0.755	104.59	<0.001
Trait	4	2.623	0.656	90.83	<0.001
Phenotypic (Input)	3	4.965	1.655	229.20	<0.001
Model	5	0.057	0.011	1.59	0.161
Phenotypic: Location	9	2.170	0.241	33.39	<0.001
Phenotypic: Trait	12	0.480	0.040	5.54	<0.001
Phenotypic: Model	15	0.016	0.001	0.15	0.9999
Residuals	428	3.090	0.007		

Input refers to phenotypic input (Raw, Ped, Spa, or Ped+Spa). df, degrees of freedom; Sum Sq, sum of squares; Mean Sq, mean square. The analysis used 480 location–trait–input–model combination means; residual df = 428.

Table 6. Mean performance of the six genomic prediction models globally and under Ped input.

Model	Global PA	Global Top-20 Recall	PA Under Ped	Top-20 Recall Under Ped
RKHS	0.488 ± 0.183	0.578 ± 0.181	0.589	0.671
AD-GBLUP	0.481 ± 0.186	0.572 ± 0.185	0.580	0.663
SVR	0.477 ± 0.168	0.559 ± 0.159	0.570	0.635
GBLUP	0.475 ± 0.192	0.574 ± 0.191	0.579	0.674
rrBLUP	0.475 ± 0.192	0.574 ± 0.191	0.579	0.674
RF	0.453 ± 0.166	0.541 ± 0.154	0.541	0.617

Ped, pedigree-corrected phenotypes; Top-20 recall is the proportion of 20 upper-tail reference candidates recovered within the predicted top 20% of the test set. rrBLUP, ridge-regression BLUP; GBLUP, genomic BLUP; AD-GBLUP, additive-dominance GBLUP; RF, random forest; SVR, support vector regression; RKHS, reproducing kernel Hilbert space regression.

3.7. Target-Specific Genomic Variance Ratios Support the Pedigree-Correction Interpretation

Target-specific genomic variance ratios were higher under Ped than Raw for most location–trait combinations. Refitting all 80 site–trait–input variance-component models for uncertainty estimation gave an overall mean h^2_g of 0.499 under Raw and 0.665 under Ped. For GY, the across-site mean increased from 0.295 (95% CI 0.155–0.435) under Raw to 0.597 (0.403–0.790) under Ped; for EH, it increased from 0.430 (0.066–0.793) to 0.675 (0.469–0.880; Figure 4). Model-based delta-method standard errors and logit-scale 95% confidence intervals for every individual estimate are reported in Supplementary Table S1. Because variance components were estimated separately for each target, these values describe the fraction of target variance associated with the genomic relationship matrix; they do not show that adjustment changed the intrinsic biological heritability of the germplasm.

3.8. GAM Improves Spatial Correction but Remains Below RAW

Contrary to the expectation of an overall benefit from spatial adjustment, neither spatially corrected target exceeded Raw. With matched location–trait–model combinations and identical random partitions, replacing $AR(1) \times AR(1)$ Spa with a two-dimensional

GAM target increased mean predictive ability from 0.313 ± 0.162 to 0.383 ± 0.150 (mean difference 0.070, 95% CI 0.053–0.087; $t_{119} = 7.96$, $p < 0.001$). Nevertheless, GAM remained below Raw (0.383 ± 0.150 versus 0.465 ± 0.157 ; mean difference -0.082 , 95% CI -0.093 to -0.071 ; $t_{119} = -14.47$, $p < 0.001$; Figure 5). Thus, a more flexible surface recovered part, but not all, of the performance lost under spatial-only target construction. The improvement from Spatial GAM was observed at all four trial sites but varied in magnitude. The largest recovery relative to $AR(1) \times AR(1)$ Spa occurred in Suihua (+0.145), followed by Boli (+0.083), Hongxinglong (+0.039), and Harbin (+0.013). Harbin had the smallest Spatial GAM gain because the original Spa result was already close to Raw, whereas Suihua showed the clearest benefit from replacing the original spatial covariance model with a smoother spatial trend model.

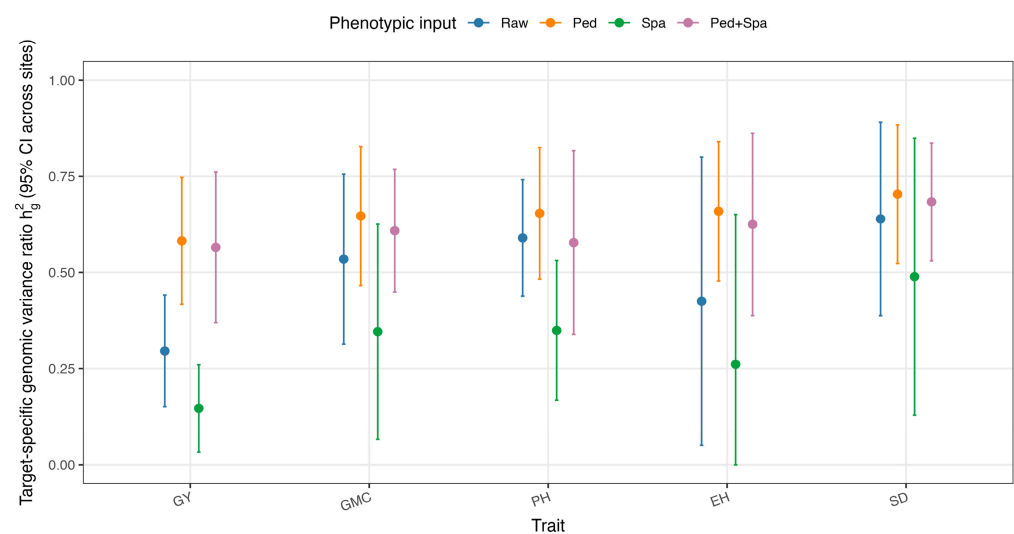


Figure 4. Target-specific genomic variance ratio (h^2g) by trait and phenotypic input. Points are means and error bars are t-based 95% confidence intervals across four refitted site-specific REML estimates; these intervals describe between-site variability. Model-based standard errors and 95% confidence intervals for individual estimates are provided in Supplementary Table S1. Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes; Spa, spatially corrected phenotypes; Ped+Spa, jointly pedigree- and spatially corrected phenotypes. GY, grain yield; GMC, grain moisture content; PH, plant height; EH, ear height; SD, silking date.

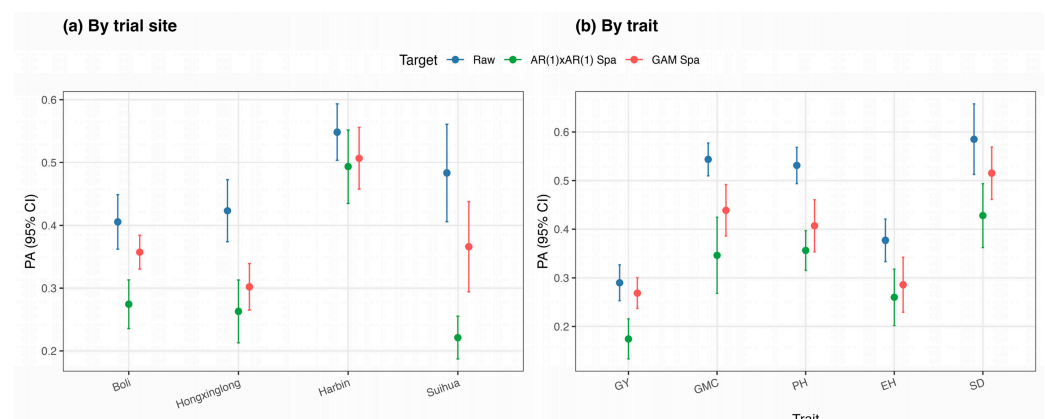


Figure 5. Predictive ability for Raw, $AR(1) \times AR(1)$ Spa, and GAM Spa targets by (a) trial site and (b) trait. Points are means and error bars are t-based 95% confidence intervals across matched combination means; all methods used identical candidate sets, partitions, genomic models, and evaluation criteria. Raw, raw plot-level observations; Ped, pedigree-corrected phenotypes; Spa, spatially corrected phenotypes; Ped+Spa, jointly pedigree- and spatially corrected phenotypes. GY, grain yield; GMC, grain moisture content; PH, plant height; EH, ear height; SD, silking date.

Trait-level patterns showed the same overall trend (Figure 6). Spatial GAM improved PA relative to $AR(1) \times AR(1)$ Spa for all five traits, with the largest gains for grain yield (+0.094), grain moisture content (+0.093), and silking date (+0.087). The smallest gain was observed for ear height (+0.026). Despite this recovery, Spatial GAM did not exceed Raw for any trait on average; the gap was smallest for grain yield (−0.021) and largest for plant height (−0.124).

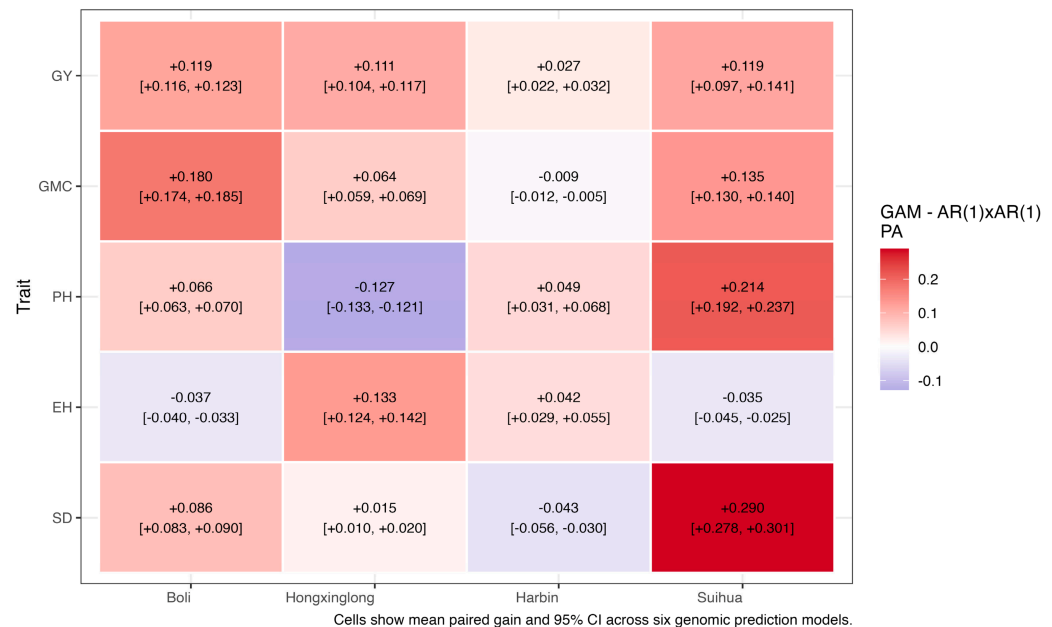


Figure 6. GAM-minus- $AR(1) \times AR(1)$ predictive ability gain by trial site and trait. Each cell reports the mean paired difference and its t-based 95% confidence interval across six genomic prediction models after averaging the 50 splits within each combination. Positive values favor GAM Spa. GY, grain yield; GMC, grain moisture content; PH, plant height; EH, ear height; SD, silking date.

4. Discussion

4.1. Primary Findings in Relation to the Study Expectations

The results were consistent with the expectation that pedigree correction would improve both predictive ability and Top-20 recall relative to Raw across sites, traits, and prediction models. In contrast, the expected overall benefit of spatial adjustment was not observed: both $AR(1) \times AR(1)$ Spa and GAM Spa remained less predictable than Raw, although GAM reduced the deficit. The proposed complementarity of pedigree and spatial information was only partly evident because Ped+Spa exceeded Raw and Spa but not Ped. As expected, the phenotypic input produced a much larger and highly significant performance contrast than the fixed model choice, and the Ped-minus-Raw changes in predictive ability and Top-20 recall were strongly associated. Together, these outcomes identify target construction—not algorithm complexity alone—as the main experimental factor in this dataset.

4.2. Why Ped Succeeded While Spatial-Only and Joint Correction Did Not

Pedigree adjustment can stabilize single-plot phenotypes by borrowing information among related hybrids, which is especially useful when entry observations contain substantial unstructured noise. The larger Ped gains for GY and EH and the higher target-specific genomic variance ratios are consistent with this explanation. They do not prove that noise removal was the sole mechanism: pedigree smoothing can also make a target more predictable because it shares a relationship structure with genomic kernels. Consequently, the

gain should be interpreted as the improved predictability of the Ped target, not automatically as the more accurate prediction of an unknown true breeding value.

Spatial-only correction was more sensitive to identifiability and model form. The diagnostic files recorded non-convergence for Boli PH and GY and Hongxinglong PH and GY, boundary solutions with raw/BLUE fallback for Harbin SD and GMC, and near-zero residual components in several fitted models. The GAM replacement for Hongxinglong PH converged and explained 15.3% of deviance, and the all-trait GAM analysis consistently improved on $AR(1) \times AR(1)$ Spa. However, neither spatial method exceeded Raw on average. A plausible explanation is that, without replication or randomization-based controls sufficient to separate entry and local field effects, spatial smoothing removed a mixture of environmental and entry-associated variation. Ped+Spa may likewise have gained little beyond Ped because pedigree and field terms competed for the same limited single-plot information. These mechanisms remain interpretations rather than directly identified causal effects.

4.3. Relation to Recent Genomic-Prediction Studies

The conditional spatial response agrees with recent evidence indicating that spatial information is useful only when its measurement and covariance structure represent the underlying field process. Morales et al. [19] used longitudinal UAV phenotypes in hybrid maize and found that spatiotemporal permanent-environment effects improved model fit and genotypic-effect estimation relative to baseline spatial models. Our single-season row-column data contain much less information and produced weaker spatial separation. In multi-environment maize, Costa et al. [12] showed that selection history and genotype-by-environment variance changed indirect prediction accuracy, while Barreto et al. [13] reported that low-heritability environments were the most difficult to predict and that statistical and machine-learning performance was case dependent. These findings help explain the large site and trait interactions observed here. Recent ranking-oriented studies [22–24] also show that correlation alone does not characterize recovery at breeding thresholds, supporting the use of Top-N recall as a complementary decision metric.

4.4. Fixed Model Configurations and the Scope of Model Comparisons

The absence of a significant model main effect does not imply that all genomic algorithms are universally equivalent. The six implementations used common fixed settings so that target construction could be compared under a controlled computational baseline. RKHS was marginally highest and AD-GBLUP was second overall, but the gaps were small. A 2026 hybrid maize study used explicit hyperparameter optimization and dominance-informed marker transformations to improve machine-learning prediction [39], illustrating that tuned nonlinear methods address a different question from our fixed benchmark. Because RF and SVR were not tuned here, their lower ranks cannot be interpreted as inherent limitations. For routine early screening, rrBLUP or GBLUP remain transparent baselines, whereas claims about optimal nonlinear performance require nested tuning and independent validation.

4.5. Implications for Breeding Programs and Future Research

For the evaluated early-stage trials, pedigree adjustment provided a practical way to construct a more predictable target and improved recovery of high-yield candidates within a 20% retention window. A defensible workflow is therefore to compare Raw and Ped as primary targets, inspect spatial diagnostics before accepting Spa or Ped+Spa, and report both correlation and decision-threshold metrics. Future work should validate these findings with replicated checks, multiple years and environments, independent populations, and predeclared spatial models; compare training-set proportions; and evaluate predictions

against replicated multi-environment reference values. Trait-specific selection should use high-tail recall for GY, low-tail recall for traits such as GMC or SD when low/early values are desired, and distance-to-ideotype measures for PH and EH when intermediate targets balance lodging, biomass, and harvest requirements. Cost–benefit analysis should then determine whether the incremental recovery of elite candidates justifies pedigree curation and spatial data collection.

4.6. Limitations

This study has seven principal limitations. First, trials were conducted in one year and four locations, limiting inference across seasons and broader target populations of environments. Second, all entries were unreplicated, so genetic and spatial signals could not always be separated. Third, Boli and Hongxinglong shared hybrids, whereas Harbin and Suihua used independent sets, confounding location with population composition and marker density. Fourth, several spatial or joint models showed non-convergence, boundary estimates, or fallbacks. Fifth, predictive ability was evaluated against each constructed target rather than a replicated reference phenotype or true breeding value; greater Ped predictability may partly reflect a shared relationship structure. Sixth, the 30% training fraction represents a demanding early-screening scenario and may not generalize to larger training sets. Seventh, Top-N was calculated from the upper tail for all traits. It directly represents favorable selection only for GY; lower GMC, earlier SD, or lower/target PH and EH require direction-specific or ideotype-distance metrics. These limitations mean that the superiority of Ped and the inferiority of Spa should not be generalized to replicated trials, different populations, species, layouts, or spatial models without validation.

5. Conclusions

Across four single-replicate maize hybrid trials, pedigree correction improved predictive ability and upper-tail Top-N recall relative to Raw. Spatial-only correction did not improve overall performance, although GAM recovered part of the deficit relative to $AR(1) \times AR(1)$. Ped+Spa exceeded Raw and Spa but not Ped, providing only partial evidence that pedigree and spatial information contributed complementary gains. Phenotypic target construction had a much larger influence than fixed baseline model choice, and gains in predictive ability were strongly associated with gains in Top-20 recall. The results justify Ped as the most useful target among those evaluated, but they do not establish universal superiority across years, populations, trial designs, or tuned algorithms. Genomic-prediction reports for early screening should therefore define the target, document correction diagnostics and fallbacks, separate target predictability from true-value accuracy, and align Top-N metrics with the favorable direction or ideotype of each trait.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy16151428/s1>, Table S1: model-based uncertainty; Table S2: GAM parameters; Table S3: site specific tests; Table S4: rrBLUP/GBLUP audit; Table S5: GAM-AR1 matched comparison.

Author Contributions: Conceptualization, S.C.; Writing—original draft, H.L., W.P., and S.C.; Supervision, H.L.; Visualization, W.P.; Investigation, S.J. and Z.N.; Resources, S.C., X.Z., and Y.W.; Data curation, W.P., Y.S., and Z.N.; Methodology, H.L., X.Z., T.Y., G.Y., W.L., and H.W.; Software, H.L. and Z.N.; Formal analysis, S.J. and Y.S.; Validation, T.Y. and G.Y.; Writing—review & editing, S.C., T.Y., G.Y., W.L., and H.W.; Project administration, S.C. and Y.W.; Funding acquisition, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Heilongjiang Seed Industry Innovation and Development Project (grant no. 2024HZYCXNK06). The funding agencies had no role in the study design, data collection, analysis and interpretation, or manuscript preparation.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (OpenAI; GPT-5.5) for language editing. The authors reviewed and edited the outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors Shuqin Jiang and Xiaoguang Zhai are employed by Qingdao KingAgroot Seed Science Co., Ltd. and Molbreeding Biotechnology Co., Ltd., respectively. All authors jointly declare that there are no commercial or financial relationships relevant to this study that could be regarded as potential conflicts of interest.

Abbreviations

Ped	Pedigree-corrected phenotype
Spa	Spatially corrected phenotype
GAM	Generalized Additive Model
GEBV	Genomic Estimated Breeding Value
Raw	Raw plot-level observation
Ped+Spa	Jointly pedigree- and spatially corrected phenotype
BLUP	Best linear unbiased prediction
GBLUP	Genomic best linear unbiased prediction
AD-GBLUP	Additive-dominance genomic best linear unbiased prediction
RF	Random forest
SVR	Support vector regression
RKHS	Reproducing kernel Hilbert space
AR(1)	First-order autoregressive model

References

1. Varshney, R.K.; Roorkiwal, M.; Sorrells, M.E. Genomic Selection for Crop Improvement: An Introduction. In *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*; Varshney, R.K., Roorkiwal, M., Sorrells, M.E., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 1–6, ISBN 978-3-319-63170-7.
2. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)] [[PubMed](#)]
3. Albrecht, T.; Wimmer, V.; Auinger, H.J.; Erbe, M.; Knaak, C.; Ouzunova, M. Genome-Based Prediction of Testcross Values in Maize. *Theor. Appl. Genet.* **2011**, *123*, 339–350. [[CrossRef](#)] [[PubMed](#)]
4. Cullis, B.R.; Smith, A.B.; Coombes, N.E. On the Design of Early Generation Variety Trials with Correlated Data. *J. Agric. Biol. Environ. Stat.* **2006**, *11*, 381–393. [[CrossRef](#)]
5. Galli, G.; Sabadin, F.; Yassue, R.M.; Galves, C.; Carvalho, H.F.; Crossa, J.; Montesinos-López, O.A.; Fritsche-Neto, R. Automated Machine Learning: A Case Study of Genomic “Image-Based” Prediction in Maize Hybrids. *Front. Plant Sci.* **2022**, *13*, 845524. [[CrossRef](#)] [[PubMed](#)]
6. Alemu, A.; Åstrand, J.; Montesinos-López, O.A.; Isidro y Sánchez, J.; Fernández-González, 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](#)] [[PubMed](#)]
7. Mothukuri, S.R.; Beyene, Y.; Gültas, M.; Burgueño, J.; Griebel, S. Optimization of Sparse Phenotyping Strategy in Multi-Environmental Trials in Maize. *Theor. Appl. Genet.* **2025**, *138*, 62. [[CrossRef](#)] [[PubMed](#)]
8. Technow, F.; Schrag, T.A.; Schipprack, W.; Bauer, E.; Simianer, H.; Melchinger, A.E. Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize. *Genetics* **2014**, *197*, 1343–1355. [[CrossRef](#)] [[PubMed](#)]
9. Crossa, J.; Perez-Rodriguez, P.; Cuevas, J.; Montesinos-Lopez, O.; Jarquín, D.; de los Campos, G. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [[CrossRef](#)] [[PubMed](#)]

10. 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](#)]
11. Velazco, J.G.; Malosetti, M.; Hunt, C.H.; Mace, E.S.; Jordan, D.R.; van Eeuwijk, F.A. Combining Pedigree and Genomic Information to Improve Prediction Quality: An Example in Sorghum. *Theor. Appl. Genet.* **2019**, *132*, 2055–2067. [[CrossRef](#)] [[PubMed](#)]
12. Costa, M.; Holland, J.B.; de Leon, N.; Kaeppler, S.M. Impact of Genotype $\times$ Environment Interaction and Selection History on Genomic Prediction in Maize (*Zea mays* L.). *Crop Sci.* **2024**, *64*, 3293–3310. [[CrossRef](#)]
13. 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](#)]
14. Piepho, H.P.; Moehring, J.; Melchinger, A.E.; Buechse, A. BLUP for Phenotypic Selection in Plant Breeding and Variety Testing. *Euphytica* **2008**, *161*, 209–228. [[CrossRef](#)]
15. Velazco, J.G.; Rodriguez-Alvarez, M.X.; Boer, M.P.; Jordan, D.R.; Eilers, P.H.C.; Malosetti, M.; van Eeuwijk, F.A. Modelling Spatial Trends in Sorghum Breeding Field Trials Using a Two-Dimensional P-Spline Mixed Model. *Theor. Appl. Genet.* **2017**, *130*, 1375–1392. [[CrossRef](#)] [[PubMed](#)]
16. Elias, A.A.; Rabbi, I.; Kulakow, P.; Jannink, J.-L. Improving Genomic Prediction in Cassava Field Experiments Using Spatial Analysis. *G3 Genes Genomes Genet.* **2018**, *8*, 53–62. [[CrossRef](#)] [[PubMed](#)]
17. Oakey, H.; Verbyla, A.P.; Cullis, B.R.; Wei, X.; Pitchford, W.S. Joint Modeling of Additive and Non-Additive Genetic Line Effects in Multi-Environment Trials. *Theor. Appl. Genet.* **2007**, *114*, 1319–1332. [[CrossRef](#)] [[PubMed](#)]
18. Gilmour, A.R.; Cullis, B.R.; Verbyla, A.P. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. *J. Agric. Biol. Environ. Stat.* **1997**, *2*, 269–293. [[CrossRef](#)]
19. Morales, N.; Anche, M.T.; Kaczmar, N.S.; Lepak, N.; Ni, P.; Romay, M.C.; Santantonio, N.; Buckler, E.S.; Gore, M.A.; Mueller, L.A.; et al. Spatio-Temporal Modeling of High-Throughput Multi-Spectral Aerial Images Improves Agronomic Trait Genomic Prediction in Hybrid Maize. *Genetics* **2024**, *227*, iyae037. [[CrossRef](#)] [[PubMed](#)]
20. Ornella, L.; Pérez, P.; Tapia, E.; González-Camacho, J.M.; Burgueño, J.; Zhang, X.; Singh, S.; Vicente, F.S.; Bonnett, D.; Dreisigacker, S.; et al. Genomic-Enabled Prediction with Classification Algorithms. *Heredity* **2014**, *112*, 616–626. [[CrossRef](#)] [[PubMed](#)]
21. Blondel, M.; Onogi, A.; Iwata, H.; Ueda, N. A Ranking Approach to Genomic Selection. *PLoS ONE* **2015**, *10*, e0128570. [[CrossRef](#)] [[PubMed](#)]
22. Heinrich, F.; Lange, T.M.; Ramzan, F.; Gültas, M.; Schmitt, A.O. Normalized Cumulative Gain as an Alternative Evaluation Measure for Genomic Selection Models. *Genet. Sel. Evol.* **2025**, *57*, 70. [[CrossRef](#)] [[PubMed](#)]
23. Montesinos-López, O.A.; Kismiantini; Montesinos-López, A. Two Simple Methods to Improve the Accuracy of the Genomic Selection Methodology. *BMC Genom.* **2023**, *24*, 220. [[CrossRef](#)] [[PubMed](#)]
24. Montesinos-López, O.A.; Kismiantini; Alemu, A.; Montesinos-López, A.; Montesinos-López, J.C.; Crossa, J. Balancing Sensitivity and Specificity Enhances Top and Bottom Ranking in Genomic Prediction of Cultivars. *Plants* **2025**, *14*, 308. [[CrossRef](#)] [[PubMed](#)]
25. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)] [[PubMed](#)]
26. Endelman, J.B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* **2011**, *4*, 250–255. [[CrossRef](#)]
27. VanRaden, P.M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)] [[PubMed](#)]
28. Vitezica, Z.G.; Varona, L.; Legarra, A. On the Additive and Dominant Variance and Covariance of Individuals Within the Genomic Selection Scope. *Genetics* **2013**, *195*, 1223–1230. [[CrossRef](#)] [[PubMed](#)]
29. Gianola, D.; Van Kaam, J.B.C.H.M. Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. *Genetics* **2008**, *178*, 2289–2303. [[CrossRef](#)] [[PubMed](#)]
30. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27.
32. Terrailon, J.; Roeber, F.K.; Flachenecker, C.; Frisch, M. Training Set Designs for Prediction of Yield and Moisture of Maize Test Cross Hybrids with Unreplicated Trials. *Front. Plant Sci.* **2023**, *14*, 1080087. [[CrossRef](#)] [[PubMed](#)]
33. Mao, X.; Dutta, S.; Wong, R.K.W.; Nettleton, D. Adjusting for Spatial Effects in Genomic Prediction. *J. Agric. Biol. Environ. Stat.* **2020**, *25*, 699–718. [[CrossRef](#)]
34. Cullis, B.; Gogel, B.; Verbyla, A.; Thompson, R. Spatial Analysis of Multi-Environment Early Generation Variety Trials. *Biometrics* **1998**, *54*, 1–18. [[CrossRef](#)]
35. Stefanova, K.T.; Smith, A.B.; Cullis, B.R. Enhanced Diagnostics for the Spatial Analysis of Field Trials. *J. Agric. Biol. Environ. Stat.* **2009**, *14*, 392–410. [[CrossRef](#)]
36. Hawinkel, S.; De Meyer, S.; Maere, S. Spatial Regression Models for Field Trials: A Comparative Study and New Ideas. *Front. Plant Sci.* **2022**, *13*, 858711. [[CrossRef](#)] [[PubMed](#)]

37. Rodriguez-Alvarez, M.X.; Boer, M.P.; van Eeuwijk, F.A.; Eilers, P.H.C. Correcting for Spatial Heterogeneity in Plant Breeding Experiments with P-Splines. *Spat. Stat.* **2018**, *23*, 52–71. [[CrossRef](#)]
38. Gogel, B.; Welham, S.; Cullis, B. Empirical Comparison of Time Series Models and Tensor Product Penalised Splines for Modelling Spatial Dependence in Plant Breeding Field Trials. *Front. Plant Sci.* **2023**, *13*, 1021143. [[CrossRef](#)] [[PubMed](#)]
39. Osatohanmwen, B.E.; Vieira, I.C.; Gholami, M.; Westhues, C.C.; Sharifi, A.R.; Beissinger, T.M. Predicting Maize Hybrid Performance with Machine Learning and a Locus-Specific Weighted Degree of Dominance Transformation. *Front. Plant Sci.* **2026**, *17*, 1694707. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.