Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting GxE through phenomic prediction: A benchmark study across three species

Clément Bienvenu
CIRAD
Hugues de Verdal
CIRAD
Vincent Segura
Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier
David Pot
david.pot@cirad.fr

CIRAD

Research Article

Keywords: G × E modeling, Phenomic prediction, Genomic prediction, Benchmark

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11626000/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Predicting $G \times E$ through phenomic prediction: A benchmark study across three species

Clément Bienvenu[1,2], Hugues de Verdal[1,2†], Vincent Segura[2,3†], David Pot[1,2*†]

[1]UMR AGAP Institut, CIRAD, Montpellier, F-34398, France.
[2]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, Montpellier, F-34398, France.
[3]Geno-Vigne, IFV-INRAE-Institut Agro, Montpellier, F-34398, France.

*Corresponding author(s). E-mail(s): david.pot@cirad.fr;
Contributing authors: bienvenu.clement@gmail.com;
hugues.de_verdal@cirad.fr; vincent.segura@inrae.fr;
†These authors contributed equally to this work.

Abstract

Phenomic prediction (PP) is a genetic value prediction method inspired by genomic prediction (GP). One key feature of PP compared to GP is its potential ability to predict interactions between genotypes and environments ($\boldsymbol{G} \times \boldsymbol{E}$). Using kernel-based models, many relationship matrix calculation methods for the $\boldsymbol{G}$ and $\boldsymbol{G} \times \boldsymbol{E}$ terms have been proposed but never fairly compared. In this study, we exhaustively tested different combinations of published methods to compute these matrices, and implemented a new method supposed to better split the genetic ($\boldsymbol{G}$) and $\boldsymbol{G} \times \boldsymbol{E}$ information in their respective matrices. Using three datasets on maize, rice, and triticale, results show that the $\boldsymbol{G}$ effect matrix is the main driver of predictive abilities (PA). Computing the $\boldsymbol{G}$ effect matrix using BLUPs of wavelengths across environments was better when predicting new genotypes in known environments. To predict known genotypes in new environments or in a sparse testing scenario, it was better to compute one relationship matrix per environment and average them into a single $\boldsymbol{G}$ effect matrix. We found little differences between models including a $\boldsymbol{G} \times \boldsymbol{E}$ term and models without it. Nevertheless, dense $\boldsymbol{G} \times \boldsymbol{E}$ matrices including relationships within and across environments were generally the best choice. Finally, combining genomic information for the $\boldsymbol{G}$ effect, and phenomic information for the $\boldsymbol{G} \times \boldsymbol{E}$ effect gave the best results. Clear guidance of the best ways to take advantage of phenomic

prediction to predict $\boldsymbol{G} \times \boldsymbol{E}$ will likely require a simulation approach to better control the $\boldsymbol{G} \times \boldsymbol{E}$ information in datasets.

Keywords: $\boldsymbol{G} \times \boldsymbol{E}$ modeling, Phenomic prediction, Genomic prediction, Benchmark

Acknowledgment and funding

The work of Clément Bienvenu was supported by a doctoral allowance from the French Ministry of Higher Education and Research. Authors would like to thank Tobias Würschum for making the triticale dataset available, and Aaron J. DeSalvio for sharing the maize dataset on Git.

ORCID

Clément Bienvenu: https://orcid.org/0009-0004-9234-6216
David Pot: https://orcid.org/0000-0001-6144-8448
Hugues de Verdal: https://orcid.org/0000-0002-1923-8575
Vincent Segura: https://orcid.org/0000-0003-1860-2256

1 Introduction

Phenomic prediction (PP) (Rincent et al. 2018) is a genetic value prediction method inspired by genomic prediction (GP) (Meuwissen et al. 2001). GP uses molecular marker information to make predictions while PP uses high throughput phenotypic (HTP) information such as near infrared spectroscopy (NIRS) on different plant tissues, or hyperspectral images acquired with drones (UAV) or other unmanned ground vehicle (UGV) at the plot level. Some authors argued that phenomic data can capture endophenotypic information such as transcriptomic information or other (-omic) signals that serve as intermediate between DNA sequence and phenotypic expression (Zhu et al. 2021; Rincent et al. 2018). Indeed, NIR spectra or hyperspectral images are phenotypes resulting from additive genetic effects but also environmental, dominance (depending on the genetic context), epistasis, and pleiotropic effects. Thus, one key feature of PP is that the predictor data contains more than just genetic information. Consequently, PP may be useful in breeding programs to specifically predict genotype by environment interactions ($G \times E$).

Several studies have used PP to predict GxE interactions or environment specific genetic values of genotypes, all reporting accurate predictions using PP. These studies generally concluded that using HTP data alone or as additional information in a GP model could accurately predict genetic values (Adak et al. 2023a, 2024, 2023b, 2022; Cuevas et al. 2019; De Salvio et al. 2024; de Verdal et al. 2024; Fumia et al. 2023; Galán et al. 2021; Jackson et al. 2023; Klein et al. 2026; Krause et al. 2019; Lane et al. 2020; Montesinos-López et al. 2024; Parmley et al. 2019; Pereira et al. 2025; Robert et al. 2022; Tadesse et al. 2026; Togninalli et al. 2023; Winn et al. 2023; Sapkota et al. 2025).

Among these studies, some compared the accuracy of GP and PP. Most studies found that PP had similar or marginally higher predictive ability (PA) than GP (Adak et al. 2024, 2023b, 2022; Klein et al. 2026; Krause et al. 2019; Robert et al. 2022). Others found that PP had significantly higher (Winn et al. 2023; Jackson et al. 2023; Galán et al. 2021), or lower (de Verdal et al. 2024; Cuevas et al. 2019) PA than GP, or both depending on the prediction scenario and the trait predicted (De Salvio et al. 2024; Sapkota et al. 2025).

Some studies also tested models combining phenomic and genomic information, which did not significantly outperform genomic-only or phenomic-only models in most cases (Adak et al. 2024, 2023b; Cuevas et al. 2019; Jackson et al. 2023; Krause et al. 2019). On the contrary, other studies found that models combining genomic and phenomic information had the best PA (Togninalli et al. 2023; Klein et al. 2026; Sapkota et al. 2025).

Finally, using phenomic information, few studies focused on the impact of computing models with or without a $G \times E$ term. Some studies compared multi-environments models with a $G \times E$ effect to single-environments models, or multi-environment models without a $G \times E$ effect. They showed that multi-environment models including a $G \times E$ effect could have higher PA than single environment models (Cuevas et al. 2019;

de Verdal et al. 2024), and that multi-environment models with a $G \times E$ term had generally higher PA than multi-environment models without a $G \times E$ term (Robert et al. 2022). In Lane et al. (2020), a similar comparison was made but models with or without a $G \times E$ term were not compared in similar cross validations scenarios. In their study, models without a $G \times E$ term had better PA, but this may be due to a more favorable cross validation scenario.

The reference GBLUP/HBLUP model is widely used in GP and PP studies with $G \times E$ prediction (Montesinos-López et al. 2024; de Verdal et al. 2024; Robert et al. 2022; De Salvio et al. 2024; Lane et al. 2020; Cuevas et al. 2019; Krause et al. 2019; Galán et al. 2021). Computing relationship matrices between genotypes within and across environments is a critical step when using it. GP uses marker data which have only one SNP profile per genotype making computation methods straightforward for relationship matrices of $G \times E$ effects. PP on the other hand uses NIRS for which each genotype can have one or multiple spectra per environment. Thus, spectra are generally pre-processed using linear models to compute genetic values of each wavelength which are then assembled to construct the genetic values of spectra. Either one value per genotype across environments or one value per genotype per environment can be computed to have spectral relationship matrices between genotypes within or across environments. Methods from the literature to extract genetic values of spectra and compute relationship matrices in PP are summarized in table 1. Different ways of describing the same method are found in different studies as there is no clear and homogeneous established formalism for such models in the PP community. They were homogenized and written in the most general form in table 1 explaining the discrepancies between their formalism in the original studies and the one in this work. Figure 1 summarizes the different types of matrices described and helps visualizing the information present in these matrices. Readers may refer to this figure while reading the next paragraph.

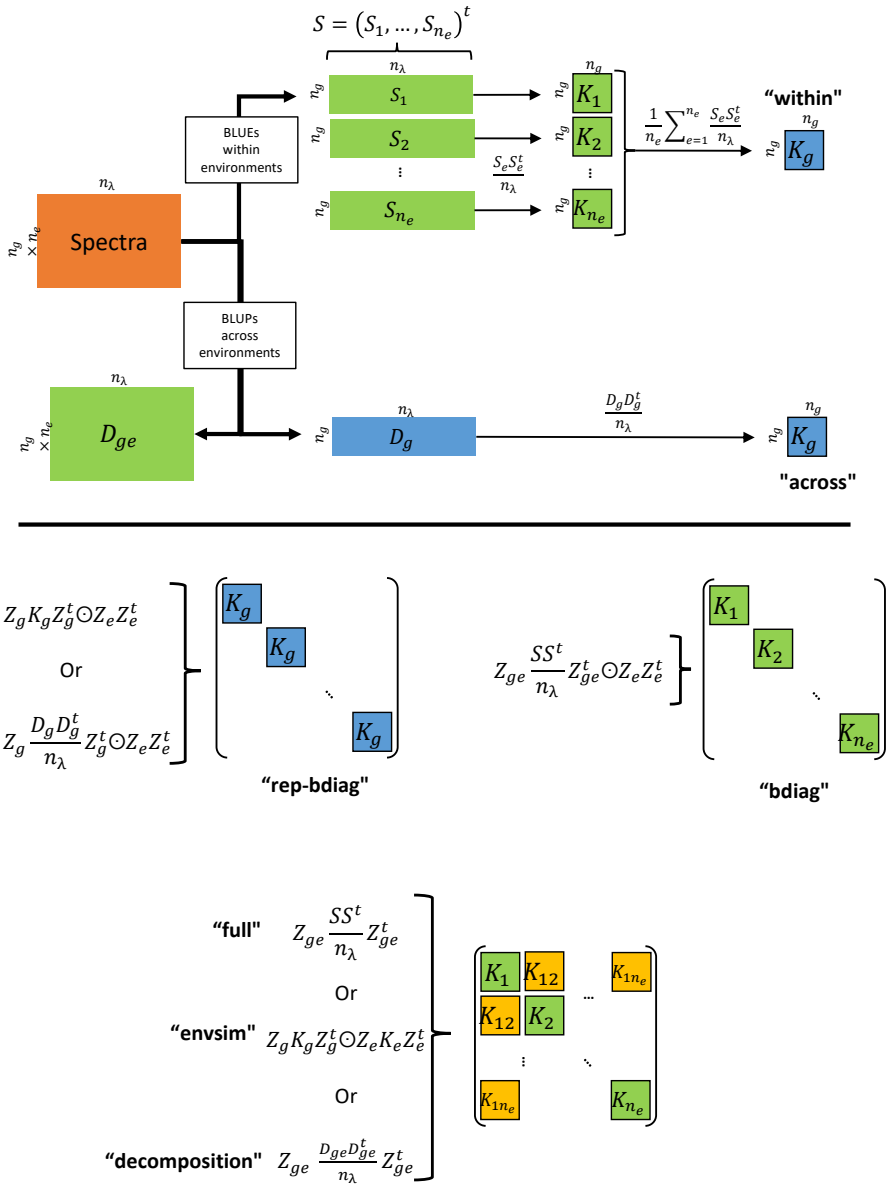


Fig. 1 Visualization of the different matrices used in prediction models. n_λ = number of wavelengths, n_g = number of genotypes, n_e = number of environments, BLUE/BLUP = best linear unbiased estimator/predictor of genetic values, S_e = BLUEs of genotypes within environment e, ^{t} = matrix transposition, D_g = BLUPs of genotypes across environments, D_{ge} = BLUPs of the $G \times E$ effect, $K_1, ..., K_{n_e}$ = relationship matrices between genotypes within environments 1 to n_e, K_g = relationship matrix between genotypes across environments, Z_g design matrix of the main genetic effect in the prediction model, Z_e design matrix of the environmental effect in the prediction model, Z_{ge} = design matrix of the $G \times E$ effect in the prediction model, $\odot$ = Hadamard product, $K_{12}, ..., K_{1n_e}$ relationship matrix between genotypes in environment 1 and 2, to 1 and n_e.

Table 1 Summary of the methods used to compute relationship matrices for $G \times E$ phenomic prediction models. S_e = BLUEs/BLUPs of wavelengths for environment e, $S = (S_1, ..., S_{n_e})^t$ where ^{t} indicate the transpose matrix, n_e = number of environments, D_g = BLUEs/BLUPs of wavelengths matrix across environments, n_λ = number of wavelengths of S_e, S, or D, $\odot$ = Hadamard product, Z_g = design matrix for the G effect, Z_{ge} design matrix for the $(G \times E)$ effect, Z_e = design matrix for the E effect, K_g = relationship matrix used for the G effect, K_{ge} = relationship matrix used for the $(G \times E)$ effect, K_e = relationship matrix used for the E effect, "\" = not specified. Grand means and residuals are present in all prediction models but were not specified here to gain space. [1]Other models combining marker and spectral information, or using only spectra from one environment are also specified in this study but not relevant here. [2]In these studies, for the main genetic effect, the $Z_g K_g Z_g^t$ for the $Z_g u_g$ vector is described instead of K_g matrix for the u_g vector. [3]K_e is computed using environmental covariables, which are also accounted for as fixed effect in the $X_{ik}\beta_k$ effects while multi-spectral information is accounted for as fixed effects in the $M_{ikl}\beta_{M,l}$ effects. [4]K_e is computed by doing a principal component analysis (PCA) on phenomic data within each environment and calculating the Pearson correlation between the first components of each PCA, other models using identity matrix for the G, E or $G \times E$ effect but were not considered relevant here.

Adjustment Model	Prediction Model	K_g	K_{ge}	Publication
BLUEs per environment in single-environment models	$Y = E + (G \times E)$	\	$Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$ *(f)*	Krause et al. (2019)
	$Y = G + E + (G \times E)$	With marker data	$Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t \odot Z_e Z_e^t$ *(e)*	
BLUEs per environment in single-environment models	$Y = G + E + (G \times E)$	$\frac{1}{n_e}\sum_{e=1}^{n_e}\frac{S_e S_e^t}{n_\lambda}$ *(a)*	$Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$ *(f)*	Lane et al. (2020)
BLUEs across environments	$Y = G + E + (G \times E)$	$\frac{D_g D_g^t}{n_\lambda}$ *(b)*	$Z_g K_g Z_g^t \odot Z_e Z_e^t$ *(d)*	Galán et al. (2021)
BLUPs per environment in single-environment models	$Y = G + E + (G \times E)$	$\frac{1}{n_e}\sum_{e=1}^{n_e}\frac{S_e S_e^t}{n_\lambda}$ *(a)*	$Z_g K_g Z_g^t \odot Z_e Z_e^t$ *(d)*	Robert et al. (2022)[1]
BLUPs per environment in single-environment models	Models combining genomic and two sources of phenomic information	$Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$ *(c)*	$Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t \odot Z_e Z_e^t$ *(e)*	Adak et al. (2023b)[2]
BLUPs per environment in a multi-environment model	$Y = G + E + (G \times E)$	$Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$ *(c)*	$Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t \odot Z_e Z_e^t$ *(e)*	De Salvio et al. (2024)[2]
BLUPs across environments	$Y = E + G$	$\frac{D_g D_g^t}{n_\lambda}$ *(b)*	\	de Verdal et al. (2024)
	$Y = G + E + (G \times E)$	$\frac{D_g D_g^t}{n_\lambda}$ *(b)*	$Z_g K_g Z_g^t \odot Z_e Z_e^t$ *(d)*	
\	$Y = G + E + (G \times E) + \sum_{k=1}^{r} X_{ik}\beta_k + \sum_{l=1}^{s} M_{ikl}\beta_{M,l}$	With marker data	$Z_g K_g Z_g^t \odot Z_e K_e Z_e^t$ *(g)*	Montesinos-López et al. (2024)[3]
BLUPs per environment in one multi-environment model	$Y = G + E + (G \times E)$	With marker data	$Z_g K_g Z_g^t \odot Z_e K_e Z_e^t$ *(g)*	Aono and Chawade (2026)[4]

Using phenomic information, three ways to compute relationship matrices between genotypes for the main genetic effect (G) can be found in the literature: (a) BLUEs/BLUPs of spectra are first computed within each environment (with S_e the BLUEs/BLUPs of spectra for environment e) yielding one value per genotype per environment, one relationship matrix is computed within each environment ($\frac{S_e S_e^t}{n_\lambda}$ where n_λ is the number of wavelengths and t means transpose) and these matrices are then averaged ($\frac{1}{n_e}\sum_{e=1}^{n_e}\frac{S_e S_e^t}{n_\lambda}$ where n_e is the total number of environments) to have one final relationship matrix between genotypes across environments (Lane et al. 2020; Robert et al. 2022), which corresponds to the "within" matrix in figure 1.

(b) Best linear unbiased estimators/predictors (BLUEs/BLUPs) of spectra are first computed across environment (D_g) yielding one single value per genotype and are then used to compute one relationship matrix between genotypes ($\frac{D_g D_g^t}{n_\lambda}$) (Galán et al. 2021; de Verdal et al. 2024). This corresponds to the "across" matrix in figure 1.

Finally, (c) some authors reported relationship matrices for the vector of total genetic effects instead of the vector of genotypic BLUEs/BLUPs (De Salvio et al. 2024; Adak et al. 2023b). For example, in a model, if Zu is the genetic term, articles using methods (a) and (b) reported matrices related to the variance of u ($var(u) = K\sigma_g$), while De Salvio et al. (2024); Adak et al. (2023b) reported matrices related to the variance of Zu ($var(Zu) = ZKZ^t\sigma_g$) where ZKZ^t is the matrix described. In both articles, the described matrix is computed by concatenating all environment specific BLUEs/BLUPs of spectra (with $S = (S_1, ..., S_{n_e})^t$ the concatenation of within environment BLUEs/BLUPs of spectra) and using them to compute the similarity between all genotypes in all environments ($Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$, where Z_{ge} is the design matrix of the $G \times E$ effect), which corresponds to the "full" matrix in figure 1. Markers were used for the G effect in Krause et al. (2019); Aono and Chawade (2026); Montesinos-López et al. (2024). Additionally, Krause et al. (2019) also proposed models with no G effect.

Four ways of computing the relationship matrix for the $G \times E$ effect can be found in the literature:

(d) the first approach uses the G effect matrix in a Hadamard product with the design matrices of the effect of the environment (Robert et al. 2022; de Verdal et al. 2024; Galán et al. 2021): $Z_g K_g Z_g^t \odot Z_e Z_e^t$ (where Z_g is the design matrix of the G effect and Z_e is the design matrix of the E effect. It corresponds to the "rep-bdiag" matrix in figure 1. With data sorted by environments, this would yield a block diagonal matrix, each block being identical and equal to the relationship matrix between genotypes as computed for the G effect. This assumes that the relationship between genotypes is identical in each environment (two genotypes have the same relationship value in all

environments), and that genotypes in different environments are independent (their covariance is 0).

(e) In the second approach, a large relationship matrix is computed using the concatenated BLUEs/BLUPs of spectra from all environments ($Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$) and use this matrix in a Hadamard product with Z_e (Krause et al. 2019; De Salvio et al. 2024; Adak et al. 2023b): $Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t \odot Z_eZ_e^t$, which corresponds to the "bdiag" matrix in figure 1. With data sorted by environments, this would yield a block diagonal matrix where each block is different and corresponds to the environment specific relationship matrices between genotypes. This assumes that relationships between genotypes are environment specific (two genotypes have different relationship values in different environments) and that genotypes from different environments are independent (their covariance is 0).

(f) The third approach is based on the calculation of a large relationship matrix using the concatenated BLUEs/BLUPs of spectra from all environments (Krause et al. 2019; Lane et al. 2020): $Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$, which corresponds to the "full" matrix in figure 1. With data sorted by environments, this would yield a dense matrix with block diagonal parts corresponding to the environment specific relationship matrices between genotypes, and off-block-diagonal parts corresponding to relationships between genotypes in different environments. This assumes that relationships between genotypes are environment specific (two genotypes have different relationship values in different environments), and that genotypes in different environments are related depending on their own relationship and the relationship between the environments they were grown into.

(g) The fourth strategy uses the G effect matrix in a Hadamard product with an environmental similarity matrix (Montesinos-López et al. 2024; Aono and Chawade 2026): $Z_gK_gZ_g^t \odot Z_eK_eZ_e^t$ (where K_e is a relationship matrix between environments), which corresponds to the "envsim" matrix in figure 1. In Montesinos-López et al. (2024), K_e

is computed using environmental variables, and in Aono and Chawade (2026), K_e is computed by doing a principal component analysis (PCA) on phenomic data within each environment and calculating the Pearson correlation between the first components of each PCA. With data sorted by environments, this would yield a dense matrix with the same properties and assumptions as the "full" matrix described above, but with different relationship values.

In models with a G and a $G \times E$ interaction effect, matrices used for each effect should reflect their respective variance-covariance structures. However, existing methods use redundant information in the G and $G \times E$ effect matrices. Indeed, the $G \times E$ matrices are computed either by re-using the G effect matrix as is (like in genomic prediction), or by using the same genetic values of spectra that were already used to compute the matrix of the G effect. In the latter case, within environment genetic values are used, which contain both genetic and $G \times E$ information and not only $G \times E$ information.

In this study, we propose to exhaustively test the methods used in PP to predict $G \times E$ interactions by testing all combinations between the different ways of computing the G and $G \times E$ effect matrices. We also tested a way of computing $G \times E$ matrices that explicitly uses only $G \times E$ information of the spectra without also accounting for G effect variance covariance structure. This was done using three datasets already used in the PP literature representing three species: rice (de Verdal et al. 2024), maize (Lane et al. 2020; De Salvio et al. 2024), and triticale (Zhu et al. 2022). These datasets encompass different panel sizes (from 190 rice accessions to 1216 triticale ones), different genotype structure (pure lines and doubled haploids for rice and triticale, hybrid testcrosses for maize), different LD decay patterns and population structures, and different $G \times E$ levels offering a diversity of situations relevant to explore the behaviors of the different prediction models.

2 Material and methods

2.1 Datasets

The three datasets used in this study were previously used in PP studies (de Verdal et al. 2024; Zhu et al. 2022; Lane et al. 2020; De Salvio et al. 2024). We retained traits that were common between datasets, namely plant height (PH), thousand grain/kernel weight (GW), phenology (PHENO) and yield (YIELD). These datasets were chosen as they had phenotypic and NIRS data available for almost all genotypes in almost all environments. A data curation step was done to have datasets in which all phenotypes and all NIRS data were available for all genotypes in all environments.

2.1.1 Rice dataset

This is the dataset used for PP in de Verdal et al. (2024), details on field experiment, NIRS measurements, and genotyping can be found in Rakotoson et al. (2021, 2017); Rakotomalala et al. (2021).

A panel of 190 rice accessions representing the working collection of the upland rice breeding program conducted in Madagascar by FOFIFA (Malagasy National Research Center for Rural Development) and CIRAD (French Agricultural Research Center for International Development) was used. These genotypes were grown during two consecutive seasons (2014-2015 and 2015-2016) using split-block designs with two replications in an alpha lattice layout. Each replication consisted of 14 blocks, further subdivided into two sub-blocks for evaluating nitrogen fertilization levels (low nitrogen and high nitrogen fertilization). In this dataset, environments were considered as combinations between fertilization levels and years.

The traits used were plant height (PH), grain yield (YIELD), thousand grain weight (GW), and days to flowering (PHENO).

NIR spectra were collected on grains ground to 1 mm. Spectra were acquired for each

plot (each accession in each sub-block) and include data from 1,500 wavelengths ranging between 1,000 and 2,500 nm with a 1 nm step. Four technical repetitions were averaged to obtain the spectra of each plot.

2.1.2 Maize dataset

This dataset was used for PP in Lane et al. (2020); De Salvio et al. (2024), and described in Farfan et al. (2015) where details on field experiment, NIRS measurements, and genotyping can be found.

A set of 346 hybrid testcrosses were grown in 2011 and 2012 with two watering conditions (well watered and water stress) in two replicates with a randomized complete block design (RCBD) at the Texas A&M AgriLife Research Farm. In this dataset, environments were considered as combinations between water conditions and years.

The traits used were plant height (PH), grain yield (YIELD), 500-kernel weight (GW), and days to silking (PHENO).

NIR spectra were collected on harvested whole kernels, and consist of every wavenumber from 4,000 to 10,000 cm^{-1}, which correspond to an interval between 1,000 and 2,500 nm. Spectra were acquired for each plot (each genotype in each environment in each repetition).

2.1.3 Triticale dataset

This dataset was described and used for PP in Zhu et al. (2022) where details on field experiment, and NIRS can be found. Contrary to the previous datasets, no genotyping data was available for this one. It consists of 1280 genotypes of triticale from three different populations (one diversity panel and two doubled haploid populations). Field experiments were divided in two trials grown in four locations. One trial was evaluated during the two years (2014 and 2015), and the other during one year only (2015). In this dataset, environments were considered as combinations between locations and years.

The traits used were, plant height (PH), grain yield (YIELD), thousand grain weight (GW) and phenology (PHENO).

NIR spectra were collected on harvested grains after being dried to 14% humidity. Intensity of reflectance was measured for each wavelength in the spectral range from 1200 to 2400 nm with a step size of 1 nm. Spectra are missing for two locations in trial 1 for the year 2015.

2.2 Spectra pre-processing

Four classical pre-processing methods were used on raw spectra of all datasets: normalization by the standard normal variate method (snv), detrend (dt), first order derivative after snv (d1), and second order derivative after snv (d2). All those pre-processings were done after a smoothing of raw spectra using a Savitsky-Golay algorithm (filter length of 21, polynomial order of 1, and derivation order of 0). Derivatives were also done with the Savitsky-Golay algorithm. For d1, parameters were: filter length of 41 wavelengths, polynomial order of 2, and derivation order of 1. For d2, parameters were: filter length of 81 wavelengths, polynomial order of 3, and derivation order of 2. In the results, the best performing pre-processing was kept for each dataset, that is, d1 for the rice dataset, and d2 for the maize and triticale datasets.

2.3 Genotypic BLUEs of phenotypes and spectra

For each dataset, the genetic best linear unbiased estimators (BLUEs) of genotype effects were computed for phenotypic data and pre-processed spectra. Estimation models were built upon the ones published in their respective articles. Some adjustments were made to have genotypic values always estimated with fixed genetic effects as recommended in Holland and Piepho (2024), which was not the case in all original models. For all datasets, genotypic BLUEs were estimated within environment to have one value per genotype per environment for each phenotypic variable and each wavelength. The genotypic BLUEs were estimated with the main effect of genotypes in the

models below. Finally, Pearson correlations between BLUEs of each environment were computed.

2.3.1 Rice dataset

As in de Verdal et al. (2024), genotypic BLUEs were estimated within environments (year/nitrogen treatment combination) with the following model:

$$Y_{ijk} = \mu + G_i + R_j + \underline{B_{(j)k}} + \underline{\varepsilon_{ijk}} \tag{1}$$

Where Y_{ijk} is the observation of genotype i in block k of repetition j. G_i is the fixed effect of genotype i, R_j is the fixed effect of repetition j, $B_{(j)k} \overset{iid}{\sim} \mathcal{N}(0, \sigma_b^2)$ is the random effect of block k nested into repetition j, σ_b^2 being the block variance, $\varepsilon_{ijk} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ is the random residual, σ^2 being the residual variance.

2.3.2 Maize dataset

As in Lane et al. (2020), genotypic BLUEs were estimated within environments (year/water treatment combination). In this article a model with a genotypic and a block effect was computed. However, no block information could be found in the available data. Instead, a row, a range, and a repetition column were available. In De Salvio et al. (2024), which used the same dataset, genotypic adjusted means were computed in a model using the row, range and repetition information. Therefore, the following model was computed in each environment:

$$Y_{ijkl} = \mu + G_i + R_j + \underline{RANGE_{(j)k}} + \underline{ROW_{(j)l}} + \underline{\varepsilon_{ijkl}} \tag{2}$$

Where Y_{ijkl} is the observation of genotype i in environment j, in range k and row l of repetition j. G_i is the fixed effect of genotype i, R_j is the fixed effect of repetition j, $RANGE_{(j)k} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{range})$ is the random effect of range k nested into repetition j, σ^2_{range} being the range variance, $ROW_{(j)l} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{row})$ is the random effect of row

l nested into repetition j, σ^2_{row} being the row variance, $\varepsilon_{ijkl} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ is the random residual, σ^2 being the residual variance.

In De Salvio et al. (2024), the effect of repetition was considered random as the range and row effects. In the present study, we computed the repetition as fixed because only two repetitions per environment were characterized.

2.3.3 Triticale dataset

The triticale dataset contains three different populations. As in Zhu et al. (2022), the estimation of genotypic BLUEs was done across populations but prediction models were computed within populations. For the estimation of BLUEs Zhu et al. (2022) used both across and within environments models. For our study, we used their within environment model.

$$Y_{ijkl} = \mu + G_i + T_j + R_{(j)k} + \underline{B_{(jk)l}} + \underline{\varepsilon_{ijkl}} \tag{3}$$

Where Y_{ijkl} is the observation of genotype i in block l of repetition k in trial j, G_i is the fixed effect of genotype i, T_j is the fixed effect of trial j, $R_{(j)k}$ is the fixed effect of repetition k nested into trial j, $B_{(jk)l} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_b)$ is the random effect of block l nested into repetition k of trial j, σ^2_b being the block variance, $\varepsilon_{ijkl} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ is the random residual, σ^2 being the residual variance.

All effects were considered fixed in Zhu et al. (2022), which is not the case in the present study. The block effect is considered random because it has many levels despite not being a variable of interest. Moreover, since some environments only included one trial in Zhu et al. (2022), the model was adapted by removing the trial effect in these environments.

2.4 Splitting the G and $G \times E$ parts of spectra

As stated in the introduction, the relationship matrices used in the G and $G \times E$ effects share information. Indeed, the interaction matrix is computed either by re-using the G effect matrix as is (like in genomic prediction), or by computing a new matrix based on the same BLUEs/BLUPs of spectra already used to compute the G effect matrix. To avoid having an interaction matrix which reflects the same information as the G effect matrix, we tested models to decompose these two pieces of information separately, to then compute two matrices that use different parts of spectral information.

To do so, we used the models that had been used to decompose the variance of phenotypic data and wavelengths in the original articles. These models were computed across environments. The genetic part of spectra was estimated as the across environment main genetic BLUPs and the $G \times E$ part of spectra was estimated as the $G \times E$ BLUPs. In addition, as in the original articles, we also used these models to estimate the variance components of phenotypic records and wavelengths.

2.4.1 Rice dataset

In de Verdal et al. (2024), an across year and nitrogen treatment model was described to compute the parts of variance of each effect on phenotypes as well as on spectra. It was adapted here by considering year and nitrogen treatment combinations as environments, and by considering all effects random:

$$Y_{ijkl} = \mu + \underline{G_i} + \underline{E_j} + \underline{(G \times E)_{ij}} + \underline{R_{(j)k}} + \underline{B_{(jk)l}} + \underline{\varepsilon_{ijkl}} \tag{4}$$

Where Y_{ijkl} is the observation of genotype i in block l of repetition k of environment j. $G_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_g^2)$ is the random effect of genotype i, σ_g^2 being the genetic variance, $E_j \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$ is the random effect of environment j, σ_e^2 being the environmental variance $(G \times E)_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_{ge}^2)$ is the random effect of the interaction between genotypes and environments, σ_{ge}^2 being its associated variance, $R_{(j)k} \overset{iid}{\sim} \mathcal{N}(0, \sigma_r^2)$ is

the random effect of repetition k nested into environment j, σ_r^2 being the repetition variance, $B_{(jk)l} \overset{iid}{\sim} \mathcal{N}(0, \sigma_b^2)$ is the random effect of block l nested into repetition k in environment j, σ_b^2 being the block variance, $\varepsilon_{ijkl} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ is the random residual, σ^2 being the residual variance.

The genetic part of spectra was estimated as the BLUPs of the G effect while the interaction part of spectra was estimation as the BLUPs of the $G \times E$ effect.

2.4.2 Maize dataset

In De Salvio et al. (2024), an across environments model was described to compute the parts of variance of each effect on phenotypes as well as on spectra. The same model was used here.

$$Y_{ijklm} = \mu + \underline{G_i} + \underline{E_j} + \underline{(G \times E)_{ij}} + \underline{R_{(j)k}} + \underline{RANGE_{(jk)l}} + \underline{ROW_{(jk)m}} + \underline{\varepsilon_{ijklm}} \quad (5)$$

Where Y_{ijklm} is the observation of genotype i in environment j, in range l and row m of repetition l. $G_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_g^2)$ is the random effect of genotype i, σ_g^2 being the genetic variance, $E_j \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$ is the random effect of environment j, σ_e^2 being the environmental variance $(G \times E)_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_{ge}^2)$ is the random effect of the interaction between genotypes and environments, σ_{ge}^2 being its associated variance, $R_{(j)k} \overset{iid}{\sim} \mathcal{N}(0, \sigma_r^2)$ is the random effect of repetition k nested into environment j, σ_r^2 being the repetition variance, $RANGE_{(jk)l} \overset{iid}{\sim} \mathcal{N}(0, \sigma_{range}^2)$ is the random effect of range l nested into repetition k of environment j, σ_{range}^2 being the row variance, $ROW_{(jk)m}$ is the random effect of row m nested into repetition k and environment j, σ_{row}^2 being the row variance, $\varepsilon_{ijklm} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ is the random residual, σ^2 being the residual variance.

The genetic part of spectra was estimated as the BLUPs of the G effect while the interaction part of spectra was estimation as the BLUPs of the $G \times E$ effect.

2.4.3 Triticale dataset

In Zhu et al. (2022), an across environment estimation model was described and was used here to decompose triticale spectra but considering all effects as random instead of fixed.

$$Y_{ijklm} = \mu + \underline{G_i} + \underline{E_j} + \underline{(G \times E)_{ij}} + \underline{T_{(j)k}} + \underline{R_{(jk)l}} + \underline{B_{(jkl)m}} + \underline{\varepsilon_{ijklm}} \tag{6}$$

Where Y_{ijklm} is the observation of genotype i in block m of repetition l in trial k of environment j, $G_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_g^2)$ is the random effect of genotype i, σ_g^2 being the genetic variance, $E_j \overset{iid}{\sim} \mathcal{N}(0, \sigma_e^2)$ is the random effect of environment j, σ_e^2 being the environmental variance, $(G \times E)_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma_{ge}^2)$ is the random effect of the interaction between genotypes and environments, σ_{ge}^2 being its associated variance, $T_{(j)k} \overset{iid}{\sim} \mathcal{N}(0, \sigma_t^2)$ is the random effect of trial k nested into environment j, σ_t^2 being the trial variance, $R_{(jk)l} \overset{iid}{\sim} \mathcal{N}(0, \sigma_r^2)$ is the random effect of repetition l nested into trial k and environments j, σ_r^2 being the repetition variance, $B_{(jkl)m} \overset{iid}{\sim} \mathcal{N}(0, \sigma_b^2)$ is the random effect of block m nested into repetition l, trial k and environment j, σ_b^2 being the block variance, $\varepsilon_{ijklm} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ is the random residual, σ^2 being the residual variance.
The genetic part of spectra was estimated as the BLUPs of the G effect while the interaction part of spectra was estimation as the BLUPs of the $G \times E$ effect.

2.5 Prediction model

For genomic and phenomic prediction, the following model was used:

$$Y = \mu + Z_e\beta_e + Z_g u_g + Z_{ge} u_{ge} + \varepsilon \tag{7}$$

Where Y is the $(n_g.n_{env} \times 1)$ vector of genetic values estimated with models described in section 2.3, $u_g \sim \mathcal{N}(0, K_g\sigma_g^2)$ is the $(n_g \times 1)$ vector of random effect

of genotypes, σ_g^2 being the genetic variance, and K_g being the $(n_g \times n_g)$ relationship matrix calculated with markers or spectra, β_e is the $(n_e \times 1)$ vector of fixed environmental effects, $u_{ge} \sim \mathcal{N}(0, K_{ge}\sigma_{ge}^2)$ is the $(n_g.n_{env} \times 1)$ vector vector of the random effects of $G \times E$ interaction, σ_{ge}^2 being its associated variance, and K_{ge} the $(n_g.n_{env} \times n_g.n_{env})$ relationship matrix between genotypes in different environments calculated with markers or spectra, and $\varepsilon \sim \mathcal{N}(0, I\sigma^2)$ the $(n_g.n_{env} \times 1)$ vector of random residuals, I being the identity matrix, and σ^2 the residual variance. Z_g, Z_e, and Z_{ge} are the design matrices of the genetic, environmental, and $G \times E$ effects respectively.

Moreover, for the triticale dataset, predictions were performed within each population. Thus, trit-dh1, trit-dh2, and trit-div will refer to the predictions in the doubled haploid population one, double haploid population two, and in the diversity panel respectively.

2.6 Relationship matrix computation

A total of 20 combinations of K_g and K_{ge} matrices were tested and summarized in table 2. Among them, two used only genomic information and were therefore GP models (methods REF1 and REF2), 14 used phenomic information and were therefor PP models (methods M1 to M14) and four were combinations of genomic information to predict the G effect and phenomic information to predict de $G \times E$ effect, and were GP + PP models (methods M16 to M18). For the PP models, 12 methods used matrices previously described in the literature and two methods used matrices computed with a new method. Among GP + PP models, three meyhods used matrices previously used in PP literature and one method used matrices computed with a new method. Readers may refer to figure 1 for a better understanding.

Methods REF1 and REF2 are reference genomic prediction models. In these models, K_g was computed using the A.mat function of the sommer package (Covarrubias-Pazaran 2016): $K_g = \frac{MM'}{c}$, where M is the column centered marker matrix with alleles coded as {-1,0,1}, and c is the mean value of the diagonal elements of MM'. REF1 is a reference model without a $G \times E$ term, and in REF2, $K_{ge} = Z_g K_g Z_g^t \odot Z_e Z_e^t$, corresponds to a block diagonal matrix where K_g is repeated on each block diagonal element if data is sorted by environments ($\odot$ is the Hadamard product). This corresponds to the "rep-bdiag" matrix in figure 1.

In methods M1 to M5, $K_g = \frac{1}{n_e} \sum_{e=1}^{n_e} \frac{S_e S_e^t}{n_\lambda}$, where n_e is the number of environments, S_e are the genetic BLUEs of wavelengths for environment e and n_λ is the number of columns of S_e, that is, the number of wavelengths (Lane et al. 2020; Robert et al. 2022). This way of computing the G relationship matrix corresponds to the "within" matrix in figure 1.

In M1, no $G \times E$ effect was used, meaning that only the $Z_g u_g$ and the $Z_e \beta_e$ terms were specified in the model. In M2, $K_{ge} = Z_g K_g Z_g^t \odot Z_e Z_e^t$, which is the same method as for genomic prediction in REF2 ("rep-bdiag" in figure 1). In this matrix, there are relationships between genotypes belonging to the same environment but not between genotypes in different environments. Moreover, the relationship between two genotypes in one environment is equal to the relationship between these two genotypes in another environment. If the dataset was sorted by environments, this corresponds to a block diagonal matrix where K_g is repeated on each block diagonal element. This method was used in Galán et al. (2021); Robert et al. (2022); de Verdal et al. (2024).

In M3, $K_{ge} = Z_{ge} \frac{SS^t}{n_\lambda} Z_{ge}^t \odot Z_e Z_e^t$, where $S = (S_1, ..., S_{n_e})^t$ is the concatenation of the genetic BLUEs of wavelengths from all environments (S_1 are the genetic BLUEs from environment 1, S_{n_e} from environment n_e). This relationship matrix corresponds to the "bdiag" matrix in figure 1. It is a matrix where there is no relationships between

genotypes belonging to different environments, but the relationships between genotypes are environment-specific, meaning that the relationship between two genotypes in one environment is different from the relationship between these two genotypes in another environment. If the dataset was sorted by environment, this yields a block diagonal matrix where each block diagonal element is the relationship matrix computed using data from one environment only (for example, the first block diagonal element would be equivalent to $\frac{S_1 S_1^t}{n_\lambda}$). This method was used in Krause et al. (2019); Adak et al. (2023b); De Salvio et al. (2024).

In M4, $K_{ge} = Z_{ge} \frac{SS^t}{n_\lambda} Z_{ge}^t$ where S is defined as for M3. This corresponds to the "full" matrix in figure 1. It is a dense matrix containing relationships between genotypes from the same environments and between genotypes belonging to different environments. As in model M3, relationships between genotypes in one environment are different from the relationships between the same genotypes in another environment, moreover, relationships between genotypes from different environments exist and are unique to each genotype-environment combination. If the dataset was sorted by environment, this would yield a matrix where block diagonal elements are the same as in model M3, but there would also be all off-block-diagonal values. This method was used in Krause et al. (2019); Lane et al. (2020).

In M5, $K_{ge} = Z_g K_g Z_g^t \odot Z_e K_e Z_e^t$, where K_e is a relationship matrix between environments (see next paragraph for the computation of K_e). This corresponds to the "envsim" matrix in figure 1. It is a matrix similar to the one used in M4, that is, a dense matrix where there are relationships between genotypes from the same environments and between genotypes belonging to different environments. Also, as in M4, relationships between genotypes in one environment are different from the relationships between the same genotypes in another environment, moreover, relationships between genotypes from different environments exist and are unique to each genotype-environment combination. If the dataset was sorted by environment, this would yield

a matrix similar to the one used in M4, but relationship values would be different. This was used in Aono and Chawade (2026); Montesinos-López et al. (2024).

In Montesinos-López et al. (2024), K_e was computed using environmental variables, and in Aono and Chawade (2026), it is computed by doing a principal component analysis (PCA) on phenomic data within each environment and calculating the Pearson correlation between the first components of each PCA. In this study, no environmental data was available, so K_e was computed with the BLUPs from the environmental effects from models presented in in section 2.4. Let SE be the matrix containing the BLUPs of the environment effect for each wavelength (of dimension $(n_e \times n_\lambda)$), then $K_e = \frac{(SE)(SE)^t}{n_\lambda}$.

Methods M6 to M10 use the same relationship matrices for the $G \times E$ effect as M1 to M5 respectively, but they all use a different matrix for the G effect: $K_g = \frac{D_g D_g^t}{n_\lambda}$, where D_g is the matrix of genotypic BLUPs of wavelengths across environments obtained with models described in section 2.4. This corresponds to the matrix "across" in figure 1. This method was used in Galán et al. (2021); de Verdal et al. (2024).

Method M11 is a methods that uses the genotypic BLUPs across environments to compute the G relationship matrix, and the $G \times E$ BLUPs only to compute the relationship matrix for the $G \times E$ effect. This was done to have a non redundant information between the K_g and K_{ge} matrices. In this case, K_g is computed as in methods M6 to M10, but $K_{ge} = \frac{D_{ge} D_{ge}^t}{n_\lambda}$, where D_{ge} is the matrix of the $G \times E$ BLUPs of wavelengths obtained with models described in section 2.4. This mehod has never been used before to our knowledge. It will be called the "decomposition" method later on as the goal is to decompose the genetic and $G \times E$ parts of the spectra more appropriately.

Methods M12 to M14 do not include a G effect, as it can be found in Krause et al. (2019). This mean that only the $Z_e \beta_e$ and $Z_{ge} u_{ge}$ terms were specified in the model. Models M12, M13, and M14 use the same methods for computing K_{ge} as methods

M3, M4, and M11 respectively. Methods derived from M2 and M5 (or equally M7 and M10) could not be used here because they require the computation of a G effect.

Finally, methods M15 to M18 combine marker information for the G effect and spectral information for the $G \times E$ effect using the same relationship matrices as for REF1-2 for the G effect and M8-11 for the $G \times E$ effect.

2.7 Cross validation scenarios

Three cross validation scenarios were implemented. The observed genotypes in observed environments scenario (oGoE) which corresponds to sparse testing, the new genotypes in observed environments scenario (nGoE) where all environments are represented in the training population and new genotypes are predicted in every environment, and the observed genotypes in new environment scenario (oGnE) where all genotypes are represented in the training set, and they are predicted in a new environment. For convenience, in all datasets, only the subset of all genotypes present in all environments was used.

The training set size was calculated for each dataset as $tsize = 0.7 \times n_g \times (n_{env} - n_{out})$ where n_g is the number of genotypes in the global population available, n_{env} is the number of environments in the dataset, and n_{out} is the number of environments let out in the oGnE scenario. A too large training set size would not allow to have genotypes not represented in it, and would not allow to have repetitions of cross validation scenarios with different combination of genotypes in the training set. This formula ensures a training set size which is not too large and avoids these issues.
For the rice dataset, $n_g = 183$, $n_{env} = 4$ and $n_{out} = 1$, so $tsize = 384$. For the maize dataset, $n_g = 143$, $n_{env} = 4$ and $n_{out} = 1$, so $tsize = 300$. For the triticale dataset, three populations are present, two doubled haploid populations (trit-dh1 and trit-dh2), and one diversity panel (trit-div). For this dataset, predictions were performed within populations. Moreover, for this dataset, traits were not phenotyped in

Table 2 Summary of all combinations of relationship matrices tested. M = column centered marker matrix with alleles coded as {-1,0,1}, c = mean value of the diagonal elements of MM', S_e = BLUEs of wavelengths for environment e, $S = (S_1, ..., S_{n_e})^t$, n_e = number of environments, D_g = genotypic BLUPs of wavelengths across environments, $D_{ge} = G \times E$ BLUPs of wavelengths, n_λ = number of wavelengths of S_e, S, or D, $\odot$ = Hadamard product, Z_g = design matrix for the G effect, Z_{ge} design matrix for the $(G \times E)$ effect, Z_e = design matrix for the E effect, K_g = relationship matrix used for the G effect, K_{ge} = relationship matrix used for the $(G \times E)$ effect, K_e = relationship matrix used for the E effect, "\" = no G or no $(G \times E)$ effect in the model.

K_g	K_{ge}	**Method**
$\frac{MM'}{c}$	\	REF1
	$Z_g K_g Z_g^t \odot Z_e Z_e^t$	REF2
$\frac{1}{n_e}\sum_{e=1}^{n_e} \frac{S_e S_e^t}{n_\lambda}$	\	M1
	$Z_g K_g Z_g^t \odot Z_e Z_e^t$	M2
	$Z_{ge}\frac{SS^t}{n_\lambda} Z_{ge}^t \odot Z_e Z_e^t$	M3
	$Z_{ge}\frac{SS^t}{n_\lambda} Z_{ge}^t$	M4
	$Z_g K_g Z_g^t \odot Z_e K_e Z_e^t$	M5
$\frac{D_g D_g^t}{n_\lambda}$	\	M6
	$Z_g K_g Z_g^t \odot Z_e Z_e^t$	M7
	$Z_{ge}\frac{SS^t}{n_\lambda} Z_{ge}^t \odot Z_e Z_e^t$	M8
	$Z_{ge}\frac{SS^t}{n_\lambda} Z_{ge}^t$	M9
	$Z_g K_g Z_g^t \odot Z_e K_e Z_e^t$	M10
$\frac{D_g D_g^t}{n_\lambda}$	$\frac{D_{ge} D_{ge}^t}{n_\lambda}$	M11
\	$Z_{ge}\frac{SS^t}{n_\lambda} Z_{ge}^t \odot Z_e Z_e^t$	M12
	$Z_{ge}\frac{SS^t}{n_\lambda} Z_{ge}^t$	M13
	$\frac{D_{ge} D_{ge}^t}{n_\lambda}$	M14
$\frac{MM'}{c}$	$Z_{ge}\frac{SS^t}{n_\lambda} Z_{ge}^t \odot Z_e Z_e^t$	M15
	$Z_{ge}\frac{SS^t}{n_\lambda} Z_{ge}^t$	M16
	$Z_g K_g Z_g^t \odot Z_e K_e Z_e^t$	M17
	$\frac{D_{ge} D_{ge}^t}{n_\lambda}$	M18

all environments, and the populations were not always characterized in the same environments. Thus, for trit-dh1, $n_g = 177$, $n_{env} = 5$ and $n_{out} = 1$ so $tsize = 496$. For trit-dh2, $n_g = 176$, $n_{env} = 5$ and $n_{out} = 1$ so $tsize = 493$. For trit-div, all traits and all genotypes were not phenotyped in the same number of environments. There were between 714 and 777 genotypes phenotyped in either seven or eight environments, and two environments were let out for the oGnE scenario. Thus $tsize = 2495$ for PH, $tsize = 2720$ for GW, $tsize = 3150$ for YIELD, and $tsize = 3204$ for PHENO.

We did not implement the cross validation scenario where new genotypes are predicted in new environments (nGnE), because this scenario poses difficulties. Indeed, for methods M1 to M5, one similarity matrix per environment has to be computed to then average them. This is possible only if spectra of all genotypes in all environments are available. This is the case in all the tested scenarios but would not in a nGnE scenario where the matrix for the new environment to predict would not have any common genotypes with the matrices for the environments in the training set. Moreover, for all PP methods, this scenario would be possible only if check genotypes were present in all environments as the genotypic BLUEs or BLUPs of spectra must be estimated. Moreover, for M11, M14 and M18 this would be possible only if there were more than two environments in the testing sets (which is the case only for the triticale dataset) as $G \times E$ BLUPs have to be estimated.

Furthermore, in oGoE, nGoE or oGnE scenarios, all genotypes in all environments are present in all folds, meaning that one estimation of BLUEs/BLUPs is needed for all folds. In the nGnE scenario, each fold has a different combination of unconnected genotypes and environments. This means that to mimic a real breeding situation, BLUEs/BLUPs would have to be re-estimated for each cross validation fold which would demand more time and computation power than available to us.

2.8 Comparing methods

Methods were compared based on the Spearman correlation coefficients between the predicted values and the BLUEs estimated within each environment. Within each dataset/CV scenario/trait combination, a pairwise Wilcoxon rank test was performed to compare methods. A Bonferroni correction was applied on p-values to account for multiple testing, and the tests were paired to account for the fact that all methods were tested on the same cross validation folds. For the sake of comparing the present results to the ones presented in the articles from which the datasets originate, Pearson correlations were also computed.

2.9 Software and packages

All analyses were conducted on the R software (R Core Team 2025). Genetic value estimation models were computed using lme4 (Bates et al. 2015), prediction models were computed using sommer (Covarrubias-Pazaran 2016), spectra pre-processing were performed using rchemo (Brandolini-Bunlon et al. 2023), graphics were produced using ggplot2 (Wickham 2016), and general data handling steps (*i.e.*, data curation) were done using tidyverse (Wickham et al. 2019).

3 Results

3.1 Variance components of traits and spectra

The data quality analysis and descriptive statistics of traits can be found in the articles from where the datasets originate. Still, we recomputed the variance components of traits in online resource 1. Genotypic variances are moderate to high in the triticale and rice datasets (from 8% to 81% of total phenotypic variance), and low in the maize dataset (from 7% to 13% of total phenotypic variance variance). Variances due to the interactions between genotypes and environments were generally low, ranging

from 0.8% to 9% of total phenotypic variances. Traits had different values of the ratio between the $G \times E$ variance and the genotypic variance ranging from 0.05 to 0.53. YIELD was the trait with the highest ratio in all datasets. Otherwise, the traits having the highest $G \times E$ variance or the highest/lowest $G \times E/G$ variance were not the same across datasets.

Concerning spectra, their variance components are presented in online resource 2. In the maize dataset, spectra had the highest part of genetic variance, but low $G \times E$ variance compared to the genetic and environmental variances. In the rice dataset, most of the variance was unexplained by genetic or environmental factors and stayed in the residual. Specifically, the amount of $G \times E$ variance was close to zero in this dataset. Finally, the triticale dataset was the one with the highest part of variance due to $G \times E$. The part of variance explained by genotypes was lower than that explained by $G \times E$, but the most important part of variance in this dataset was environmental.

3.2 General comparison between combination of reference methods (REF1 to M10)

Comparisons between methods REF1, REF2, and M1 to M10 for CV scenarios oGoE, nGoE, and oGnE are available in figures 2, 3, and online resource 3 respectively. These methods use relationship matrices that were already used in the PP literature but represent all possible combinations of these different matrices. Results from the oGnE scenario were placed in supplementary material as they were very similar to results from the oGoE scenario.

For the rice and maize datasets, for which genomic data were available, GP outperformed PP in most cases. The only exceptions were for GW in the maize dataset in which a PP model was the best performing in oGoE and nGoE scenarios.

The results of PP show that the main driver of PA was the similarity matrix used in the G term. Indeed, the largest differences between methods were found between

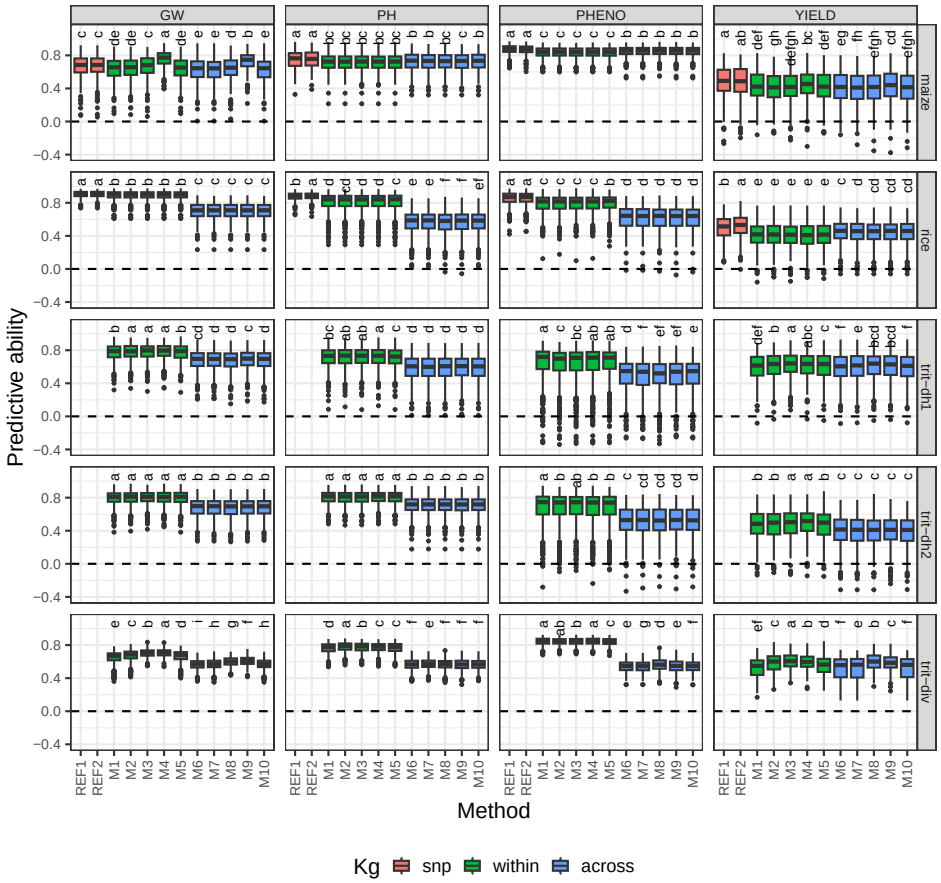


Fig. 2 Comparison between methods REF1 to M10 in the oGoE scenario. Letters indicate groups of non-significantly different methods. GW = grain weight, PH = plant height, PHENO = phenology trait. snp = genomic prediction method, within = phenomic prediction with the G relationship matrix being calculated as the cell by cell average of environment specific genetic relationship matrices which are calculated with within environment genetic BLUEs of spectra, across = phenomic prediction with the G relationship matrix being calculated using across environment BLUEs of spectra.

the M1-5 group which all share the "within" G effect matrix, and the M6-10 group which all share the "across" G effect matrix. In the oGoE and oGnE scenarios, the "within" genetic matrix (methods M1-5) reached better PA than the "across" genetic matrix (methods M6-10). On the contrary, for the nGoE scenario, the "across" matrix reached better PA than the "within" matrix.

To compare the performances of each $G \times E$ matrix, the methods were compared by groups of methods sharing the same G effect matrix. In each group the reference

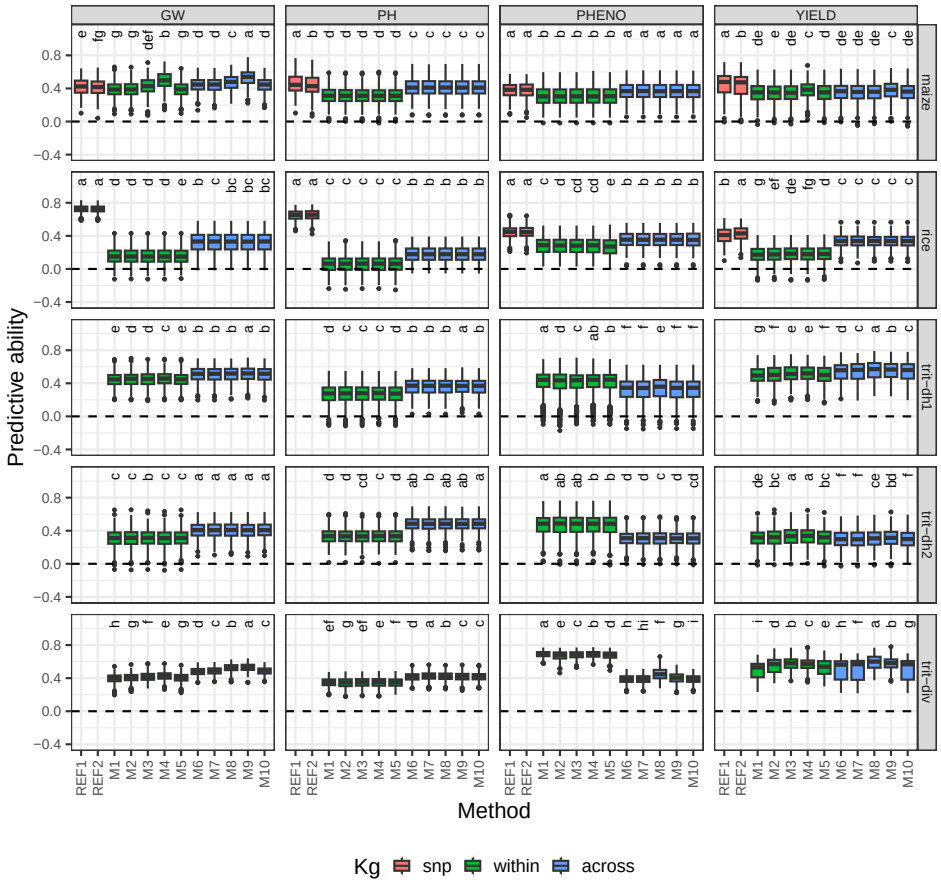


Fig. 3 Comparison between methods REF1 to M10 in the nGoE scenario. Letters indicate groups of non-significantly different methods. GW = grain weight, PH = plant height, PHENO = phenology trait. snp = genomic prediction method, within = phenomic prediction with the G relationship matrix being calculated as the cell by cell average of environment specific genetic relationship matrices which are calculated with within environment genetic BLUEs of spectra, across = phenomic prediction with the G relationship matrix being calculated using across environment BLUEs of spectra.

method to be compared to was the method with no $G \times E$ effect (REF1, M1, and M6), to see how much each $G \times E$ matrix could improve the PA. Across dataset/trait/CV scenario combinations, the number of times each method had a PA significantly higher, lower, or not significantly different from the reference of the group was counted. Moreover, in the cases when there were significant differences between methods, the number of times each method had the best predictive ability within its group was also counted (if multiple methods had the best PA and were not significantly different, they were

all granted a point). Results are summarized in table 3. Methods were compared on whether they had a high/low number of occurrences having the best PA, a high/low number of occurrences having PA higher, equal or lower than the reference method. For GP methods, models with and without a $G \times E$ effect performed similarly: REF2 was significantly higher than REF1 five times, significantly lower five times, not different 14 times, and had the best result six times while REF1 had the best results four times.

In the M1-5 group, the worst performing method was M5 with the lowest number of best performances and the highest number occurrences where its PA were not different, or lower than M1. The best performing methods were M4 having the highest number of best performances and M2 having the highest number of PA higher than M1.

For the M6-10 group, the worst performing method was M10 with the lowest number of best performances and the highest number of PA lower or not different from M6. The best method was M9 with the highest number of best performances, and of PA higher than M6. However, M9 also had the highest number of PA lower than M6.

Significant differences are to be replaced in the context of PA gain/loss. Indeed, statistical tests did find significant differences between methods even when these differences were negligible, which is typically the case for PH in the rice dataset in the oGoE scenario for example. In the tested datasets, including a $G \times E$ matrix in prediction models, whether it was for GP or PP, had little impact on predictive abilities. Indeed, methods REF1, M1 and M6 had no $G \times E$ relationship matrix and often performed as well as methods REF2, M2-5, and M7-10 respectively even if significant differences were found.

Looking back at figures 2, 3, and online resource 3, the two methods that stood out independently from the CV scenario were M4 and M9, except for PHENO and YIELD in the trit-div dataset for which results were variable across CV scenario and for which M3 and M8 also stood out to a lesser extent. M4 and M9 both use the same

Table 3 Number of times each method outperformed its reference and have the best predictive ability across traits, datasets and cross validation scenarios. Results arise from a pairwise paired Wilcoxon test between methods. Methods were compared in groups: REF1 and REF2 together, M1 to M5 together and M6 to M10 together. The "Method" column indicates the methods that is considered. The "Best" column indicates the number of times the method had the best predictive ability within its group in the cases when there were significant differences between methods (if multiple methods had the best PA and were not significantly different they were all granted a point). The "Higher", "No difference" and "Lower" columns indicate the number of times the method had a predictive ability significantly higher, not significantly different, or significantly lower compared to the method stated in the "Reference" column. Percentages are calculated within each method relatively to the total number of times each method was tested (which is 24 for REF1 and REF2, and 60 for M1 to M10).

Reference	Method	Best	Higher	No Difference	Lower
REF1	REF1	4 (16.7%)	\	\	\
	REF2	6 (25%)	5 (20.8%)	14 (58.3%)	5 (20.8%)
M1	M1	10 (16.7%)	\	\	\
	M2	19 (31.7%)	26 (43.3%)	26 (43.3%)	8 (13.3%)
	M3	24 (40%)	17 (28.3%)	33 (55%)	10 (16.7%)
	M4	29 (48.3%)	23 (38.3%)	28 (46.7%)	9 (15%)
	M5	11 (18.3%)	16 (26.7%)	34 (56.7%)	10 (16.7%)
M6	M6	14 (23.3%)	\	\	\
	M7	17 (28.3%)	14 (23.3%)	40 (66.7%)	6 (10%)
	M8	18 (30%)	16 (26.7%)	35 (58.3%)	9 (15%)
	M9	21 (35%)	17 (28.3%)	30 (50%)	13 (21.7%)
	M10	13 (21.7%)	5 (8.3%)	49 (81.7%)	6 (10%)

$G \times E$ matrix: $K_{ge} = Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$ where S contains the concatenation of all within environment BLUEs of spectra. This corresponds to the "full" matrix in figure 1.

3.3 Predictive abilities obtained for the decomposition method (M11) and methods with no main genetic effect (M12-M14)

Online resource 4 shows the results of the "decomposition" method M11 compared to methods M4 and M9 which were found to be the best options to include GxE effects in the predictive model. Online resource 5 shows the number of times M11 had significantly higher or significantly lower predictive abilities than both M4 and M9. Together, these results show that there were few differences between M11 and M4 or M9. More specifically, M11 and M9 behaved similarly as they share the same G effect

matrix. The only cases where M11 really outperformed M4 and M9 was for GW in the maize dataset where it had slightly higher PA.

Online resource 6 shows the results of methods without a G effect, M12 to M14 compared to M4 and M9. Online resource 5 shows the number of times each method had significantly higher or lower PA than both M4 and M9. These methods consistently produced lower PA than references methods. Not including a G effect in models led to poor performances.

3.4 Methods combining genomic and phenomic information to compute the G and $G \times E$ matrices

Results of methods M15 to M18 combining genomic information in the G effect, and phenomic information in the $G \times E$ effect are shown in figure 4. Only the rice and the maize datasets had available genotyping data. Online resource 7 shows the number of times each method had significantly higher, not different or lower PA than REF2, and the number of times each method had the best predictive ability when significant differences existed (if multiple methods had the best PA and were not significantly different, they were all granted a point). Looking only at Online resource 7, M17 was the best method as it had the highest number of best performances, and the highest number of times having significantly higher PA than REF2, and lowest numbers of times having a significantly lower or not significant PA than REF2. Nevertheless, as for previous results, many significant differences were found while the difference in PA between methods seemed negligible. Looking at figure 4, M15, M16 and M18 did substantially outperform REF2 for GW in the maize dataset in all CV scenarios, except M15 for the oGnE scenario. Likewise, methods M16 and M18 also outperformed REF2 for YIELD in the maize dataset in all CV scenarios, and M17 outperformed REF2 for PHENO and YIELD in the rice dataset in the oGoE and oGnE scenarios.

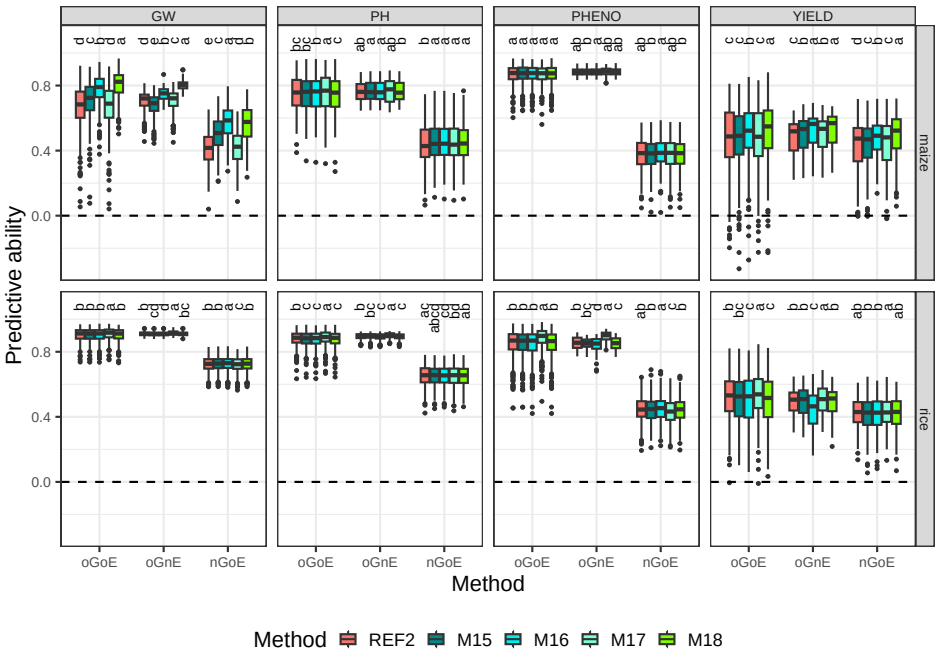


Fig. 4 Comparison between genomic prediction and methods combining genomic and phenomic information. Letters indicate groups of non-significantly different methods. GW = grain weight, PH = plant height, PHENO = phenology trait.

3.5 Comparison to results published using the same datasets

Finally, our results can be compared to the ones previously obtained (Lane et al. 2020; De Salvio et al. 2024; de Verdal et al. 2024) except for the study of Zhu et al. (2022) for which models with $G \times E$ were not implemented. For the maize dataset, Lane et al. (2020) implemented models corresponding to M4 in oGoE, M4 in oGnE, M1 in nGoE (called M1 in CV2, M1 in CV0, and M3 in CV1 respectively in their study), to predict YIELD and used Pearson correlations to evaluate their performances. For equivalent models and correlation calculation methods (results not shown), our PA were in comparable ranges.

For the same dataset, De Salvio et al. (2024) used models that have no equivalent in our study and used a comparison metric taking the number of individuals in the testing set into account. Using the same metric as theirs, their models outperformed

ours. They also reported better PA for models combining genomic and phenomic information, as was the case in the present study.

For the rice dataset, de Verdal et al. (2024) used models which correspond to methods REF1 and M6, and to methods REF2 and M7 (called MM and MDs respectively in their study). They also implemented models combining genomic and phenomic information that are not comparable to ours as they used the genomic and phenomic information to predict both the G and GxE matrices. Although their cross-validation schemes were very different, our PA are in comparable ranges. More variability was found between models with or without a $G \times E$ term in de Verdal et al. (2024) than in our study, even though the reported differences were low. They also reported low differences between genomic-only, phenomic-only, and models combining genomic and phenomic information which was not the case in our study.

4 Discussion

In this study, we tested existing methods to compute relationship matrices based on phenomic information for the prediction of $G \times E$ interactions. A total of 20 methods were tested: two reference genomic prediction models, 10 combinations of different phenomic relationship matrices for the G and $G \times E$ terms, one new method intended to better isolate the genetic and $G \times E$ information in their respective phenomic relationship matrices, three methods omitting the G effect, and four methods combining genomic information for the G effect and phenomic information for the $G \times E$ effect. This benchmark was done on three datasets from the PP literature (Zhu et al. 2022; Lane et al. 2020; De Salvio et al. 2024; de Verdal et al. 2024) for which phenotypes and NIRS acquisition were available for all genotypes in all environments. PA reported in our study were generally in the same range as PA reported in the original articles for equivalent models and PA computation methods.

In our results, GP generally outperformed PP. Differences between models including

a $G \times E$ effect and models with only a G effect were small, and the main driver of PA was the matrix used in the G effect. Depending on the cross-validation scenario, different G effect matrices had the best PA, and the best $G \times E$ matrix was generally a dense matrix with environment specific relationship between genotypes, within and across environments.

4.1 Which relationship matrices should be used for phenomic prediction in a multi-environment context

Concerning the phenomic relationship matrix to be used for the G effect, results were clear: when predicting already observed genotypes in a new environment (oGnE), or in a sparse testing context (oGoE), using the "within" matrix (mean of all environment specific matrices, $\frac{1}{n_e}\sum_{e=1}^{n_e}\frac{S_e S_e^t}{n_\lambda}$) produced better results. Conversely, when predicting new genotypes in known environments (nGoE), using the "across" matrix computed using across environments BLUPs of spectra produced better results ($\frac{D_g D_g^t}{n_\lambda}$). As the G effect matrix was the main driver of predictive abilities, we can conclude that the choice of this matrix is important and should be made according to the prediction scenario.

Concerning the phenomic relationship matrix to be used for the $G \times E$ effect in phenomic-only models, results were less clear. However, in the cases where differences between methods were obvious, one matrix did stand out: the "full" matrix ($Z_{ge}\frac{SS^t}{n_\lambda}Z_{ge}^t$). This is a dense matrix where two genotypes in one environment will have a different relationship from the two same genotypes in another environment, and two genotypes from different environments will have a relationship value too. However, differences between methods with or without $G \times E$ effects were generally small even for reference methods. This may underline some limitations of the datasets used in this study. Thus, this claim is not strongly supported by results and should be confirmed in further studies.

4.2 Combining genomic and phenomic information in prediction models can increase PA

Considering overall performances of all methods, the best PA were produced by methods using genomic information to compute the G effect matrix, and phenomic information to compute the $G \times E$ matrix. Phenomic information can be a complementary source of information to genomic information and help predict $G \times E$ interactions in some cases where genomic information alone fails to do so. It is noticeable that the phenomic $G \times E$ matrices that complement marker data the best were not necessarily the ones that performed best in phenomic only models. In particular, the "decomposition" ($\frac{D_{ge}D_{ge}^t}{n_\lambda}$) and the "envsim" ($Z_g K_g Z_g^t \odot Z_e K_e Z_e^t$) matrices reached the highest PA in more cases than in phenomic only models, in which they did not increase PA most of the time.

In previous PP studies, many models combining phenomic and genomic information were tested, but they were not fairly comparable to ours as they did not use genomic information to predict only the G effect and phenomic only to predict the $G \times E$ effect (Adak et al. 2023b; Jackson et al. 2023; Klein et al. 2026; Togninalli et al. 2023; Adak et al. 2024; Cuevas et al. 2019). Among these studies, only Togninalli et al. (2023); Klein et al. (2026) found substantial improvement in PA with models combining genomic and phenomic information. Krause et al. (2019) used models combining phenomic and genomic information in that same way that we did in our study. They also found that predicting the G effect with genomic information and the $G \times E$ effect with phenomic information generally had better PA than genomic-only or phenomic-only models.

4.3 Little effect of $G \times E$ term in phenomic-only prediction models

One major trend in our results is that in most cases, adding a $G \times E$ term in models did not improve predictive ability compared to models without a $G \times E$ term. Few studies have tested models with or without a $G \times E$ term in equivalent cross-validation scenarios. (Robert et al. 2022) found that adding a $G \times E$ term to the model did improve PA over models without $G \times E$ term, and (Cuevas et al. 2019; de Verdal et al. 2024) that multi environment models had better PA than single environment models.

One hypothesis that could explain these results is the absence/low impact of a $G \times E$ effect on phenotypic records and spectra. Considering the rice dataset, de Verdal et al. (2024) showed that within environment BLUEs of the different traits were highly correlated to one another except for yield. In addition, the variance explained by interactions between genotypic and environmental factors (year and nitrogen treatment) was generally small compared to the main genetic variance except for yield. These results are in accordance with our own estimation of variance components (online resource 1). However, even if PA of YIELD for this dataset did seem to exhibit slightly more variation between methods, they did not behave much differently from PA of other traits for this same dataset.

Considering the triticale dataset, Zhu et al. (2022) reported a substantial $G \times E$ variance compared to the G variance in spectra. Our estimations of variance components for spectra are in accordance with this result. Our estimations of variance components for the different traits showed generally higher ratios between the $G \times E$ variance and the G variance than for the other two datasets. Accordingly, the PA from the diversity panel of this dataset (trit-div) were more variable across methods than for other datasets even though differences between methods remained low. To a lesser extent, this also holds for the two double haploid populations.

Considering the maize dataset, De Salvio et al. (2024) reported that traits and spectra displayed small variances of interaction between genotypes and environment. In our estimations, the two traits that had the highest ratio between the $G \times E$ and G variance were YIELD and GW, which were the two traits with the most differences between methods in this dataset. Spectra also displayed little $G \times E$ variance in our estimations.

Finally, online resource 8 shows the correlation between BLUEs estimated in different environments for each dataset. Correlations can be as low as 0.17 in the triticale dataset, and, as expected they are generally lower for traits with a high ratio between $G \times E$ and G variance. Altogether, these results show that a slight link between the ratio of $G \times E$ and G variances, the correlation between BLUEs estimated in different environments, and the differences of performance between methods, seems to exist. However, attempts to statistically demonstrate this link failed (results not shown) probably due to the high variability of responses between datasets, traits, and cross validation models. Thus, it means that the tested methods were generally not able to correctly predict the environment specific response of genotypes in the tested datasets. However, despite having small amounts of $G \times E$ variance in spectra, models using genomic information for the G effect and spectra for the $G \times E$ effect had the best PA overall. This indicates that spectra could help improve PA of GP models and that they do contain information that is complementary to pure additive genetic signal.

It would be interesting to test these methods on datasets with larger $G \times E$ variances in phenotypic traits. Indeed, in the literature larger levels of GxE have been reported in several studies. For example, on a wheat dataset, Lado et al. (2016) found a $G \times E$ variance of 15.5 % of the total phenotypic and a ratio of 2.7 between the $G \times E$ variance and the main genetic variance. Likewise, on a rice dataset, Rebollo et al. (2023) found a $G \times E$ variance of 12.84 % and a ratio of 8.56. However, such datasets do not contain phenomic data, and we are not aware of any available dataset containing

phenomic data and this amount of $G \times E$ variance in phenotypic data. It would also be interesting to compare methods on simulated data with different known levels of $G \times E$ variance. This may be possible thanks to recent frameworks designed to simulate such phenotypic data (Bančič et al. 2024), or to simulate artificial NIR spectra (Beć and Huck 2019).

Another hypothesis that could explain the few differences between models including or not a $G \times E$ effect is that the G effect matrix and the $G \times E$ matrix contain redundant information, and the addition of a $G \times E$ matrix does not really add useful information to predict environment specific BLUEs. For GP this is likely as markers are not environment specific and the $G \times E$ relationship matrix is a repetition of the G effect matrix on block diagonal elements of a larger matrix.

In PP, two elements can corroborate this hypothesis. The first element is that models combining genomic information for the G effect and phenomic information for the $G \times E$ effect had the best performances overall. These models use different sources of information for the two effects, which are likely non-redundant. The second element is that the "full" matrix stood out in phenomic-only models. More specifically, in cases where significant differences between models occurred, models using the "full" matrix had better performances than models using block diagonal matrices ("bdiag" and "rep-bdiag"). These matrices are by construction redundant with the G effect matrix as they re-use the G effect matrix directly ("rep-bdiag") or are composed of intermediate environment specific relationship matrices that are used to compute the G effect matrix ("bdiag"). Moreover, computing block diagonal matrices means that genotypes from different environments are assumed to be independent (the covariances in off-block-diagonal elements is 0). Thus, it does not consider the fact that $G \times E$ interactions can make genotypes more or less related in different environments and the fact that genotypes in similar environments will be more related than genotypes

in environments that differ. This seems ill-advised in the context of PP in which phenomic records theoretically contain environment specific information that can be used to retrieve relationships between genotypes across environments. The "full" matrix, although its block diagonal elements are equal to the ones of the "bdiag" matrix, assumes that genotypes covary across environments and adds information that is not present in the G effect matrix.

However, the "envsim" and "decomp" matrices are also dense and assume relationships between genotypes across environments, but models including them did not outperform models using block diagonal matrices. This is especially disappointing concerning the "decomp" matrix and model M11 which was supposed to properly split the G and $G \times E$ information into their respective matrices. In the "envsim" matrix, environmental BLUPs are used to compute an environmental relationship matrix which is then used in the $G \times E$ matrix computation. For model M11, genotypic BLUPs are used to compute the G effect matrix, and $G \times E$ BLUPs are used to compute the $G \times E$ matrix. Thus, it is possible that BLUPs estimation models were unable to correctly estimate, extract, and differentiate the G, E, and $G \times E$ information in wavelengths. Indeed, wavelengths can be noisy and very correlated to each other, and it could be interesting to use meaningful subsets of wavelengths such as principal components to extract genetic and environmental information from spectra. It also could be interesting to explore matrix orthogonalization methods (Roger and Boulet 2018) to construct G and $G \times E$ matrices that have no collinearity.

Finally, even for the "full" matrix or models combining genomic and phenomic information, most improvements in PA compared to other methods were modest. This could be due to a sub-optimal way of computing matrices that inaccurately represents relationships between genotypes within or across environments. For example, Cuevas et al. (2019) showed that using a Gaussian or a nonlinear arc-cosine kernel instead of the classic relationship matrix could improve PP results.

It is also noticeable that cases where differences between models were observed correspond in most cases to the YIELD and GW traits. Although YIELD had the highest $G \times E$ variance in all datasets, GW had variable $G \times E$ variances across datasets. Moreover, spectra were collected on grain material in all datasets. Thus, it is possible that the tissue of NIRS acquisition played a role in the results as the traits that exhibited the most variance between methods were related to grains. NIRS of grain may contain $G \times E$ information that is most useful to predict $G \times E$ interactions happening during the reproductive stage but not the vegetative stage which is a main driver of the variability observed for PH and PHENO. It is also likely that the stage of NIRS acquisition makes a difference on the information contained in spectra and that this information could be useful to predict $G \times E$ interactions for a specific trait. Indeed, some studies showed that different pieces of information can be captured by multi-spectral sensors mounted on UAV/ UGV at different dates, through correlation of vegetation indices throughout a growing season (Washburn et al. 2024), or by modeling the effect of the flight dates on vegetation indices (Adak et al. 2023a). Some studies also showed that different dates of acquisition achieved different predictive abilities (Adak et al. 2022; Washburn et al. 2024; Kaushal et al. 2024; Sleper et al. 2025; Shafiee et al. 2024) and that combining data from different acquisition dates can marginally improve prediction results compared to single time point acquisitions (Washburn et al. 2024).

5 Conclusion and perspectives

In this study, we exhaustively tested different relationship matrices for modeling the G and $G \times E$ effects in phenomic prediction models. In our results, adding a $G \times E$ matrix generally had a very low impact on prediction results compared to models without a $G \times E$ term. This is likely due to a high redundancy in the information carried by the G and $G \times E$ matrices and to the low amounts of $G \times E$ variance in

traits of the tested datasets. Still, we showed that having complementary sources of information to compute the G and $G \times E$ matrices was beneficial as models combining genomic information for the genetic effect and phenomic information for the $G \times E$ effect had the highest predictive abilities. We also showed that the best G effect relationship matrix was dependent on the cross-validation scenario. Finally, we showed that the best phenomic derived $G \times E$ matrix was generally a dense matrix in which relationships between genotypes are environment specific, and relationships between genotypes in different environments exist. It would be interesting to make such analyses on datasets in which traits and spectra have more $G \times E$ variance than the ones of this study, or on simulated data with different known levels of $G \times E$ variance in traits.

Declarations

Author contributions

All authors contributed to the conceptualization and investigation. Data curation, formal analysis, methodology and the writing of the original draft were performed by Clément Bienvenu. Funding acquisition, supervision and validation were done by Hugues de Verdal, Vincent Segura and David Pot. All authors contributed to the review and editing of the manuscript. All authors have read and approved the final manuscript.

Acknowledgment

Authors would like to thank Tobias Würschum for making the triticale dataset available, and Aaron J. DeSalvio for sharing the maize dataset on Git.

Funding

The work of Clément Bienvenu was supported by a doctoral allowance from the French Ministry of Higher Education and Research.

Conflict of interest statement

The authors have no relevant financial or non-financial interests to disclose.

Data availability statement

Codes are available on Gitlab : https://gitlab.cirad.fr/agap/giv/predicting-gxe-through-phenomic-prediction-a-benchmark-study-across-three-species/-/tree/90baf3709cdb6e7ae87ab2b0c359fc4a2c86b583/.

All data used in this study are available online or upon request:

- Rice: All files are available from the FigShare database (https://doi.org/10.6084/m9.figshare.27282459.v1).
- Maize: All files are available on GitHub (https://github.com/ajdesalvio/Maize-NIRS-GBS
- Triticale: See Zhu et al. (2022)

ORCID

Clément Bienvenu: https://orcid.org/0009-0004-9234-6216

David Pot: https://orcid.org/0000-0001-6144-8448

Hugues de Verdal: https://orcid.org/0000-0002-1923-8575

Vincent Segura: https://orcid.org/0000-0003-1860-2256

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

References

Adak, A., Anderson, S.L., Murray, S.C.: Pedigree-management-flight interaction for temporal phenotype analysis and temporal phenomic prediction. The Plant Phenome Journal **6**(1), 20057 (2023) https://doi.org/10.1002/ppj2.20057

Aono, A.H., Chawade, A.: Phenomics-assisted sparse testing for potato breeding. Scientific Reports **16**(1), 22033 (2026) https://doi.org/10.1038/s41598-026-59202-6

Adak, A., DeSalvio, A.J., Arik, M.A., Murray, S.C.: Field-based high-throughput phenotyping enhances phenomic and genomic predictions for grain yield and plant height across years in maize. G3 Genes|Genomes|Genetics **14**(7), 092 (2024) https://doi.org/10.1093/g3journal/jkae092

Adak, A., Kang, M., Anderson, S.L., Murray, S.C., Jarquin, D., Wong, R.K.W., Katzfuß, M.: Phenomic data-driven biological prediction of maize through field-based high-throughput phenotyping integration with genomic data. Journal of Experimental Botany **74**(17), 5307–5326 (2023) https://doi.org/10.1093/jxb/erad216

Adak, A., Murray, S.C., Anderson, S.L.: Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. G3 Genes|Genomes|Genetics, 294 (2022) https://doi.org/10.1093/g3journal/jkac294

Bančič, J., Gorjanc, G., Tolhurst, D.J.: A framework for simulating genotype-by-environment interaction using multiplicative models. Theoretical and Applied Genetics **137**(8), 197 (2024) https://doi.org/10.1007/s00122-024-04644-7

Beć, K.B., Huck, C.W.: Breakthrough Potential in Near-Infrared Spectroscopy: Spectra Simulation. A Review of Recent Developments. Frontiers in Chemistry **7** (2019) https://doi.org/10.3389/fchem.2019.00048

957 Brandolini-Bunlon, M., Jallais, M., Roger, J.M., Lesnoff, M.: R package rchemo:
958 Dimension Reduction, Regression and Discrimination for Chemometrics. (2023)
959 https://doi.org/https://github.com/ChemHouse-group/rchemo

960 Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models
961 using lme4. Journal of Statistical Software **67**(1), 1–48 (2015) https://doi.org/10.
962 18637/jss.v067.i01

963 Cuevas, J., Montesinos-López, O., Juliana, P., Guzmán, C., Pérez-Rodríguez, P.,
964 González-Bucio, J., Burgueño, J., Montesinos-López, A., Crossa, J.: Deep Kernel
965 for Genomic and Near Infrared Predictions in Multi-environment Breeding Tri-
966 als. G3 Genes|Genomes|Genetics **9**(9), 2913–2924 (2019) https://doi.org/10.1534/
967 g3.119.400493

968 Covarrubias-Pazaran, G.: Genome assisted prediction of quantitative traits using the
969 R package sommer. PLoS ONE **11**, 1–15 (2016)

970 De Salvio, A.J., Adak, A., Murray, S.C., Jarquín, D., Winans, N.D., Crozier, D.,
971 Rooney, W.L.: Near-infrared reflectance spectroscopy phenomic prediction can per-
972 form similarly to genomic prediction of maize agronomic traits across environments.
973 The Plant Genome **17**(2), 20454 (2024) https://doi.org/10.1002/tpg2.20454

974 de Verdal, H., Segura, V., Pot, D., Salas, N., Garin, V., Rakotoson, T., Raboin, L.-M.,
975 VomBrocke, K., Dusserre, J., Pacheco, S.A.C., Grenier, C.: Performance of phenomic
976 selection in rice: Effects of population size and genotype-environment interactions
977 on predictive ability. PLOS ONE **19**(12), 0309502 (2024) https://doi.org/10.1371/
978 journal.pone.0309502

979 Farfan, I.D.B., Fuente, G.N.D.L., Murray, S.C., Isakeit, T., Huang, P.-C., Warbur-
980 ton, M., Williams, P., Windham, G.L., Kolomiets, M.: Genome Wide Association

Study for Drought, Aflatoxin Resistance, and Important Agronomic Traits of Maize Hybrids in the Sub-Tropics. PLOS ONE **10**(2), 0117737 (2015) https://doi.org/10.1371/journal.pone.0117737

Fumia, N., Nair, R., Lin, Y.-P., Lee, C.-R., Chen, H.-W., von Wettberg, E.B., Kantar, M., Schafleitner, R.: Leveraging genomics and phenomics to accelerate improvement in mungbean: A case study in how to go from GWAS to selection. The Plant Phenome Journal **6**(1), 20088 (2023) https://doi.org/10.1002/ppj2.20088

Galán, R.J., Bernal-Vasquez, A.-M., Jebsen, C., Piepho, H.-P., Thorwarth, P., Steffan, P., Gordillo, A., Miedaner, T.: Early prediction of biomass in hybrid rye based on hyperspectral data surpasses genomic predictability in less-related breeding material. Theoretical and Applied Genetics **134**(5), 1409–1422 (2021) https://doi.org/10.1007/s00122-021-03779-1

Holland, J.B., Piepho, H.-P.: Don't BLUP Twice. G3 Genes|Genomes|Genetics **14**(12), 250 (2024) https://doi.org/10.1093/g3journal/jkae250

Jackson, R., Buntjer, J.B., Bentley, A.R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E., Poupard, B., Smith, S., Hayes, C., Barber, T., Love, B., Gaynor, R.C., Gorjanc, G., Howell, P., Mackay, I.J., Hickey, J.M., Ober, E.S.: Phenomic and genomic prediction of yield on multiple locations in winter wheat. Frontiers in Genetics **14** (2023) https://doi.org/10.3389/fgene.2023.1164935

Klein, A., Freitas, V., Wade, A., Aubert, G., Naudet-Huart, M., Touratier, M., Régulier, J., Herbommez, J.-F., Kraut, C., Roullet, G., Carlier-Lézie, A., Tayeh, N., Rincent, R., Dufour, P., Burstin, J., Touzy, G.: Combining phenomic and genomic selection for pea breeding improvement. The Plant Phenome Journal **9**(1), 70076 (2026) https://doi.org/10.1002/ppj2.70076

Kaushal, S., Gill, H.S., Billah, M.M., Khan, S.N., Halder, J., Bernardo, A., Amand, P.S., Bai, G., Glover, K., Maimaitijiang, M., Sehgal, S.K.: Enhancing the potential of phenomic and genomic prediction in winter wheat breeding using high-throughput phenotyping and deep learning. Frontiers in Plant Science **15** (2024) https://doi.org/10.3389/fpls.2024.1410249

Krause, M.R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R.P., Dreisigacker, S., Poland, J., Rutkoski, J., Sorrells, M., Gore, M.A., Mondal, S.: Hyperspectral Reflectance-Derived Relationship Matrices for Genomic Prediction of Grain Yield in Wheat. G3: Genes, Genomes, Genetics **9**(4), 1231–1247 (2019) https://doi.org/10.1534/g3.118.200856 . Chap. Genomic Prediction

Lado, B., Barrios, P.G., Quincke, M., Silva, P., Gutiérrez, L.: Modeling Genotype × Environment Interaction for Genomic Selection with Unbalanced Data from a Wheat Breeding Program. Crop Science **56**(5), 2165–2179 (2016) https://doi.org/10.2135/cropsci2015.04.0207

Lane, H.M., Murray, S.C., Montesinos-López, O.A., Montesinos-López, A., Crossa, J., Rooney, D.K., Barrero-Farfan, I.D., Fuente, G.N.D.L., Morgan, C.L.S.: Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. The Plant Phenome Journal **3**(1), 20002 (2020) https://doi.org/10.1002/ppj2.20002

Montesinos-López, O.A., Herr, A.W., Crossa, J., Montesinos-López, A., Carter, A.H.: Enhancing winter wheat prediction with genomics, phenomics and environmental data. BMC Genomics **25**(1), 544 (2024) https://doi.org/10.1186/s12864-024-10438-4

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E.: Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics **157**(4), 1819–1829 (2001) https:

//doi.org/10.1093/genetics/157.4.1819

Pereira, H.D., Nonato, J.V.A., Duarte, R.C.R.M., Gerhardt, I.R., Dante, R.A., Arruda, P., Yassitepe, J.E.d.C.T.: Temporal field phenomics of transgenic maize events subjected to drought stress: Cross-validation scenarios and machine learning models. The Plant Phenome Journal **8**(1), 70015 (2025) https://doi.org/10.1002/ppj2.70015

Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., Singh, A.K.: Development of Optimized Phenomic Predictors for Efficient Plant Breeding Decisions Using Phenomic-Assisted Selection in Soybean. Plant Phenomics (2019) https://doi.org/10.34133/2019/5809404

R Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria (2025)

Rebollo, I., Aguilar, I., Vida, F.P., Molina, F., Gutiérrez, L., Rosas, J.E.: Genotype by environment interaction characterization and its modeling with random regression to climatic variables in two rice breeding populations. Crop Science **63**(4), 2220–2240 (2023) https://doi.org/10.1002/csc2.21029

Roger, J.-M., Boulet, J.-C.: A review of orthogonal projections for calibration. Journal of Chemometrics **32**(9), 3045 (2018) https://doi.org/10.1002/cem.3045

Rincent, R., Charpentier, J.-P., Faivre-Rampant, P., Paux, E., Le Gouis, J., Bastien, C., Segura, V.: Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar. G3 Genes|Genomes|Genetics **8**(12), 3961–3972 (2018) https://doi.org/10.1534/g3.118.200760

Rakotoson, T., Dusserre, J., Letourmy, P., Ramonta, I.R., Cao, T.-V., Ramanantsoanirina, A., Roumet, P., Ahmadi, N., Raboin, L.-M.: Genetic variability of nitrogen use efficiency in rainfed upland rice. Field Crops Research **213**, 194–203 (2017) https://doi.org/10.1016/j.fcr.2017.07.023

Rakotoson, T., Dusserre, J., Letourmy, P., Frouin, J., Ratsimiala, I.R., Rakotoarisoa, N.V., cao, T.-V., Vom Brocke, K., Ramanantsoanirina, A., Ahmadi, N., Raboin, L.-M.: Genome-Wide Association Study of Nitrogen Use Efficiency and Agronomic Traits in Upland Rice. Rice Science **28**(4), 379–390 (2021) https://doi.org/10.1016/j.rsci.2021.05.008

Robert, P., Goudemand, E., Auzanneau, J., Oury, F.-X., Rolland, B., Heumez, E., Bouchet, S., Caillebotte, A., Mary-Huard, T., Le Gouis, J., Rincent, R.: Phenomic selection in wheat breeding: Prediction of the genotype-by-environment interaction in multi-environment breeding trials. Theoretical and Applied Genetics **135**(10), 3337–3356 (2022) https://doi.org/10.1007/s00122-022-04170-4

Rakotomalala, J., Vom Brocke, K., Frouin, J., Pot, D., Rabekijana, R.R., Ramanantsoanirina, A., Ratsimiala Ramonta, I., Cao, T.-V.: Breeding for nitrogen use efficiency: Lessons from genomic prediction experiments based on a broad-based population of upland rice. Rice Research: Open Access (2021)

Sapkota, P., Fonseca, J., Perumal, R., Crossa, J., Rooney, W.L.: Near-Infrared Spectroscopy-Based Phenomics Data Can Improve Genomic Prediction of Agronomic and Grain Quality Traits Across Multi-Environment Sorghum Hybrid Trials. Plants **14**(18), 2871 (2025) https://doi.org/10.3390/plants14182871

Shafiee, S., Montesinos-López, O.A., Crossa, J., Burud, I., Dieseth, J.A., Alsheikh, M., Lillemo, M.: Optimizing wheat crop performance: Genomic and phenomic insights into yield and days to maturity prediction using multi-temporal UAV imagery. The

Plant Phenome Journal **7**(1), 70004 (2024) https://doi.org/10.1002/ppj2.70004

Sleper, J.A., Zheng, C., Ji, L., Wang, X., Abd-Elrahman, A., Whitaker, V.M.: Exploring the efficacy of phenomic and genomic selection for yield and fruit quality traits in strawberry. The Plant Genome **18**(4), 70156 (2025) https://doi.org/10.1002/tpg2.70156

Tadesse, T., Wilde, P., Ejerso, W., Borchardt, D., Thorwarth, P., Miedaner, T.: Enhancing Predictive Ability of Agronomic and Quality Traits in Ethiopian Malting Barley (Hordeum vulgare L.) Using Spectral Variable Selection Methods. Plant Breeding **n/a**(n/a) (2026) https://doi.org/10.1111/pbr.70115

Togninalli, M., Wang, X., Kucera, T., Shrestha, S., Juliana, P., Mondal, S., Pinto, F., Govindan, V., Crespo-Herrera, L., Huerta-Espino, J., Singh, R.P., Borgwardt, K., Poland, J.: Multi-modal deep learning improves grain yield prediction in wheat breeding by fusing genomics and phenomics. Bioinformatics **39**(6), 336 (2023) https://doi.org/10.1093/bioinformatics/btad336

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H.: Welcome to the tidyverse. Journal of Open Source Software **4**(43), 1686 (2019) https://doi.org/10.21105/joss.01686

Washburn, J.D., Adak, A., DeSalvio, A.J., Arik, M.A., Murray, S.C.: High temporal resolution unoccupied aerial systems phenotyping provides unique information between flight dates. The Plant Phenome Journal **7**(1), 20113 (2024) https://doi.org/10.1002/ppj2.20113

Winn, Z.J., Amsberry, A.L., Haley, S.D., DeWitt, N.D., Mason, R.E.: Phenomic versus

genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines. The Plant Phenome Journal **6**(1), 20084 (2023) https://doi.org/10.1002/ppj2.20084

Wickham, H.: Ggplot2: Elegant Graphics for Data Analysis. Springer, ??? (2016)

Zhu, X., Leiser, W.L., Hahn, V., Würschum, T.: Phenomic selection is competitive with genomic selection for breeding of complex traits. The Plant Phenome Journal **4**(1), 20027 (2021) https://doi.org/10.1002/ppj2.20027

Zhu, X., Maurer, H.P., Jenz, M., Hahn, V., Ruckelshausen, A., Leiser, W.L., Würschum, T.: The performance of phenomic selection depends on the genetic architecture of the target trait. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik **135**(2), 653–665 (2022) https://doi.org/10.1007/s00122-021-03997-7

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- gxe5supp.pdf