Phenomic selection in wheat breeding: prediction of the genotype-by-environment interaction in multi-environment breeding trials

Pauline Robert, Ellen Goudemand, Jérôme Auzanneau, François-Xavier Oury, Bernard Rolland, Emmanuel Heumez, Sophie Bouchet, Antoine Caillebotte, Tristan Mary-Huard, Jacques Le Gouis, et al.

Phenomic selection in wheat breeding: prediction of the genotype-by-environment interaction in multi-environment breeding trials

Pauline Robert[1,2,3,4], Ellen Goudemand[4,] Jérôme Auzanneau[3], François-Xavier Oury[2], Bernard Rolland[5], Emmanuel Heumez[6], Sophie Bouchet[2], Antoine Caillebotte[1], Tristan Mary-Huard[1,7], Jacques Le Gouis[2], Renaud Rincent[1,2, #]

[1] Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France

[2] INRAE - Université Clermont-Auvergne, UMR1095, GDEC, 5 chemin de Beaulieu, 63000 Clermont-Ferrand, France

[3] Agri-Obtentions, Ferme de Gauvilliers, 78660, Orsonville, France

[4] Florimond-Desprez Veuve & Fils SAS, 3 rue Florimond-Desprez, BP 41, 59242, Cappelle-en-Pévèle, France

[5] INRAE–Agrocampus Ouest-Université Rennes 1, UMR1349, IGEPP, Domaine de la Motte, 35653 Le Rheu, France

[6] INRAE, UE 972, Grandes Cultures Innovation Environnement, 2 Chaussée Brunehaut, 80200 Estrées-Mons, France

[7] MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005 Paris, France

Corresponding authors: renaud.rincent@inrae.fr

ORCID : P. Robert : 0000-0002-6314-180X / J. Le Gouis : 0000-0001-5726-4902 / S. Bouchet : 0000-0001-5868-3359 / R. Rincent : 0000-0003-0885-0969

Abstract

Key message

Phenomic prediction of wheat grain yield and heading date in different multi-environmental trial scenarios is accurate. Modelling the genotype by environment interaction effect using phenomic data is a potentially low-cost complement to genomic prediction.

The performance of wheat cultivars in multi-environmental trials (MET) is difficult to predict because of the genotype by environment interactions (GxE). Phenomic selection is supposed to be efficient for modelling the GxE effect because it accounts for non-additive effects. Here, phenomic data are near infrared (NIR) spectra obtained from plant material. While phenomic selection has recently been shown to accurately predict wheat grain yield in single environments, its accuracy needs to be investigated for MET. We used four datasets from two winter wheat breeding programs to test and compare the predictive abilities of phenomic and genomic models for grain yield and heading date in different MET scenarios. We also compared different methods to model the GxE using different covariance matrices based on spectra. On average, phenomic and genomic prediction abilities are similar in all the different MET scenarios. Better predictive abilities were obtained when GxE effects were modelled with NIR spectra than without them, and it was better to use all the spectra of all genotypes in all environments for modelling the GxE. To facilitate the implementation of phenomic prediction, we tested MET designs where the NIR spectra were measured only on the genotype-environment combinations phenotyped for the target trait. Missing spectra were predicted with a weighted multivariate ridge regression. Intermediate predictive abilities for grain yield were obtained in a sparse testing scenario and for new genotypes, which shows that phenomic selection is an efficient and practicable prediction method for dealing with GxE.

Key words:

Bread wheat, Genomic selection (GS), Multi-Environment Trial (MET), Near infrared spectroscopy (NIRS),

Genotype by environment interaction (GxE), Phenomic selection (PS), Plant breeding, *Triticum aestivum*.

Declarations

Funding: This work was funded by Agri-Obtentions, Florimond-Desprez, and the Association Nationale de la Recherche et de la Technologie (ANRT, grant number 2019/0060).

Conflicts of interest/Competing interests: Co-author Ellen Goudemand was employed by Florimond-Desprez, and co-author Jérôme Auzanneau was employed by Agri-Obtentions. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Availability of data and material: The datasets generated during and/or analysed during the current study are not publicly available due to the breeding program privacy policy, but are available from the corresponding author on reasonable request.

Code availability: Code used to run the analysis is available from the corresponding author on request.

Author contribution statement: JA, FXO, BR and EH designed the field trials and collected the phenotypic data from Agri-Obtentions and INRAE. EGD provided the phenotypic data and genotyping data from the Florimond Desprez company. SB provided the genotyping data from Agri-Obtentions and INRAE and participated in discussions of this study. AC and TMH developed the method of multivariate weighted ridge regression prediction of NIR spectra. RR initiated the project and with JLG supervised the study and helped improving the manuscript. PR analysed the data and wrote the manuscript. All authors approved the final manuscript.

Introduction

To select superior crop cultivars, plant breeders mainly compare the recorded phenotypes of candidates then select the best ones. However, phenotypic information only documents the relative performance of a cultivar in that environment, which depends on the genome in interaction with the environment (GxE). The relative performance of cultivars commonly differs in different environments, leading to differences in how cultivars are ranked (Comstock and Moll 1963; Allard and Bradshaw 1964). To take GxE into account, breeders evaluate candidate lines across several environments (combinations of year × site × management). These multi-environment trials (MET) enable breeders to identify high-performing selection candidates in each environment. METs are very expensive because of the phenotyping costs, so only a limited number of candidates and environments can be tested at a time, which slows genetic progress.

Genomic selection (GS) (Bernardo 1994; Whittaker et al. 2000; Meuwissen et al. 2001) is a method developed to predict the performances of candidates using a predictive model calibrated with a genotyped and phenotyped training population, potentially taking GxE into account. Single-environment predictions on breeding material were promising, because the predictive abilities (PA) were higher than those obtained using pedigree information only (de los Campos et al. 2009; de los Campos et al. 2010; Crossa et al. 2010). In the past decade, specific GS models have been proposed for predicting GxE. Burgueño et al. (2012) found that integrating the information that was common to different environments could improve the PA of the model by 17.5% compared to models with a main genetic effect only. To do this, an environmental covariance matrix was introduced in the prediction model. However, these models were still only applicable to environments in which some phenotypes had been measured. Other models were proposed that relied on environmental covariates , either pedoclimatic data (e.g. temperature, radiation, soil characteristics) or stress indexes (Ly et al. 2017; Rincent et al. 2019). Environments with similar environmental covariates values are assumed to interact with genotypes in a similar way, so the model can be used to predict plant behaviour in new environments (Heslot et al. 2014; Jarquín et al. 2014; Malosetti et al. 2016; Lado et al. 2016; Ly et al. 2018). However, further improvements in predictions based on environmental covariates are marginal, as modelling GxE this way does not account for complex interactions between plants and environments. Some environmental covariates may not be responsible for GxE while the covariates responsible for the GxE may differ from one genotype to another. The response of a plant to its environment can be characterised at the molecular level using endophenotypes such as transcripts, proteins or metabolites. An endophenotype is essentially a measurable molecular trait that is intermediate between the genome and the final phenotype. It has been shown that transcripts (Frisch et al. 2010; Fu et al. 2012; Zenke-Philippi et al. 2017; Azodi et al. 2020), and a combination of transcripts and metabolites (Guo et al. 2016; Westhues et al. 2017; Schrag et al. 2018; Wang et al. 2019) can be used in accurate genomic-like omics based (GLOB) prediction of the performances of hybrid maize or hybrid rice, especially for complex traits. Prediction of grain yield (GY) using such omics data was on average just as accurate as those obtained with GS models using molecular markers. However, when only information on metabolites was used to inform the model, PA were lower than when using molecular markers to predict GY in hybrid maize and in barley (Riedelsheimer et al. 2012; Xu et al. 2016; Gemmer et al. 2020). Practically, GLOB prediction is similar

to genomic prediction with the difference that scoring molecular markers is replaced by omics data collection for the chosen endophenotypes (Robert et al. 2022b). To limit costs, the omics data are generally collected once and for all on material grown in controlled conditions. The control conditions are radically different than the environments of MET, so endophenotype data may be of limited use in predicting GxE.

Phenomic selection (PS) is a recently developed method (Rincent et al., 2018) similar to GS and GLOB, that replaces molecular markers or endophenotypes with information from near infrared spectroscopy (NIRS). By measuring the reflectance or absorbance of a tissue sample at different wavelengths of light from visible to near infrared (NIR), NIRS quantifies the chemical bonds and hence the molecular composition of the tissue. Like the endophenotypes, this molecular composition is under genetic control and is a consequence of the expression of the genotype in a particular environment. NIR spectra are not themselves molecular traits, but they are influenced by the molecular composition, and are thus able to capture genetic similarity between genotypes. NIRS, which is already routinely used by cereal breeders to predict quality traits such as grain protein content, has the advantage of being low-cost, high throughput and non-destructive (Osborne 2006). PS was shown to outperform GS in two prediction scenarios. The first used NIR spectra acquired in the same environment as the phenotyping, and the second used NIR spectra acquired in a different environment than the one in which the training population was phenotyped (Rincent et al., 2018). For single environment scenarios, PS has shown promising results for the prediction of GY in bread wheat (Rincent et al. 2018; Krause et al. 2019; Cuevas et al. 2019; Robert et al. 2022a), rye (Galán et al. 2020), maize (Lane et al. 2020), triticale (Zhu et al. 2021b) and soybean (Parmley et al. 2019; Zhu et al. 2021a). More recently, Robert et al. (2022a) found that it might be preferable to use the NIRS data from the same environment as the target trait (e.g. for GY) or to combine several spectra from various environments (e.g. for heading date (HD)) depending on the interactivity of the underlying genes with the environment. One of the key advantage of PS is the low cost of NIRS data acquisition for any species, in contrast to genotyping. For some major crops such as wheat, NIRS data are already routinely collected in breeding programs, which means that PS can already be applied without any additional cost for some applications.

Like endophenotypes, NIR spectra are the result of a genotype, an environment, and the interaction between them. Therefore, it is reasonable to expect that the spectra capture the genetic effect and the GxE even for complex traits such as GY. By comparing PA of models based on NIR spectra collected in the same or in different environments from the one in which the training population was phenotyped, it was shown that the information from NIRS was indeed able to capture part of the GxE (Rincent et al., 2018; Robert et al., 2022a). A few examples of phenomic prediction in the context of MET have now been reported. Krause et al. (2019) compared different best linear unbiased prediction (BLUP) models in GS (G-BLUP) and in PS (H-BLUP) and a model combining molecular markers and NIR spectra (multi-omics model) on MET data. They found that the maximal PA obtained was with the multi-omics model where the main genetic effect is modelled by markers or pedigree and the GxE effect is modelled by NIR spectra collected in each environment. Lane et al. (2020) integrated the wavelength-by-environment interaction effect with all spectra collected in each environment of a MET in a H-BLUP model or in functional regression. They then tested these models in a MET scenario where one environment was new in the trial network and no genotypes were phenotyped. On average, the PA reached 0.40 for H-BLUP and 0.53 for the functional regression. These two demonstrations show the promise of integrating information from NIR spectra with GxE estimates.

Our aim here was to compare different PS, GS and multi-omics models (combining molecular markers and NIR spectra) for predicting GY and HD of wheat breeding candidates in different MET scenarios. More precisely, we investigated whether NIR spectra can model the GxE effect in predictive models by comparing different covariance structures based on spectral data. For the modelling of GxE using NIR spectra to be practicable, breeders would ideally acquire spectra in all environments of the MET and for all genotypes, whether they were used to calibrate the model or not, which would require specific nurseries for growing the reference genotypes in each environment. To overcome this limitation, we also tested two approaches for predicting missing spectra and evaluated them in two original MET scenarios.

Materials and Methods

Plant material, genotyping and NIRS acquisition

The plant material, genotyping and NIR spectra acquisition in the present study have been previously described in Robert et al. (2022a). Briefly, four datasets of winter bread wheat breeding lines were used: Set1-2016, Set2-2019, Set4-2018, Set4-2019. Set1 corresponds to lines developed in the breeding program of Florimond Desprez (France) and the other sets to lines developed in the breeding program of Agri-Obtentions (France) in collaboration with INRAE (France). Set1 and Set2 correspond to the first year of trial evaluation for the candidate lines and the two Set4 sets to the second year of trial evaluation. These datasets were chosen based on the number of environments

of the MET where NIR spectra were acquired and the number of candidate lines phenotyped and genotyped. Each dataset was analysed independently.

In more detail, Set1-2016 is composed of 152 candidate lines genotyped with the 35K breeder Bristol array (Axiom™ Wheat Breeder's Genotyping Array). NIR spectra of 10 g of wheat flour were acquired with the NIR 6500 FOSS spectrometer (FOSS NIR Systems, Silver Spring, MD, USA) over the range 400 to 2500 nm in steps of 2 nm. Final spectra are the average of 32 repeated measurements. Set2-2019 is composed of 325 lines genotyped with the 35K BreedWheat array (Axiom™ BreedWheat Genotyping Array) which is a subset of the TaBW410k SNPs array (Kitt et al. 2021) including 280k SNPs from the TaBW280k SNP array (Rimbert et al. 2018). NIR spectra of 150 g of grain were acquired with the XDS NIR Analysers FOSS spectrometer (FOSS NIR Systems, Silver Spring, MD, USA) in the range 400 to 2500 nm in steps of 2 nm. Final spectra are the average of 16 repeated measurements. Set4-2018 and Set4-2019 are composed of 71 and 100 lines, respectively, genotyped with the 35K BreedWheat array (Axiom™ BreedWheat Genotyping Array) which is a subset of the TaBW410k SNP array including 280k SNPs from the TaBW280k SNP array (Rimbert et al. 2018). NIR spectra on 350 g of grain were acquired with the MPA II FT-NIR analyser (Bruker Optics, Ettlingen, Germany) ranging from 3594.92 cm^{-1} to 12489.60 cm^{-1} in steps of 7.7 cm^{-1}. Final spectra are the average of 64 repeated measurements. NIRS data of Set4 were harmonised with the other NIRS data using a conversion into nm in steps of 2 nm and a common window was defined from 802 to 2492 nm for all the NIR spectra.

Analysis of the phenotypes measured in the METs, genotypic data and NIRS data

Description of the different METs and the phenotypic data

The trial information and designs of the different METs (two to four environments) recorded in the datasets are described in Robert et al. (2022a). Briefly, Set1-2016 is composed of two environments (two sites × one year) with augmented design trials in Houville and Lectoure for the year 2016. Lines were phenotyped for GY and HD. Set2-2019 is composed of two environments (two sites × one year) with augmented design trials in Estrée-Mons and randomised block trials in Genlis for the year 2019. Intensive management practices were used at both of the latter sites, except that lines were not treated with fungicide at Estrée-Mons. Lines were phenotyped for GY. Set4-2018 and Set4-2019 are composed of four environments (two sites × two treatments) with three complete randomised block trials in Estrée-Mons and Le Moulon and intensive practices or low input treatment at each, in the years 2018 and 2019. Set4-2018 and Set4-2019 are thus independent datasets. For Set4-2018, one environment was removed from the analysis due to a mis-association between genotypes and phenotypes, so only three environments were analysed here. Lines in Set4-2018 and Set4-2019 were phenotyped for GY and HD.

Adjustment of the phenotypic and NIRS data and estimation of the variance components

When possible, GY, HD and the NIR spectra at each wavelength were adjusted with specific spatial models described in Robert et al. (2022a). Briefly, statistical models were used to compute adjusted means corrected for spatial effects, or block and replicate effects, or all three depending on the experimental design. The same model was applied for target traits (GY and HD) and the absorbance at each of the 845 wavelengths.

Spectra were visualised to filter out any spectra with abnormal absorbances resulting from technical errors. For trials which included repetitions, adjusted means of NIR absorbance were computed as for GY and HD. Finally, all the spectra were pre-treated (first derivative of the normalized spectra) to eliminate noise inherent in the absorbance measurement. The different corrections are described in Robert et al. (2022a).

For the datasets with replicates (Set4-2018 and Set4-2019) we estimated the different variance components for GY, HD and NIRS wavelength readouts based on the following equation:

$$\hat{Y}_{ijk} = \mu + E_j + G_i + GE_{ij} + \epsilon_{ijk} \quad (1)$$

$\hat{Y}_{ijk}$ is the trait value corrected for spatial effects for the replicate k of the genotype i in the environment j; μ is the intercept; E_j, G_i and GE_{ij} are random environmental, genetic and interaction effects, respectively; and ϵ_{ijk} is the random residual effect. We assumed that the random effects are independent and identically distributed (i.i.d) and normally distributed such that $E_j \overset{i.i.d.}{\sim} N(0, \sigma_E^2)$, $G_i \overset{i.i.d.}{\sim} N(0, \sigma_G^2)$, $GE_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_{GE}^2)$ and $\epsilon_{ijk} \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$. Broad-sense heritabilities within each environment for traits and wavelengths were described in Robert et al. (2022a).

For all the datasets without replicates we followed the Rincent et al. (2018) approach where the genomic variance components were estimated from the following bivariate mixed model across two environments:

$$\hat{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = X\beta + Zu + e \quad (2)$$

where y_1 and y_2 are the absorbances for a given wavelength in each environment, corrected for spatial effects. **X** and **Z** are the design matrices of the corresponding effect, **β** is the vector of fixed environment effect, **u** is the random polygenic effect with $\mathrm{var}(u) = \begin{bmatrix} \sigma^2_{u1} & \sigma_{u12} \\ \sigma_{u21} & \sigma^2_{u2} \end{bmatrix} \otimes \mathbf{K}$, and **e** is the vector of the residuals with $\mathrm{var}(\mathbf{e}) = \begin{bmatrix} \sigma^2_{e1} & 0 \\ 0 & \sigma^2_{e2} \end{bmatrix} \otimes \mathbf{I}$, **K** being the genomic relationship matrix described below, and **I** the identity matrix.

The variance components were then estimated with the approach of Yamada et al. (1998):

$$\sigma^2_G = \hat{\sigma}_{u12},\ \sigma^2_{G\times E} = \frac{1}{2}(\hat{\sigma}_{u1} + \hat{\sigma}_{u2}) - \hat{\sigma}_{u12} \text{ and } \sigma^2_{\epsilon} = \frac{1}{2}(\hat{\sigma}^2_{e1} + \hat{\sigma}^2_{e2})$$

This decomposition was conducted on paired environments from two different sites. For Set1-2016 and Set2-2019, the two environments of each dataset were used to compute the variance components. For Set4-2018, only the data from a low input treatment were used. Finally, for Set4-2019, data were paired by environments from the same treatment (intensive practices or low input) but from different sites (Estrée-Mons and Le Moulon).

From now on, we consider that Y_{ij} stands for the adjusted mean of the considered trait for the genotype i in environment j. It is recommended to use the precision of the estimates of this first stage in subsequent analysis (Damesa et al. 2017). But, as precision could not be estimated for all the datasets and was homogeneous in the others, we did not use it in the second stage of our analysis (genomic and phenomic predictions).

Quality analysis of the genomic data

Molecular marker results were checked for quality. Markers were eliminated if the minor allele frequency was less than 5%, or the heterozygosity rate or missing value rate were greater than 5%. A total of 5 824 SNPs for Set1-2016, 12 303 SNPs for Set4-2018, and 19 512 SNPs for both Set2-2019 and Set4-2019 remained after filtering. On average, 1.1% of the SNPs were missing and were imputed with the average allele frequency of the corresponding marker. All the data quality analysis was conducted with the sommer R package (Covarrubias-Pazaran 2016).

Estimation of the covariance matrices describing the similarities between genotypes or environments

The predictive models developed further in this study rely on covariance matrices to describe the covariance between the genotypes and between the GxE interactions. Relationship matrices between varieties were estimated based on molecular markers (kinship matrix K) or NIR spectra (hyperspectral relationship matrix H).

Genomic and hyperspectral relationship matrices

The kinship matrix **K** was computed following the Endelman and Jannink (2012) equation:

$$\mathrm{K} = \frac{\mathbf{A}\mathbf{A}'}{2\sum \mathrm{p_k}(1-\mathrm{p_k})}$$

where **A** is a centred genotypic matrix with dimensions $N_G \times M$, N_G is the number of genotypes, and M the number of molecular markers. For the i^{th} individual and the k^{th} marker, $A_{ik} = X_{ik} + 1 - 2p_k$ with X the genotype matrix, coded in {-1,0,1} and p_k the frequency of allele 1 at marker k. K was computed with the rrBLUP R package (Endelman 2011).

The hyperspectral relationship matrix specific to environment j was computed as:

$$\mathbf{H_j} = \frac{\mathbf{Sp_j^*}\,\mathbf{Sp_j^{*\prime}}}{\mathrm{L}}$$

where $\mathbf{Sp_j^*}$ is the centred and scaled matrix of NIR spectra (dimension $N_G \times L$) from environment j, L the number of wavelengths, and j the environmental index with $j \in \{1, \ldots, N_E\}$, N_E being the number of environments of the MET.

The combined hyperspectral relationship matrix was computed as

$$\mathbf{H_{cb}} = \frac{\mathbf{S^*_{cb}}\mathbf{S^{*\prime}_{cb}}}{\mathrm{L}}$$

where $\mathbf{S}^{*}_{\mathbf{cb}}$ is the centred and scaled matrix of NIR spectra added one next to the other from NIR spectrum 1 matrix (S_1) to NIR spectrum N_E matrix ($\mathbf{S}_{\mathbf{N_E}}$)

$$\mathbf{S_{cb}} = (\mathbf{S_1} \quad ... \quad \mathbf{S_{N_E}})$$

$\mathbf{S_{cb}}$ had $N_G \times \sum_{j=1}^{N_E} \lambda_j$ dimensions where λ_j designates the number of wavelengths for the NIR spectrum j, and N_G the number of genotypes.

The $\mathbf{K}$, $\mathbf{H_j}$ and $\mathbf{H_{cb}}$ matrices have the same dimensions ($N_G \times N_G$).

Genotype-by-environment covariance matrix

We then decided to characterise the covariance between all the genotype-environment combinations based on NIR spectra. It was shown in literature that environmental covariates can be used to estimate such covariances. We assumed that wavelengths constituting the NIR spectrum are similar to environmental covariates and relevant to describe GxE because they capture the expression of the genotype in a specific environment.

Each wavelength of a spectrum corresponds to an environmental covariate in which the absorbance is dependent on an environment j and a genotype i. The matrix of the environmental covariates is then all the NIR spectra of the MET, called $\mathbf{S_{all}}$ with dimension $N_{GE} \times L$, N_{GE} corresponding to the number of genotype-environment combinations.

$$\mathbf{S_{all}} = \begin{pmatrix} \mathbf{S_1} \\ ... \\ \mathbf{S_{N_E}} \end{pmatrix}$$

The GxE interaction similarity matrix $\mathbf{P}$ was computed based on the full matrix of NIR spectra $\mathbf{S_{all}}$. First, we calculated the Euclidian distance ($\mathbf{D_{S_{all}}}$) between the different combinations of genotype × environment terms with $\mathbf{S_{all}}$. Then, we estimated the genotype × environment covariance matrix following the equation:

$$\mathbf{P} = \mathbf{1_{GE}} - \frac{\mathbf{D_{S_{all}}}}{\max(\mathbf{D_{S_{all}}})}$$

where $\mathbf{1_{GE}}$ is a matrix of 1 of size ($N_G \times N_E$, $N_G \times N_E$). Here we assume that two genotype-environment combinations will covary if the two corresponding spectra are similar. The underlying hypotheses is that spectra capture the similarity between genotypes as well as the genetic similarity between environments.

$\mathbf{K}$, $\mathbf{H_j}$, $\mathbf{H_{cb}}$ and $\mathbf{P}$ were scaled to have a sample variance of 1 to avoid biased parameter estimations due to different scales (Kang et al. 2010; Forni et al. 2011).

Scenarios of prediction and corresponding cross-validation schemes

Six prediction scenarios were considered to simulate the different prediction objectives of genomic and phenomic selection (Figure 1). In four scenarios, all the NIR spectra are available for all varieties in each environment of the MET, whether phenotyped or not for the target trait. In the other two scenarios, one environment of the MET is the environment of reference where all the varieties have their NIR spectrum acquired, while in the other environments the NIRS are collected only on the varieties phenotyped for the target trait. The phenotype and NIR spectra of other varieties are hence missing in the non-reference environments. For each prediction scenario a cross-validation scheme is described below.

For the first category of scenarios, NIR spectra are available for all varieties in each environment, whether or not phenotyped for the target trait. **oGoE** corresponds to a sparse testing scenario. This scenario was tested in a CVRandom scheme with 5-fold cross-validation, which consisted in randomly splitting the MET data in five folds of the same size. **nGoE** corresponds to new varieties in observed environments, which mimics the evaluation of new varieties in the MET without any phenotyping of the target trait. Concretely, missing data were attributed in all the environments for the varieties designated as new. This scenario was tested in a CVNewG scheme with 5-fold cross-validations, in which varieties are randomly split in five groups. **oGnE** corresponds to observed varieties in a new environment as it mimics the prediction in a new environment. The predicted environment is new in the sense that the target trait is not measured, but the NIR spectra are known. This scenario was tested in a CVNewE scheme with a leave-one-environment-out scheme, in which one environment of the MET is predicted with the others. Finally, **nGnE** corresponds to new varieties in a new environment. This scenario was tested in a CVNewGE scheme with a leave-one-environment-out scheme and a 5-fold cross-validation to determine the varieties considered as new.

For the second category of scenarios, in which NIR spectra are only available for the plots phenotyped for the target trait, **oGoEref** and **nGoEref** correspond to the same scenarios as **oGoE** and **nGoE** respectively, with the exception that one environment was randomly designated the environment of reference in which the phenotyping and NIR spectra are available for all varieties. In the other environments, the new varieties were not observed at all (neither for the target trait nor for NIRS). This mimics the situation where traditionally a breeder evaluates all the varieties in a single reference environment with sparse testing in the other environments. The scenario **oGoEref** was tested in a CVRandom_Ref scheme with one environment considered as reference and a 5-fold cross-validation, by randomly splitting the rest of the MET data in five folds of the same size. The scenario **nGoEref** was tested in a CVNewG_Ref scheme with one reference environment and a 5-fold cross-validation, by randomly splitting varieties in five folds of the same size.

To compare the performance of the different predictive models in the different cross-validation schemes, we calculated the predictive ability (PA) of each model as the Pearson's correlation coefficient between the predicted values and the adjusted means of the target trait (GY or HD) on the validation set, for each environment of prediction in the MET. For CVRandom, CVNewG and CVNewGE $5 \times N_E$ PA were calculated once the five folds were designated as the validation sets. For CVNewE, N_E PA were calculated because we applied a leave-one-environment-out method. In scenarios **oGoEref** and **nGoEref**, the PA was not calculated for the reference environment. For CVRandom_Ref and CVNewG_Ref, $5 \times (N_E - 1) \times N_E$ PA were then calculated on the 5-fold cross-validation multiplied by the number of environments, which are in turn considered as the new environment or reference environment. All the different cross-validation schemes were repeated 50 times to get stable results, except for the CVNewE, which is a simple leave-one-environment-out validation.

Models developed for the different multi-environment prediction scenarios

Genomic prediction models

In the different prediction scenarios, we compared models with different levels of complexity, using molecular markers, NIR spectra or both. All the different models used in further analysis are summarised in Table 1.

For genomic prediction, we used two reference models based on the genomic relationship matrix (kinship) allowing to share information between varieties but not between environments (Jarquín et al. 2014).

$$\mathbf{EG}: \ Y_{ij} = \mu + E_j + G_i + \epsilon_{ij} \quad (3)$$

$$\mathbf{EG\_GxE}: \ Y_{ij} = \mu + E_j + G_i + GE_{ij} + \epsilon_{ij} \quad (4)$$

Y_{ij} is the phenotype for variety i in environment j; μ is the intercept; G_i is a random polygenic effect of the variety i following a normal distribution $G \sim N(0, K\sigma^2_G)$; and GE_{ij} is a random effect corresponding to the interaction between variety i and environment j following a normal distribution $GE \sim N(0, [\mathbf{ZKZ'}] \odot [\mathbf{Z_E Z_E'}]\sigma^2_{GE})$. So, in this model (4) no information is shared between environments in the GE term. $\mathbf{K}$ is the relationship matrix calculated with molecular markers (see above), $\mathbf{Z}$ the incidence matrix for the genetic effect and $\mathbf{Z_E}$ the incidence matrix for the effects of the environments. $\odot$ corresponds to the Hadamard product. Finally, ϵ_i is the random residual effect that follows the normal distribution $\epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2_\epsilon)$. We assumed a same residual variance for all environments. The different random effects of the models are assumed to be independent.

Single-NIRS phenomic models

For phenomic prediction we used two kinds of predictive models, involving a NIR spectrum acquired in only one environment (Single-NIRS) or from multiple NIR spectra acquired in different environments of the MET (Multi-NIRS). Like the genomic prediction models (3) and (4), we compared two reference models based on the hyperspectral relationship matrix computed from a single environment.

$$\mathbf{EW}: \ Y_{ij} = \mu + E_j + W_i + \epsilon_{ij} \quad (5)$$

$$\mathbf{EW\_WxE}: \ Y_{ij} = \mu + E_j + W_i + WE_{ij} + \epsilon_{ij} \quad (6)$$

Y_{ij} is the phenotype for variety i in environment j; μ is the intercept; W_i is a random genetic effect of the variety i following a normal distribution $W \sim N(0, H_S\sigma^2_W)$; and WE_{ij} is a random effect corresponding to the interaction between variety i and environment j following a normal distribution $WE \sim N(0, [\mathbf{ZH_S Z'}] \odot [\mathbf{Z_E Z_E'}]\sigma^2_{GE})$. $\mathbf{H_S}$ is the relationship matrix calculated with NIR spectra (see above) from the environment S with $S \in \{1, \ldots, N_E\}$. Finally, ϵ_i is the random residual effect following a normal distribution $\epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2_\epsilon)$. The different random effects of the

models are assumed to be independent. In Models (5) and (6) the H_S matrix of each environment was tested successively, resulting in N_E Models (5) and (6).

Multi-NIRS phenomic models

The next objective was to assess the effect of using a hyperspectral relationship matrix combining NIR spectra acquired from different environments to predict GY and HD. To do this, we used the $\mathbf{H_{cb}}$ matrix (the hyperspectral relationship matrix computed with S_{cb}) described above. Multi-NIRS models were adapted from the single-NIRS models (5) and (6) with models called $\mathbf{EW_{cb}}$ and $\mathbf{EW_{cb\_}W_{cb}xE}$ respectively:

$$\mathbf{EW_{cb}}: Y_{ij} = \mu + E_j + Wcb_i + \epsilon_{ij} \quad (7)$$

$$\mathbf{EW_{cb\_}W_{cb}xE}: Y_{ij} = \mu + E_j + Wcb_i + WcbE_{ij} + \epsilon_{ij} \quad (8)$$

The description of the different effects is the same as in the previous models, and Wcb_i is a random genetic effect of the variety i following a normal distribution $Wcb \sim N(0, H_{cb}\sigma^2_{Wcb})$, $WcbE_{ij}$ is a random effect corresponding to the interaction between variety i and environment j following a normal distribution $WcbE \sim N(0, [\mathbf{ZH_{cb}Z'}] \odot [\mathbf{Z_E Z_E'}]\sigma^2_{GEcb})$. $\mathbf{H_{cb}}$ is the relationship matrix calculated with S_{cb} NIR spectra. Finally, ϵ_i is the random residual effect which follows the normal distribution $\epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2_\epsilon)$. The different random effects of the models are assumed to be independent.

We tested several other models to evaluate how well NIR spectra capture and predict GxE interactions. First, we specified in the WE_{ij} covariance matrix a block-diagonal structure (Malosetti et al. 2016), in which each block corresponds to the $\mathbf{H_j}$ matrix of the corresponding environment. In this model we consider that within one environment, varieties covary according to their hyperspectral similarity in the same environment. The assumption is that NIR spectra capture information on local adaptation. The corresponding model is:

$$\mathbf{EW_{cb\_}W_{diag}xE}: Y_{ij} = \mu + E_j + Wcb_i + WEd_{ij} + \epsilon_{ij} \quad (9)$$

The description of the different effects is the same as in Model (7), and WEd_{ij} follows a normal distribution $WEd \sim N(0, \mathbf{H_{diag}}\sigma^2_{WEd})$, with $\mathbf{H_{diag}} = \begin{pmatrix} H_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & H_{N_E} \end{pmatrix}$ and $\mathbf{H_1}$ to $\mathbf{H_{N_E}}$ correspond to the N_E hyperspectral relationship matrices available in the N_E environments of the MET.

Secondly, we described GxE by using the **P** similarity matrix based on all the NIR spectra:

$$\mathbf{EW_{cb\_}P}: Y_{ij} = \mu + E_j + Wcb_i + WEp_{ij} + \epsilon_{ij} \quad (10)$$

The description of the different effects is the same as in Model (7), and $\mathbf{WEp_{ij}}$ follows a normal distribution $WEp \sim N(0, P\,\sigma^2_{WEp})$. **P** is the genotype-by-environment similarity matrix computed with the NIR spectra (see above).

Results of models (9) and (10) are presented in the Supplementary section.

Multi-omics models

We developed models combining information from molecular markers and NIR spectra. We tested different models including genetic effects involving covariance matrices **K** or $\mathbf{H_{cb}}$ and GxE involving covariance matrices **K**, $\mathbf{H_{cb}}$, $\mathbf{H_{diag}}$ or **P**.

$$\mathbf{EG\_P}: Y_{ij} = \mu + E_j + G_i + P_{ij} + \epsilon_{ij} \quad (11)$$

$$\mathbf{EGW_{cb}}: Y_{ij} = \mu + E_j + G_i + Wcb_i + \epsilon_{ij} \quad (12)$$

$$\mathbf{EGW_{cb\_}GxE}: Y_{ij} = \mu + E_j + G_i + Wcb_i + GE_{ij} + \epsilon_{ij} \quad (13)$$

$$\mathbf{EGW_{cb\_}W_{cb}xE}: Y_{ij} = \mu + E_j + G_i + Wcb_i + WcbE_{ij} + \epsilon_{ij} \quad (14)$$

$$\mathbf{EGW_{cb\_}GxE\_W_{cb}xE}: Y_{ij} = \mu + E_j + G_i + Wcb_i + GE_{ij} + WcbE_{ij} + \epsilon_{ij} \quad (15)$$

The description of the different effects is the same as in the previous models. Alternatives to models (13) and (14) were defined by replacing $WcbE_{ij}$ with $\mathbf{W_{diag}xE}$ (models $\mathbf{EGW_{cb\_}W_{diag}xE}$ and $\mathbf{EGW_{cb\_}GxE\_W_{diag}xE)}$, or with $WEp \sim N(0, \mathbf{P}\,\sigma^2_{WEp})$ (models $\mathbf{EGW_{cb\_}P}$ and $\mathbf{EGW_{cb\_}GxE\_P}$). Results from these alternatives to models (13) and (14) are presented in the Supplementary section.

All the models presented here were fitted with the R package BGLR (Pérez and de los Campos 2014) with a burn-in of 2 000 iterations on a running total of 10 000 iterations to reach convergence. Starting parameters for df_0 and S_0 were chosen as 5 and 0.5 by default.

Imputation of missing spectra for unphenotyped varieties

All the models developed above assume that NIR spectra have been obtained from all the varieties, whether phenotyped or not, in all the environments of the MET. However, in more ambitious scenarios such as **oGoEref** and **nGoEref**, the genotype-environment combinations to be predicted are not characterised by NIRS, except in the reference environment. It is necessary to predict the missing spectra for the models involving local NIR spectra (Multi-NIRS and Multi-omics).

We used two approaches to predict the missing NIR spectra of the genotype-environment combinations to be predicted using the spectra of all genotypes in the reference environment, and the spectra of the training genotypes in the environment to be predicted. These predicted spectra were then used to compute the environment-specific or the combined hyperspectral relationship matrices compatible with all the phenomic models presented above. For the single-NIRS models, the NIR spectra were acquired on all genotypes only in the reference environment. So, Models (5) and (6) were adapted using $\mathbf{H_{ref}}$ instead of $\mathbf{H_S}$, leading to models $\mathbf{EW}_{ref}$ and $\mathbf{EW}_{ref\_}\mathbf{W}_{ref}\mathbf{xE}$, respectively. For multi-NIRS models and the modelling of the GxE interaction with local or combined spectra, missing spectra were added by imputation using one of two methods.

The first method is a nearest neighbour method. For any given environment each missing genotype is attributed the same spectrum as the most similar genotype. This similarity is based on the lowest Euclidian distance between the NIR spectra in the environment of reference (Figure S1). Results from this method are presented in the Supplementary section.

The second method is a weighted multivariate ridge regression (WMRR) following the equation:

$$\mathbf{Y} = \mathbf{X\beta} + \epsilon \quad (14)$$

$\mathbf{Y}$ is the matrix of spectra to be predicted and $\mathbf{X}$ the matrix of spectra from the reference environment, and both matrices have dimensions $N_G \times L$. $\boldsymbol{\beta}$ is the matrix of estimated wavelength effects with dimensions $L \times L$, while $\boldsymbol{\epsilon}$ is the matrix of random error effect such as $vec(\epsilon) \sim N(0, \sigma^2_\epsilon \mathbf{I_{N_G \times L}})$.

We weighted the wavelengths so that some are more explicative than others. For this, we applied a ridge regression to the training set to predict the target trait in the environment to be predicted. This was possible because NIR spectra were acquired on the training set in the environment to be predicted. We then used the estimated β (absolute value) to define a weight matrix: $\mathbf{Q} = \begin{pmatrix} q_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & q_L^2 \end{pmatrix}$

To minimise the norm: $\|\mathbf{Y} - \mathbf{XB}\|^2 = \|\mathbf{YQ} - \mathbf{XBQ}\|^2$, the estimator $\mathbf{B}$ was calculated with the following equation:

$$\widehat{\boldsymbol{B}} = [(\boldsymbol{Q} \otimes \boldsymbol{X})^T (\boldsymbol{Q} \otimes \boldsymbol{X}) + \lambda \boldsymbol{I}_{\boldsymbol{L}\times\boldsymbol{L}}]^{-1} (\boldsymbol{Q} \otimes \boldsymbol{X})^T \left(\boldsymbol{Q} \otimes \boldsymbol{I}_{\boldsymbol{N_G}}\right)\boldsymbol{Y}_{\boldsymbol{vec}}$$

$\otimes$ stands for the Kronecker product.

The shrinkage parameter λ was selected based on a 10-fold cross-validation. Multivariate ridge regression and estimation of the weights were run with the glmnet package (Friedman et al. 2010).

Contrary to the nearest neighbour method, the WMRR predicts a new spectrum for all the genotypes (training and validation sets) in the specific environment. The predicted spectra of both calibration and validation sets were used to compute the hyperspectral matrix. Predictive ability of models involving predicted spectra were compared to a theoretical "Optimum" model, based on the observed NIR spectra instead of the predicted ones.

Table 1 *Summary of all genomic, phenomic and multi-omics models. For each model the different effects and associated covariances matrices are indicated, as well as the prediction scenarios in which they were tested. Effects were either included (×) or excluded (-). For random effects, the covariance matrices are indicated. Models were tested on of two scenarios (O, WMRR) or not (-). For* ***oGoE, nGoE, oGnE*** *and* ***nGnE****, models used a* ***H*** *matrix computed on observed NIRS (O), whereas for* ***oGoEref*** *and* ***nGoEref****, the multi-NIRS models and multi-omics models, used an* ***H*** *matrix computed on all spectra predicted by the weighted multivariate ridge regression (WMRR).* ***Eref*** *corresponds to an environment of reference in which the phenotyping and NIR spectra are available for all varieties. The* $\boldsymbol{H_S}$ *matrix was computed on spectra from one single environment. The* $\boldsymbol{H_{ref}}$ *matrix was computed on spectra from the environment of reference. Finally, the* $\boldsymbol{H_{cb}}$ *was computed on spectra from multiple environments.*

	Model effects					*Prediction scenarios*					
Models	E	G	W	G×E	W×E	oG oE	nG oE	oG nE	nG nE	oG oEref	nG oEref
Genomic models											
EG	×	$\mathbf{K}$	-	-	-	O	O	O	O	O	O
EG_GxE	×	$\mathbf{K}$	-	$[\mathbf{ZKZ}'] \odot [\mathbf{Z_E Z_E}']$	-	O	O	-	-	O	O
Single NIRS models											
EW	×	-	$\mathbf{H_S}$	-	-	O	O	O	O	-	-
EW_WxE	×	-	$\mathbf{H_S}$	-	$[\mathbf{ZH_S Z}'] \odot [\mathbf{Z_E Z_E}']$	O	O	-	-	-	-
$\mathbf{EW_{ref}}$	×	-	$\mathbf{H_{ref}}$	-	-	-	-	-	-	O	O
$\mathbf{EW_{ref}\_W_{ref}xE}$	×	-	$\mathbf{H_{ref}}$	-	$[\mathbf{ZH_{ref} Z}'] \odot [\mathbf{Z_E Z_E}']$	-	-	-	-	O	O
Multi-NIRS models											
$\mathbf{EW_{cb}}$	×	-	$\mathbf{H_{cb}}$	-	-	O	O	O	O	WMRR	WMRR
$\mathbf{EW_{cb}\_W_{cb}xE}$	×	-	$\mathbf{H_{cb}}$	-	$[\mathbf{ZH_{cb} Z}'] \odot [\mathbf{Z_E Z_E}']$	O	O	-	-	WMRR	WMRR
Multi-omics models											
EG_P	×	$\mathbf{K}$	-	-	$\mathbf{P}$	O	O	O	O	WMRR	WMRR
$\mathbf{EGW_{cb}}$	×	$\mathbf{K}$	$\mathbf{H_{cb}}$	-	-	O	O	O	O	WMRR	WMRR
$\mathbf{EGW_{cb}\_GxE}$	×	$\mathbf{K}$	$\mathbf{H_{cb}}$	$[\mathbf{ZKZ}'] \odot [\mathbf{Z_E Z_E}']$	-	O	O	-	-	WMRR	WMRR
$\mathbf{EGW_{cb}\_W_{cb}xE}$	×	K	H_{cb}	-	$[ZH_{cb}Z'] \odot [Z_E Z_E']$	O	O	-	-	WMRR	WMRR
$\mathbf{EGW_{cb}\_GxE\_W_{cb}xE}$	×	K	H_{cb}	$[ZKZ'] \odot [Z_E Z_E']$	$[ZH_{cb}Z'] \odot [Z_E Z_E']$	O	O	-	-	WMRR	WMRR

Results

Characterisation of the METs based on phenotypes

We characterised and compared four previously established sets of data from METS of two winter wheat breeding programs in France. Within each dataset, correlations were calculated between the adjusted means for GY for each pair of environments in the MET, that is, two environments in each of Set1-2016 and Set2-2019, three for Set4-2018, and four for Set4-2019 (Figure 2). GY correlations between environments were very variable, from as low as 0.01 in Set1 to a maximum of 0.72 in Set4-2019. Where possible, correlations were also calculated for the adjusted means for HD between each pair of environments within a MET (Figure 2). Unlike GY, the HD correlations were strong and did not vary widely. For example, in Set4-2019 the correlation ranged from 0.89 to 0.97.

For Set4-2019, the variance in GY and HD was decomposed using Equation (1). For Set4-2018, only the variance for GY was decomposed (Table S1, B). The other two sets did not have any replicate for GY and HD so the decomposition of variance could not be done. For GY, the variance in the GxE effect was low for Set4-2019 (12.81) and high for Set4-2018 (47.58) compared to the genetic effect (16.77 and 12.76, respectively). For GY, the environment explained most of the variance. For HD, the variance was mainly explained by the genetic effect.

Decomposition of the variance along NIR spectrum

We decomposed the genetic variance for all spectra based on the genomic kinship (Figure 3). Spectra from Set1-2016 had the lowest proportion of genomic variance (18% on average) and Set2-2019 had the highest proportion (67% on average). In all sets, the total variance in absorbances included a non-negligible proportion of GxE variance. Two datasets, Set1-2016 and Set4-2018-LI had on average, more GxE variance (30% and 28%, respectively) than genetic variance (18% and 23%, respectively). We also decomposed the variance of the absorbances based on Equation (1) and using NIRS acquisition replicates for Set4-2018 and Set4-2019 (Figure S2). The E effect explains most of the absorbance variances across all wavelengths. The G variance was also larger across the spectrum (on average 21.5% and 24.3% of total variance) than the GxE variance (representing 12.8% and 17.3%) in the respective sets.

Comparison of genomic, phenomic and multi-omics models in classical MET scenarios for the prediction of GY and HD

To predict the performance of individual genotypes in particular environments, several models (Table 1) were tested using different scenarios according to the objective of the prediction and which data were available or necessary for training and validation of the models (Figure 1).

Comparison of the scenarios oGoE and nGoE

For the oGoE and nGoE scenarios NIRS data were available for all genotype-environment combinations, and predictions were sought for certain unobserved combinations or new genotypes in all environments, respectively. The PA of models for GY were highly variable between each other, between the two scenarios and between datasets. Taking the average PA for the four datasets and comparing CVRandom (Table 2) and CVNewG (Table 3) validation scenarios, the highest PA were reached by the model EGW_{cb}_GxE_W_{cb}xE in the CVRandom (range 0.4-0.67) and in the CVNewG (range 0.42-0.67) validations.. On average, several models reached very high PA in validation scenario CVRandom (range 0.81-0.94) as well as high PA in CVNewG (0.57-0.74).

We considered first the models without a GxE effect. On average, for the prediction of GY, single NIRS model EW gave similar PA as multi-NIRS model EW_{cb} in both scenarios. For the prediction of HD, EW_{cb} gave higher PA than EW with an average gain of 16%. The best phenomic models in the CVRandom scenario gave similar PA to genomic models for the prediction of GY and HD. In scenario CVNewG, the best phenomic model was better than the best genomic model for the prediction of GY (0.44 versus 0.38) and for the prediction of HD (0.51 versus 0.43). Combining information from molecular markers and NIR spectra gave the best PA for GY in both scenarios (0.54 for CVRandom and 0.50 for CVNewG) and was equivalent to genomic and phenomic models for HD in scenario CVRandom. For CVNewG, multi-omics models performed better than genomic and phenomic models for the prediction of HD (0.64 versus 0.43 and 0.51).

Then we considered the models with a GxE effect. For GY, the predictive models that included the GxE effect always gave better predictions than those without. For the prediction of HD, the PA were similar for both kinds of models. We therefore focused on GY prediction only. In genomic models, a single covariance matrix was used to integrate the GxE effect, which increased the gain in PA by 24% on average. In single-NIRS and multi-NIRS models, the integration of the GxE effect, described in Table 1, increased the gain in PA by 23% on average. For phenomic models integrating the GxE effect, we also compared the results of using different covariance matrix

structures (Tables S2 and S3). In multi-NIRS models, all the models with a GxE effect performed similarly. Finally, the multi-omics models with higher PA were the ones combining all main genetic effects (G and W_{cb}) and all the GxE interaction effects (GxE and W_{cb}xE or W_{diag}xE).

Comparison in the scenarios oGnE and nGnE

In the scenarios oGnE and nGnE, there are NIRS data for every genotype-environment combination but some environments do not have any phenotypic data, and for nGnE some genotypes are not phenotyped. The PA for GY were much lower in oGnE (Tables S4) and nGnE (Tables S5) than in oGoE and nGoE. Considering the average of the different sets and comparing CVNewE and CVNewGE validation scenarios, the highest PA were reached with the multi-omics model EG_P (0.32 and 0.27, respectively), higher than with the EG model (0.24 and 0.16, respectively). Models integrating the GxE effect (modelled by P) had better PA than those without it, with an average gain of 36% and 57% for the corresponding scenarios. EG_GxE was not compared to the other models because the GxE could not be estimated with this model for new environments. This was also the case for models with W_{cb}xE and W_{diag}xE. The best models were EG_W_{cb} and EG_W_{cb}_P with PA reaching 0.91 for CVNewE and around 0.62 for CVNewGE. Adding the GxE effect in the models did not increase the PA for HD.

Comparison of genomic, phenomic and multi-omics models in two original MET scenarios for the prediction of GY and HD

We compared the previous developed models in two original scenarios oGoEref (Table 4) and nGoEref (Table 5) where the NIR spectra were missing for the validation set in the environments for which GY was to be predicted. We imputed the missing spectra by weighted multivariate ridge regressions (WMRR). The results of this imputing methods was lower than the "optimum" for both CV schemes when all the spectra were known. For the best multi-NIRS model with WMRR, the loss in PA was around 28% for both scenarios and for the best multi-omics model, the loss in PA was around 11% for CVRandom_Ref and 14% for CVNewG_Ref. In comparison to the nearest neighbour imputation (Tables S6 and S7), the WMRR gave on average, better PA for multi-NIRS and multi-omics models in both scenarios.

Using multi-NIRS models with imputed spectra gave slightly better PA than using the single NIRS model based on spectra of reference, while using multi-omics models gave slightly better PA than genomic models. The best multi-omics models integrated all the G, W, GxE and WxE effects (Table 4, Table 5) in both scenarios. On average, EGW_{cb}_GxE_W_{cb}xE with imputed spectra gave the best PA for both scenarios, 10% better than genomic models. Models including the GxE effect gave higher PA than those without. The different types of covariance structures to model the GxE effect performed similarly (Tables S6 and S7). Finally, the PA of the best model in CVRandom_Ref and in CVNewG_Ref were lower than those of the best model in CVRandom and CVNewG with a decrease of 11% and 14% respectively.

Table 2. *Predictive abilities of the different genomic, single NIRS, multi-NIRS and multi-omics models in scenario* ***oGoE*** *for GY and HD. CVRandom is a five-fold cross-validation scheme in which random data were missing from the dataset. This CV corresponds to a sparse-testing design. Averaged predictive abilities based on 50 repetitions are presented for each set, and averaged over all sets for GY or three sets for HD. The best predictive abilities by set and by model category (based on Average) are shown in bold.*

oGoE scenario (CVRandom)	*GY prediction*					*HD prediction*			
Models	**Set1 2016**	**Set2 2019**	**Set4 2018**	**Set4 2019**	**Average**	**Set2 2019**	**Set4 2018**	**Set4 2019**	**Average**
Genomic models									
EG	0.13	0.35	0.31	0.55	0.34	0.76	0.91	**0.94**	**0.87**
EG_GxE	0.17	0.55	**0.44**	0.6	**0.44**	0.72	0.92	**0.94**	0.86
Single NIRS models									
EW	0.19	0.38	0.28	0.38	0.31	0.69	0.79	0.93	**0.80**
EW_WxE	0.29	0.46	0.28	0.41	**0.36**	0.67	0.73	0.93	0.78
Multi-NIRS models									
EW_{cb}	0.21	0.43	0.33	0.45	0.36	0.78	0.87	0.93	**0.86**
EW_{cb}_W_{cb}xE	0.38	0.6	0.33	0.49	**0.45**	0.76	0.88	0.93	**0.86**
Multi-omics models									
EG_P	0.32	0.57	0.35	0.59	0.46	0.76	0.92	**0.94**	0.87
EGW_{cb}	0.18	0.43	0.32	0.57	0.38	**0.83**	0.91	**0.94**	0.89
EGW_{cb}_W_{cb}xE	0.37	0.61	0.35	0.64	0.49	0.82	**0.93**	**0.94**	**0.9**
EGW_{cb}_GxE	0.22	0.60	0.43	0.62	0.47	0.82	0.92	**0.94**	0.89
EGW_{cb}_GxE_W_{cb}xE	**0.4**	**0.67**	0.43	**0.66**	**0.54**	0.81	**0.93**	**0.94**	0.89

Table 3. *Predictive abilities of genomic, single NIRS, multi-NIRS and multi-omics models in scenario* ***nGoE*** *for GY and HD. CVNewG is a five-fold cross-validation scheme in which ramdom genotypes are unphenotyped in all the environments of the trial network. Predictive abilities based on 50 repetitions are presented by set and are averaged over all sets for GY or three sets for HD. The best predictive abilities by set and model category (based on Average) is shown in bold.*

nGoE scenario (CVNewG)	*GY prediction*					*HD prediction*			
Models	**Set1 2016**	**Set2 2019**	**Set4 2018**	**Set4 2019**	**Average**	**Set2 2019**	**Set4 2018**	**Set4 2019**	**Average**
Genomic models									
EG	0.25	0.40	0.24	0.37	0.32	0.36	0.42	0.52	0.43
EG_GxE	0.23	0.54	0.34	0.43	**0.38**	0.36	0.42	0.52	0.43
Single NIRS models									
EW	0.22	0.38	0.22	0.31	0.28	0.63	0.29	0.32	0.41
EW_WxE	0.30	0.46	0.25	0.37	**0.34**	0.62	0.30	0.32	0.41
Multi-NIRS models									
EW_{cb}	0.25	0.44	0.27	0.36	0.33	0.73	0.36	0.43	**0.51**
EW_{cb}_W_{cb}xE	0.39	0.59	0.30	0.43	**0.43**	0.72	0.36	0.42	0.50
Multi-omics models									
EG_P	0.38	0.59	0.31	0.46	0.44	0.43	0.4	0.54	0.46
EGW_{cb}	0.29	0.48	0.26	0.47	0.38	**0.74**	**0.58**	**0.60**	**0.64**
EGW_{cb}_W_{cb}xE	0.41	0.63	0.29	0.53	0.46	**0.74**	**0.58**	**0.60**	**0.64**
EGW_{cb}_GxE	0.29	0.61	0.34	0.52	0.44	**0.74**	0.57	0.59	0.63
EGW_{cb}_GxE_W_{cb}xE	**0.42**	**0.67**	**0.36**	**0.56**	**0.50**	**0.74**	0.57	0.59	0.63

1
2

Discussion

Phenomic prediction and genomic prediction performed similarly and were outperformed by multi-omics prediction in MET designs

We studied the PA of genomic, phenomic and multi-omics models in classic MET designs for the prediction of GY and HD. We adapted the models proposed by Robert et al. (2022a) to the MET context, by modelling the genetic and environmental effects, and in some models the GxE effect too. For the prediction of GY, we found that single-NIRS models gave slightly lower PA than multi-NIRS models. For the prediction of HD, the single-NIRS models gave much lower PA than multi-NIRS models. These results are consistent with the results found by Robert et al. (2022a), in which combining several spectra from different environments increased the PA for HD, but not necessarily for GY. We suppose that this difference is due to different effects on the variance of the trait, being mainly additive for HD and additive and interactive (G×E) for GY. Also, the relationship between spectra and HD is more likely to be similar in the different environments of the MET knowing that the phenotypic correlations were very strong between environments for this trait (Figure 2). GY is strongly influenced by the response of the plants to the specific environmental conditions in which they are grown. NIR spectra are apparently able to capture part of this response, so may be useful in predicting G×E.

We compared the phenomic to the genomic prediction models in each MET scenario. In scenarios oGoE and oGnE, there was little difference in PA of phenomic and genomic prediction models for either trait. These results are consistent with those of Krause et al. (2019) who found that predictions of GY with NIR spectra in a multi-environment context were similar to the predictions with molecular markers. In scenarios nGoE and nGnE, when the behaviour of a new genotype (nG) has to be predicted, no phenotyping is available in a MET. Phenomic prediction gave slightly better PA than genomic prediction for both traits. The genomic models appeared to be more sensitive to missing information than the phenomic models. In single environment predictions, several studies also found that PA of models using NIR spectra were similar or higher than for models using molecular markers (Rincent et al. 2018; Krause et al. 2019; Galán et al. 2020; Zhu et al. 2021b; Robert et al. 2022a). In the present study, the multi-omics models combining information from molecular markers and NIR spectra were the best models whatever the MET scenario. Krause et al. (2019) also reported that the multi-kernel model performed better than single kernel models. In addition, Galán et al. (2021) found that a bivariate model across environments, incorporating molecular markers, height and NIR spectra gave better PA than the G-BLUP or H-BLUP alone.

Modelling the GxE effect with NIR spectra improves the ability to predict grain yield

Modelling the G×E effect is challenging for complex traits like GY in multiple environments, because it is influenced by both additive genetic and GxE effects (Burgueño et al. 2012). We proposed different ways to model GxE based on NIR spectra. Our results showed that modelling GxE with molecular markers or with NIR spectra always improved the PA for GY, consistent with several reports from the past decade (Heslot et al. 2013; Jarquín et al. 2014, 2017; Lopez-Cruz et al. 2015; Lado et al. 2016; Cuevas et al. 2016; Pérez-Rodríguez et al. 2017; Ly et al. 2018; Rincent et al. 2019; Robert et al. 2020). However, for HD there was no improvement in PA, probably because the interaction effect only has a slight influence on the trait variance. This was not unexpected considering the very high correlation for HD between environments (Figure 2) and the genetic decomposition of variance (Table S1b). Indeed, all experiments in the MET were autumn sown and we do not expect a high G×E for winter wheat genotypes in these conditions.

Different covariance structures computed with NIR spectra (W_{cb}xE, W_{diag}xE and P) were used to model the GxE effect. In scenarios with observed environment (oE/oEref), W_{cb}xE , W_{diag}xE and P performed similarly. These results showed that considering all the NIR spectra in the same matrix (W_{cb}) or specifying each NIR spectra for each environment (W_{diag}) contribute the same information towards describing the WxE effect, while no information is shared between environments. By contrast, the P structure considers the covariance between each genotype-environment combination of the MET, allowing the sharing of information between environments using NIR spectra. This structure did not seem to better model the WxE effect. However, in scenarios in which a new environment is predicted (oGnE and nGnE), only the P matrix allows sharing information between environments. The P matrix can be used in these scenarios to improve the predictive ability of the model. We found that models with this GxE effect gave better PA than models without it, which confirms that NIR spectra capture part of the genetic similarity between environments.

Table 4. *Predictive abilities of the genomic, single NIRS, multi-NIRS and multi-omics models in scenario* ***oGoEref*** *for GY. CVRandom_Ref is a leave-one-environment-out scheme followed by a five-fold cross-validation in which there is sparse testing in all the environments of the MET except the reference environment. Predictive abilities based on 50 repetitions are presented for each set and as the average over all sets. The optimum models are shown in red type, and the best predictive ability by set and by model category (based on Average) in bold type.*

oGoEref scenario (CVRandom_Ref)	***GY prediction***				
Models	**Set1 2016**	**Set2 2019**	**Set4 2018**	**Set4 2019**	**Average**
Genomic models					
EG	0.12	0.31	0.32	0.55	0.32
EG_GxE	0.17	0.54	**0.46**	0.60	**0.44**
Single NIRS models					
EW_{ref}	0.11	0.27	0.27	0.38	0.26
$EW_{ref}\_W_{ref}xE$	0.24	0.35	0.24	0.40	**0.31**
Multi-NIRS models					
Optimum					
EW_{cb}	0.17	0.40	0.36	0.45	0.35
$EW_{cb}\_W_{cb}xE$	0.38	0.60	0.33	0.49	0.45
Weighted Multivariate Ridge Regression					
EW_{cb}	0.13	0.28	0.27	0.38	0.26
$EW_{cb}\_W_{cb}xE$	0.28	0.40	0.24	0.41	**0.33**
Multi-omics models					
Optimum					
EGW_{cb}_GxE	0.19	0.59	**0.46**	0.62	0.46
$EGW_{cb}\_GxE\_W_{cb}xE$	**0.39**	**0.67**	0.44	**0.66**	**0.54**
Weighted Multivariate Ridge Regression					
EGW_{cb}_GxE	0.15	0.55	0.45	0.61	0.44
$EGW_{cb}\_GxE\_W_{cb}xE$	**0.29**	**0.57**	0.44	**0.64**	**0.48**

Table 5. *Predictive abilities of genomic, single NIRS, multi-omics and multi-NIRS models in scenario* ***nGoEref*** *for GY. CVRandom_Ref and CVNewG_Ref are leave-one-environment-out schemes followed by five-fold cross-validation in which new genotypes are unphenotyped in all the environments of the MET except the reference environment. Predictive abilities calculated based on 50 repetitions are presented by set and averaged over all sets. The optimum models are shown in red, and the best predictive abilities by set and by model category (based on Average) in bold.*

nGoEref scenario (CVNewG_Ref)	***GY prediction***				
Models	**Set1 2016**	**Set2 2019**	**Set4 2018**	**Set4 2019**	**Average**
Genomic models					
EG	0.25	0.39	0.24	0.38	0.32
EG_GxE	0.23	0.54	0.34	0.44	**0.39**
Single NIRS models					
EW_{ref}	0.16	0.29	0.19	0.30	0.23
$EW_{ref}\_W_{ref}xE$	0.25	0.35	0.20	0.37	**0.29**
Multi-NIRS models					
Optimum					
EW_{cb}	0.25	0.44	0.27	0.36	0.33
$EW_{cb}\_W_{cb}xE$	0.40	0.60	0.29	0.43	**0.43**
Weighted Multivariate Ridge Regression					
EW_{cb}	0.19	0.31	0.17	0.30	0.24
$EW_{cb}\_W_{cb}xE$	0.29	0.40	0.19	0.37	**0.31**
Multi-omics models					
Optimum					
EGW_{cb}_GxE	0.29	0.61	0.33	0.52	0.44
$EGW_{cb}\_GxE\_W_{cb}xE$	**0.43**	**0.67**	0.33	**0.56**	**0.50**
Weighted Multivariate Ridge Regression					
EGW_{cb}_GxE	0.24	0.56	**0.33**	0.48	0.40
$EGW_{cb}\_GxE\_W_{cb}xE$	**0.33**	**0.57**	0.31	**0.51**	**0.43**

The PA were not too low (0.32 and 0.27, Tables S4-S5) considering the difficulty of predicting GY in new environments. For some sets, the PA markedly increased when the P matrix was used to model GxE. For example, the PA of model EG was 0.03 for Set2-2019 while the PA of model EG_P was 0.27 (Table S3). We found that using a single NIRS model was less effective than using all spectra available to model GxE. The main genetic effect can be modelled by an H matrix computed on a single environment, but GxE is better modelled by combining all the NIR spectra available. The optimal situation for predicting GxE is therefore to have NIR data for each predicted environment.

Is it better to model GxE with information from NIR spectra or from molecular markers? Unlike molecular markers, the spectra characterise the effects of both the genotype and the environment on the plant material tested, so they may provide more information. After trying different combinations to model the main genetic effect with molecular markers (G) or NIRS (W) and the GxE effect with molecular markers (GxE) or NIRS (P), we found that the best combination was G for the main genetic effect and P for the GxE effect. However, the increase in PA is moderate. Similar conclusions were reported by Krause et al. (2019) for predicting GY in wheat. Montesinos-López et al. (2017) also reported that modelling GxE by the interaction between spectral band and the environment (BxE) resulted in higher PA than models without this term. To assess the benefit of using NIR spectra to model the GxE effect, larger datasets with numerous and contrasted environments would need to be studied. The size of our dataset, and in particular the low number of environments clearly made difficult the sharing of information between environments.

Our assumption here was that NIR spectra can be considered as simple environmental covariates for estimating the similarity between environments. They are indeed the results of the complex regulation and expression of genes in a specific environment, as confirmed by the large amount of GxE along the spectra (Figure 3). However, the environmental factors responsible for the absorbance variance are not necessarily the same as the ones responsible for the GY variance. It might be informative to select wavelengths which capture most of the GxE variance for the GY trait. For example, to weight the contribution of the different wavelengths, Additive Main effects and Multiplicative Interaction (AMMI) decomposition can determine the markers and ECs important for predicting the interactions (Rincent et al., 2019).

Original MET designs enable breeders to use PS at minimal cost

To extend the application of PS for breeders, we compared different MET scenarios with different assumptions. oGoE, nGoE, oGnE and nGnE are classic MET scenarios in which GS has been applied. We first tested our models in a sparse testing scenario (oGoE) and in a scenario where a new line would not have been phenotyped in the MET (nGoE). It is acknowledged that the datasets are very small, both in terms of the number of environments and the number of breeding lines. Despite this, we demonstrate the advantage of MET prediction by PS in the high PA for GY and HD achieved by models developed in oGoE and nGoE. We then tested our models in more challenging scenarios where no genotype was phenotyped in the environment of prediction (oGnE) and new genotypes were not phenotyped in the MET (nGnE). As expected, PA were lower than the previous scenarios. Similarly, Lane et al. (2020) found that for a CV0 scenario equivalent to oGnE, the PA of a multi-NIRS model including the GxE interaction for maize GY was around 0.54. The maize trial covered 4 environments (two years x two treatments) and around 300 hybrids. For wheat, we found similar PA for GY with the largest dataset, Set4-2019, but this level of accuracy was not attained with the sets with fewer genotypes (Set4-2018) or with just two environments (Set1-2016 and Set2-2019). This may be partly because the environmental conditions of the unobserved environment did not fall within the range of those of the training environments (Jarquín et al. 2017). For Set1-2016 and Set2-2019, the correlation of GY between environments is close to null (Figure 2), and the prediction of GY cannot be accurate when the training and validating environment differ so much.

In designing these scenarios, we suppose that breeders acquire NIR spectra on all lines (training and validation) in each environment of the MET. Practically this could be done by dedicating nursery rows in each environment to observation (i.e. scoring diseases and lodging) and NIR spectra acquisition of all lines, in parallel to trial plots dedicated to the measurement of GY of the training lines. This is feasible if the grains or tissue samples are in a good state with no deterioration due to biotic or abiotic factors. Even in the event of disease, we assume that biotic stress will influence the phenotype of the plant and the NIR spectra, but it would still be possible to compute the covariance between genotypes. In our study of Set2_2019, the two environments are different sites with intensive management practices, with the difference that Genlis was treated with fungicide but Estrée-Mons was not. The PA reached by our models for this dataset were the highest for GY in most scenarios (Tables 2-5). For an easier and cheaper application of PS, we also proposed two new experimental designs (oGoEref and nGoEref) in which NIR spectra are acquired only on the plots evaluated for GY, which means there is no need for a nursery in parallel to the trials. The principle is to observe and acquire NIRS from all lines in one environment. This environment serves as the reference which is usually chosen by the breeder to observe all the candidate lines for selection.

Scenario oGoEref resembles a sparse testing scenario and scenario nGoEref resembles a new genotype scenario, except that they use one environment as reference to acquire NIR spectra on all genotypes.

To compare multi-NIRS models and multi-omics models in these scenarios, we had to impute the missing spectra. Our results showed that WMRR gave better PA for GY than the NN imputation method (Table 4, Table 5). This may be because NN assumes that if two lines have a similar spectrum in one environment, then they will have similar spectra in the other environments. This is clearly not in accordance with the observation of strong GxE across the spectra (Figure 3). Two genotypes can have similar spectra in one environment and dissimilar spectra in another due to the GxE effect, which is the reason why they are able to capture GxE of the predicted trait more efficiently than markers. Inversely WMRR was expected to consider GxE in the spectra by predicting the specific spectrum of a specific genotype in a specific environment, and thus better impute the missing spectra. The expected maximum PA for GY ("optimum" scenario) were however still not reached when the WMRR method was used. The correlations between predicted and acquired spectra were very variable from one set to another, but for Set4-2019 the correlation across spectra was high (Figures S3, FigureS4). Set4-2019 was the only set for which the best model using predicted spectra almost reached the PA of the optimum (0.64 instead of 0.66 in Table 4 for CVRandom_Ref and 0.51 instead of 0.56 in Table 5 for CVNewG_Ref). The objective with the WMRR method was not to accurately predict the entire spectra, but to accurately predict the wavelengths indicative of major effects on the target trait. However, all the wavelengths are somewhat explicative, so it was not possible to isolate distinct wavelengths or peaks. PA for GY were slightly higher compared to an unweighted multivariate ridge regression method (not shown here).

The WMRR method is promising because it circumvents the potential obstacle of acquiring NIR spectra on all genotypes in each environment. It would be more convenient for breeders to directly apply PS in already defined MET designs and without any additional cost. The PA obtained with WMRR were however lower than those obtained with the GS model in our study. There is scope to improve the method as the optimum has not been reached yet. Two factors which could be adjusted is the size and the composition of the training set for predicting missing spectra. As we know, these factors influence the PA of different traits (Albrecht et al. 2011; Heffner et al. 2011; Pszczola et al. 2012; Rincent et al. 2012; Daetwyler et al. 2013) in GS as well as in PS (Parmley et al. 2019; Galán et al. 2020; Zhu et al. 2021a; Robert et al. 2022a). We assume that the prediction of the absorbances will also be influenced by these factors. Other methods of imputation could be considered. We tried the single-step method of Legarra et al. (2009) to overcome the issue of incomplete NIRS acquisition. We used the NIR spectra acquired in the environment of reference combined with the incomplete NIR spectra in another environment to have a hybrid covariance matrix between all genotypes (Michel et al. 2021). The hybrid matrix has the advantage of being quick to compute. However, PA obtained with the hybrid matrix were slightly lower compared to PA of our WMRR method.

Phenomic selection is a promising tool to improve breeding programs

More and more studies have tried to apply PS to annual crops like bread wheat (Montesinos-López et al. 2017; Rincent et al. 2018; Krause et al. 2019; Cuevas et al. 2019; Robert et al. 2022a), maize (Lane et al. 2020), rye (Galán et al. 2020, 2021), triticale (Zhu et al. 2021b) or soybean (Parmley et al. 2019; Zhu et al. 2021a), or perennial crops like poplar (Rincent et al. 2018) and recently grapevine (Brault et al. 2021). PS has the advantage, compared to other types of predictors, of being more flexible as to which data acquired. NIR spectra can be collected on different tissues (leaves, wood, grain, fruits, roots) at multiple stages of the growing cycle, which is similar to other omics methods but less costly. Conventionally, NIR spectra are collected in controlled conditions in a laboratory, with coverage of wavelengths from the visible and near infrared (Osborne 2006). With the rapid development of high throughput phenotyping and new sensors, NIR spectra can be collected at moderate to low cost directly in the fields with an unoccupied aerial vehicle, portable (micro)spectrometer, or directly in the harvester. The experimental protocol for acquiring NIR spectra has clearly to be further studied for an optimal use of PS. It is still not clear which spectrophotometer and on which tissue it is best to work with. The method is also very straightforward for breeders who are used to the prediction models already developed for GS. Multiple applications are possible in breeding like sparse testing in single environments (Rincent et al. 2012; Cuevas et al. 2019; Robert et al. 2022a) or in MET designs as we showed here. PS is also useful to predict breeding material genetically distant from the training set (Galán et al. 2021). Other avenues do not have been explored yet but have great potential to interest breeders (Opinion and review in Robert et al. 2022b) like the screening of diversity collections at lower cost, or the prediction of breeding material at early stages in nurseries.

Conclusion

We explored whether PS can be used by breeders in different multi-environmental contexts to predict GY and HD. Our main results showed that the PS models performed as well as GS models in the prediction of GY for classic MET designs. For the prediction of HD, GS performed slightly better than PS because HD variance is mainly additive. We compared different ways to model GxE and found that models integrating the GxE effect modelled with NIR spectra were more accurate than the models without this term. We incorporated a P matrix characterising the covariance between the different genotype-environment combinations of the MET, which effectively shares information between environments. For scenarios with new environments, this gave better predictions than models without the P matrix. Finally, we designed two new MET scenarios specific to PS in which NIR spectra were missing for the genotypes to be predicted. Among the different methods available for spectra imputation, the WMRR gave the best PA but not the optimum. The models developed in this paper as well as the original MET designs would enable breeders to use PS for diverse objectives and at minimal cost.

Acknowledgements

The authors thank INRAE experimental units (UE PHACC Clermont-Ferrand, UE GCIE Estrées-Mons, IE UMR GQE Le Moulon, UE Domaine de la Motte Rennes), breeders from Agri-Obtentions and Florimond Desprez. The authors are grateful to Agri-Obtentions, Florimond-Desprez, and the Association Nationale de la Recherche et de la Technologie (ANRT, grant number 2019/0060) which supported this PhD work. The authors also thank Rachel Carol (Bioscience Editing, France) for proofreading.

References

Albrecht T, Wimmer V, Auinger H-J, et al (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350. https://doi.org/10.1007/s00122-011-1587-7

Allard RW, Bradshaw AD (1964) Implications of Genotype-Environmental Interactions in Applied Plant Breeding [1]. Crop Sci 4:503–508. https://doi.org/10.2135/cropsci1964.0011183X000400050021x

Azodi CB, Pardo J, VanBuren R, et al (2020) Transcriptome-Based Prediction of Complex Traits in Maize. Plant Cell 32:139–151. https://doi.org/10.1105/tpc.19.00332

Bernardo R (1994) Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. Crop Sci 34:20–25. https://doi.org/10.2135/cropsci1994.0011183X003400010003x

Brault C, Lazerges J, Doligez A, et al (2021) Interest of phenomic prediction as an alternative to genomic prediction in grapevine. Genetics

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. Crop Sci 52:707–719. https://doi.org/10.2135/cropsci2011.06.0299

Comstock RE, Moll RH (1963) Genotype x Environment Interactions. Stat Genet Plant Breed

Covarrubias-Pazaran G (2016) Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLOS ONE 11:e0156744. https://doi.org/10.1371/journal.pone.0156744

Crossa J, Campos G de los, Pérez P, et al (2010) Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. Genetics 186:713–724. https://doi.org/10.1534/genetics.110.118521

Cuevas J, Crossa J, Soberanis V, et al (2016) Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. Plant Genome 9:plantgenome2016.03.0024. https://doi.org/10.3835/plantgenome2016.03.0024

Cuevas J, Montesinos-López O, Juliana P, et al (2019) Deep Kernel for Genomic and Near Infrared Predictions in Multi-environment Breeding Trials. G3amp58 GenesGenomesGenetics 9:2913–2924. https://doi.org/10.1534/g3.119.400493

Daetwyler HD, Calus MPL, Pong-Wong R, et al (2013) Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. Genetics 193:347–365. https://doi.org/10.1534/genetics.112.147983

Damesa T, Worku M, Möhring J, and Piepho, HP (2017) One step at a time: stage-wise analysis of a series of experiments. Agron. J. 109, 845–857. doi: 10.2134/agronj2016.07.0395

De Los Campos G, Gianola D, Rosa GJM, et al (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet Res 92:295–308. https://doi.org/10.1017/S0016672310000285

de los Campos G, Naya H, Gianola D, et al (2009) Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree. Genetics 182:375–385. https://doi.org/10.1534/genetics.109.101501

Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4:250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Endelman JB, Jannink J-L (2012) Shrinkage Estimation of the Realized Relationship Matrix. G3 GenesGenomesGenetics 2:1405–1413. https://doi.org/10.1534/g3.112.004259

Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet Sel Evol 43:1. https://doi.org/10.1186/1297-9686-43-1

Friedman J, Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 33:. https://doi.org/10.18637/jss.v033.i01

Frisch M, Thiemann A, Fu J, et al (2010) Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor Appl Genet 120:441–450. https://doi.org/10.1007/s00122-009-1204-1

Fu J, Falke KC, Thiemann A, et al (2012) Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene expression data. Theor Appl Genet 124:825–833. https://doi.org/10.1007/s00122-011-1747-9

Galán RJ, Bernal-Vasquez A-M, Jebsen C, et al (2020) Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. Theor Appl Genet 133:3001–3015. https://doi.org/10.1007/s00122-020-03651-8

Galán RJ, Bernal-Vasquez A-M, Jebsen C, et al (2021) Early prediction of biomass in hybrid rye based on hyperspectral data surpasses genomic predictability in less-related breeding material. Theor Appl Genet 134:1409–1422. https://doi.org/10.1007/s00122-021-03779-1

Gemmer MR, Richter C, Jiang Y, et al (2020) Can metabolic prediction be an alternative to genomic prediction in barley? PLOS ONE 15:e0234052. https://doi.org/10.1371/journal.pone.0234052

Guo Z, Magwire MM, Basten CJ, et al (2016) Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. Theor Appl Genet 129:2413–2427. https://doi.org/10.1007/s00122-016-2780-5

Heffner EL, Jannink J-L, Iwata H, et al (2011) Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. Crop Sci 51:2597–2606. https://doi.org/10.2135/cropsci2011.05.0253

Heslot N, Akdemir D, Sorrells ME, Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127:463–480. https://doi.org/10.1007/s00122-013-2231-5

Heslot N, Jannink J-L, Sorrells ME (2013) Using Genomic Prediction to Characterize Environments and Optimize Prediction Accuracy in Applied Breeding Data. Crop Sci 53:921–933. https://doi.org/10.2135/cropsci2012.07.0420

Jarquín D, Crossa J, Lacaze X, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607. https://doi.org/10.1007/s00122-013-2243-1

Jarquín D, Lemes da Silva C, Gaynor RC, et al (2017) Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype × Environment Interactions in Kansas Wheat. Plant Genome 10:plantgenome2016.12.0130. https://doi.org/10.3835/plantgenome2016.12.0130

Kang HM, Sul JH, Service SK, et al (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354. https://doi.org/10.1038/ng.548

Krause MR, González-Pérez L, Crossa J, et al (2019) Hyperspectral Reflectance-Derived Relationship Matrices for Genomic Prediction of Grain Yield in Wheat. G3amp58 GenesGenomesGenetics g3.200856.2018. https://doi.org/10.1534/g3.118.200856

Lado B, Barrios PG, Quincke M, et al (2016) Modeling Genotype × Environment Interaction for Genomic Selection with Unbalanced Data from a Wheat Breeding Program. Crop Sci 56:2165–2179. https://doi.org/10.2135/cropsci2015.04.0207

Lane HM, Murray SC, Montesinos-López OA, et al (2020) Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. Plant Phenome J 3:. https://doi.org/10.1002/ppj2.20002

Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. J Dairy Sci 92:4656–4663. https://doi.org/10.3168/jds.2009-2061

Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased Prediction Accuracy in Wheat Breeding Trials Using a Marker × Environment Interaction Genomic Selection Model. G3 GenesGenomesGenetics 5:569–582. https://doi.org/10.1534/g3.114.016097

Ly D, Huet S, Gauffreteau A, et al (2018) Whole-genome prediction of reaction norms to environmental stress in bread wheat (Triticum aestivum L.) by genomic random regression. Field Crops Res 216:32–41. https://doi.org/10.1016/j.fcr.2017.08.020

Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA (2016) Predicting Responses in Multiple Environments: Issues in Relation to Genotype × Environment Interactions. Crop Sci 56:2210. https://doi.org/10.2135/cropsci2015.05.0311

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Michel S, Wagner C, Nosenko T, et al (2021) Merging Genomics and Transcriptomics for Predicting Fusarium Head Blight Resistance in Wheat. Genes 12:114. https://doi.org/10.3390/genes12010114

Montesinos-López A, Montesinos-López OA, Cuevas J, et al (2017) Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. Plant Methods 13:. https://doi.org/10.1186/s13007-017-0212-4

Osborne BG (2006) Applications of near Infrared Spectroscopy in Quality Screening of Early-Generation Material in Cereal Breeding Programmes. J Infrared Spectrosc 14:93–101. https://doi.org/10.1255/jnirs.595

Parmley K, Nagasubramanian K, Sarkar S, et al (2019) Development of Optimized Phenomic Predictors for Efficient Plant Breeding Decisions Using Phenomic-Assisted Selection in Soybean. Plant Phenomics 2019:1–15. https://doi.org/10.34133/2019/5809404

Pérez P, de los Campos G (2014) Genome-Wide Regression and Prediction with the BGLR Statistical Package. Genetics 198:483–495. https://doi.org/10.1534/genetics.114.164442

Pérez-Rodríguez P, Crossa J, Rutkoski J, et al (2017) Single-Step Genomic and Pedigree Genotype × Environment Interaction Models for Predicting Wheat Lines in International Environments. Plant Genome 10:plantgenome2016.09.0089. https://doi.org/10.3835/plantgenome2016.09.0089

Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. J Dairy Sci 95:389–400. https://doi.org/10.3168/jds.2011-4338

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217

Rimbert H, Darrier B, Navarro J, et al (2018) High throughput SNP discovery and genotyping in hexaploid wheat. PLOS ONE 13:e0186329. https://doi.org/10.1371/journal.pone.0186329

Rincent R, Charpentier J-P, Faivre-Rampant P, et al (2018) Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar. G3amp58 GenesGenomesGenetics g3.200760.2018. https://doi.org/10.1534/g3.118.200760

Rincent R, Laloë D, Nicolas S, et al (2012) Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (*Zea mays* L.). Genetics 192:715–728. https://doi.org/10.1534/genetics.112.141473

Rincent R, Malosetti M, Ababaei B, et al (2019) Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. Theor Appl Genet 132:3399–3411. https://doi.org/10.1007/s00122-019-03432-y

Robert P, Auzanneau J, Goudemand E, et al (2022a) Phenomic selection in wheat breeding: identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. Theor Appl Genet. https://doi.org/10.1007/s00122-021-04005-8

Robert P, Brault C, Rincent R, Segura V (2022b) Phenomic Selection: A New and Efficient Alternative to Genomic Selection. In: Ahmadi N, Bartholomé J (eds) Complex Trait Prediction. Springer US, New York, NY, pp 397–420. https://doi.org/10.1007/978-1-0716-2205-6_14

Robert P, Le Gouis J, The BreedWheat Consortium, Rincent R (2020) Combining Crop Growth Modeling With Trait-Assisted Prediction Improved the Prediction of Genotype by Environment Interactions. Front Plant Sci 11:827. https://doi.org/10.3389/fpls.2020.00827

Schrag TA, Westhues M, Schipprack W, et al (2018) Beyond Genomic Prediction: Combining Different Types of *omics* Data Can Improve Prediction of Hybrid Performance in Maize. Genetics 208:1373–1385. https://doi.org/10.1534/genetics.117.300374

Wang S, Wei J, Li R, et al (2019) Identification of optimal prediction models using multi-omic data for selecting hybrid rice. Heredity 123:395–406. https://doi.org/10.1038/s41437-019-0210-6

Westhues M, Schrag TA, Heuer C, et al (2017) Omics-based hybrid prediction in maize. Theor Appl Genet 130:1927–1939. https://doi.org/10.1007/s00122-017-2934-0

Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. Genet Res 75:249–252. https://doi.org/10.1017/S0016672399004462

Xu S, Xu Y, Gong L, Zhang Q (2016) Metabolomic prediction of yield in hybrid rice. Plant J 88:219–227. https://doi.org/10.1111/tpj.13242

Zenke-Philippi C, Frisch M, Thiemann A, et al (2017) Transcriptome-based prediction of hybrid performance with unbalanced data from a maize breeding programme. Plant Breed 136:331–337. https://doi.org/10.1111/pbr.12482

Zhu X, Leiser WL, Hahn V, Würschum T (2021a) Phenomic selection is competitive with genomic selection for breeding of complex traits. Plant Phenome J 4:. https://doi.org/10.1002/ppj2.20027

Zhu X, Maurer HP, Jenz M, et al (2021b) The performance of phenomic selection depends on the genetic architecture of the target trait. Theor Appl Genet. https://doi.org/10.1007/s00122-021-03997-7

Figure captions

Fig. 1 Schematic overview of the six prediction scenarios representing the associated missing data to be predicted in the MET. G_i corresponds to the genotype i with $i \in \{1, \dots, N_G\}$ and E_j corresponds to the environment j of the MET, with $j \in \{1, \dots, N_E\}$. Filled squares represent data of the training set and crosses represent data of the validation set. Blue shading indicates that a phenotype and a NIR spectrum were acquired from genotype i in environment j. Yellow shading indicates that only the NIR spectrum was acquired, not the phenotyping. Finally, orange shading indicates that neither phenotyping nor NIR spectrum were acquired. **oGoE** corresponds to prediction in a sparse testing scenario where some combinations of genotypes and environments are not phenotyped. **oGnE** corresponds to a new environment in which no phenotype is available. **nGoE** corresponds to the prediction of new genotypes in observed environments. **nGnE** corresponds to new genotypes to be predicted in a new environment. **oGoEref** and **nGoEref** correspond to two scenarios where one environment is considered as a reference where NIR spectrum is acquired on all genotypes. **oGoEref** therefore is similar to scenario **oGoE** except that NIR spectra are not acquired on the validation set, while **nGoEref** is similar to scenario **nGoE** except that NIR spectra are not acquired on the validation set apart from in the reference environment.

Fig2. Correlation coefficient matrices between trial environments for heading date adjusted means (upper left) and grain yield adjusted means (lower right). A trial environment is defined as a combination of treatment × year × site. Treatments were denoted by T, treated (equivalent to intensive practices) or LI, low input. Sites were denoted by EM (Estrée-Mons), GL (Genlis), HV (Houville), LC (Lectoure), LM (Le Moulon). Asterisks indicate the significance level: *P-value <0.05, **P-value <0.01 and ***P-value <0.001.

Fig 3. Proportion of the genomic (G), the genomic x environment (GxE) and residual variances across the NIR spectra of winter wheat grains from 5 different MET sets.