ENVIROME-WIDE ASSOCIATIONS ENHANCE MULTI-YEAR GENOME-BASED PREDICTION OF HISTORICAL WHEAT BREEDING DATA

Germano Costa-Neto[1], Leonardo Crespo-Herrera[2], Nick Fradgley[3], Keith Gardner[2], Alison R. Bentley[2], Susanne Dreisigacker[2], Roberto Fritsche-Neto[4], Osval A. Montesinos-López[5,*], and Jose Crossa[2,6,*]

Affiliations:

[1] Institute for Genomics Diversity, Cornell University, Ithaca, NY, 14853, USA.

[2] International Maize and Wheat Improvement Center (CIMMYT). Km 45 Carretera México-Veracruz, El Batan, 5623, Edo. de México, México.

[3] NIAB, 93 Lawrence Weaver Road, Cambridge, CB3 0LE, United Kingdom.

[4] Louisiana State University, Baton Rouge, 70803, USA.

[5] Facultad de Telemática, Universidad de Colima, Colima, 28040, México.

[6] Colegio de Postgraduados, Montecillo, 56231, Edo. de México, México.

ABSTRACT

Linking high-throughput environmental data (enviromics) to genomic prediction (GP) is a cost-effective strategy for increasing selection intensity under genotype-by-environment interactions (G×E). This study developed a data-driven approach based on Environment-Phenotype Associations (EPA) aimed at recycling important G×E information from historical breeding data. EPA was developed in two applications: (1) scanning a secondary source of genetic variation, weighted from the shared reaction-norms of past-evaluated genotypes; (2) pinpointing weights of the similarity among trial-sites (locations), given the historical impact of each envirotyping data variable for a

given site. These results were then used as a dimensionality reduction strategy, integrating historical data to feed multi-environment GP models, which led to development of four new G×E kernels considering genomics, enviromics and EPA outcomes. The wheat trial data used included 36 locations, eight years and three target populations of environments (TPE) in India. Four prediction scenarios and six kernel-models within/across TPEs were tested. Our results suggest that the conventional GBLUP, without enviromic data or when omitting EPA, is inefficient in predicting the performance of wheat lines in future years. Nevertheless, when EPA was introduced as an intermediary learning step to reduce the dimensionality of the G×E kernels while connecting phenotypic and environmental-wide variation, a significant enhancement of G×E prediction accuracy was evident. EPA revealed that the effect of seasonality makes strategies such as "covariable selection" unfeasible because G×E is year-germplasm specific. We propose that the EPA effectively serves as a "reinforcement learner" algorithm capable of uncovering the effect of seasonality over the reaction-norms, with the benefits of better forecasting the similarities between past and future trialing sites. EPA combines the benefits of dimensionality reduction while reducing the uncertainty of genotype-by-year predictions and increasing the resolution of GP for the genotype-specific level.

Key words: genomic selection; climate change; wheat breeding; envirotyping; adaptability; target populations of environments

INTRODUCTION

Traditionally, the importance of genotype by environment interactions (G×E) has been seen as problematic source of variation that must be taken into consideration for plant breeding decisions. As such, many computational tools and statistical models have been developed and used for studying G×E in diverse scenarios and contexts (Crossa et al., 2017). Essentially, the goal of any G×E analytics is to enable plant breeders to assess the stability of target traits under multiple-environment trial (MET) conditions. Since the introduction of predictive breeding strategies, such as genomic prediction (GP),

these goals must also support the prediction of the phenotypic performance of newly developed cultivars, which are mostly evaluated under a pool of selected environmental conditions – also known as the target population of environments (TPE) of breeding programs. Although MET are useful for selecting stable and high yielding varieties (Crossa et al., 2004) for the environments included in the trialing network, they are less suitable for providing targeted recommendations for specific environments (e.g., site and year combinations), especially those where no field trials were performed.

Since the early 1930s, a diverse set of statistical methods have been developed to incorporate additional (co)variables as a proxy of the environmental causes of G×E. This rich history of "analytical" methods started with the traditional regression (or reaction-norm) models of Yates and Cochran (1938); Finlay and Wilkinson (1963); and Eberhart and Russell (1966) describe environments based on the mean performance of cultivars in the test environments. However, instead of regressing the mean grain yield of cultivar performance on the mean grain yield of the environments (environmental quality index), environmental covariates can be used to characterize both (1) the environments in the MET and also (2) new target environments (Hardwick and Wood, 1972) that were not included in the MET trials.

Classical approaches were most used to interpret G×E rather than to predict it. This can be seen in the use of fixed-effect linear-bilinear models, with the most used being the Sites Regression (SREG) (Crossa and Cornelius, 1997) and the Additive Main effect and Multiplicative Interaction (AMMI) (Gauch, 1988; Cornelius et al., 1996). Essentially, these models are generalizations of single regressions on environmental data – aimed to compute genotypes reaction-norms and in this way, study the genotypic response patterns across environments. In these models, the response patterns of genotypes and environments can be visualized graphically using biplots that allow the breeder to observe the high performing genotype(s) in a region(s) and/or sub-region(s). Partial least squares (PLS) regression is also a type of fixed-effect linear model that allows external environmental and genotypic covariables to be directly incorporated into the model and has been demonstrated to be useful at identifying the climatic causes of G×E or the genetic factors (e.g., molecular markers) influencing G×E (Aastveit and

Martens, 1986; Helland, 1988; Vargas et al., 1998; 1999; Montesinos-Lopez et al., 2022). In fact, the first mention of the use of high-throughput environmental data ("enviromics") in plant science involved the use of PLS for scanning mechanisms acting in the internal cell environment (Teixeira et al., 2011). This concept was recently popularized by some seminal papers such as Resende et al. (2021) and Costa-Neto et al. (2021b), who mostly addressed the benefits of using enviromics to aid genomic prediction based on prior studies involving the so-called reaction-norm GBLUP (Jarquin et al., 2014; Morais-Junior *et al*., 2018).

As modern plant breeding is an interdisciplinary field based on multi-dimensional data types, the importance of developing new enviromics pipelines is key for successful prediction of G×E (de los Campos et al., 2020; Crossa et al., 2021). In the field of genomics, for instance, the availability of high density, low-cost genetic markers has made it possible to saturate the genome with these markers and predict estimated breeding values (Bassi et al., 2015). This has increased the precision of genetic value prediction compared to that achieved with traditional pedigree information. Genomic data can also help assess chromosome regions, e.g., marker effects and patterns of (co)variability of marker effects linked to different environmental conditions. Since the analysis of genetic and genomic data is one of the most challenging statistical problems currently faced, different models from diverse areas of statistical research need to be integrated in order to make significant progress in understanding genetic effects and their interaction with environments. In addition, new environmental sensors and remote sensing technologies (Morisse et al., 2021) permit the real-time, dynamic collection of environmental covariables with very high resolution anywhere in the world. Thus, marker technology combined with the rapid digitalization of high density environmental covariables (EC) and in-season image capture offers a new set of opportunities for improving the performance of MET for greater targeting of plant breeding (Crossa et al., 2021).

Jarquín et al. (2014) and Heslot et al. (2014) introduced the concept of genome-based reaction norms to model G×E using many environmental covariates (e.g., weather

data, soil properties, crop growth model outcomes). Since then, a new wave of studies focused on the prediction of unobserved genotypes in new untested environments (see Crossa et al., 2022). The so-called linear kernel-based reaction norm model of Jarquin et al. (2014) was utilized by Pérez-Rodríguez et al. (2015), employing pedigree and environmental covariates in cotton trials. Similar substantial gains in prediction accuracy of genotype performance were obtained by Cuevas et al. (2016; 2017; 2018; 2019), He et al. (2019) and Costa-Neto et al. (2021a) using non-linear genomic Gaussian kernel models for modeling G×E interaction over the conventional linear GBLUP kernel. However, to date, although these studies have focused on genomic-enabled predictions, they have neglected two important components: first, the uncertainty of predictions in new environments, which can be substantial, as demonstrated by de los Campos et al. (2020); and second, the detailed enviromics assessment needed for the quantification of soil and climatic variables. Both factors are necessary for studying G×E and for the prediction of genomic estimated breeding values under different environmental conditions.

To address this situation, we made some efforts in the past to develop public envirotyping pipelines, such as *EnvRtype* (Costa-Neto et al., 2021b), tested the merit of nonlinear kernels for modeling the environmental similarities in genomic prediction (Costa-Neto et al., 2021a), and designed new "environmental markers" to add value in the prediction platforms (Costa-Neto et al., 2021c) by increasing the resolution of the predictive models in reproducing the phenotypic plasticity of the plants, while supporting the optimization of training sets with less phenotyping effort. Other authors have also addressed this situation by including crop growth models as a supervised machine learning tool, deep learning approaches using long-term field trial data and using robust experimental trialing networks for training reaction-norm models using in-field sensors. Despite the benefits of all these approaches, the lack of future environmental data is still a bottleneck in predicting yet-to-be-seen genotype-by-year interactions (G×Y). For "proof-of-concept" studies, it is easy to simulate "a new year" because we already have the environmental data observed in the past; however, for real-world breeding programs, we don't have any environmental covariate for a yet-to-be-

seen year. Because of this, in the current study, we focused on developing new tools for connecting past and future envirotyping data in order to reduce the uncertainty of G×Y predictions while using enviromics to add value in the current cultivar testing pipelines.

To achieve this goal, we returned to an old idea of connecting environmental and phenotyping data to interpret G×E interactions, considering a particular season and germplasm pool. To facilitate the interpretation of the concept, hereinafter we call it "environmental-phenotype association" (EPA). The theoretical basis of the EPA modeling approach assumes the genotypic response over the environmental variation, and consequently the emergent G×E interaction observed in the experimental network is strictly related to the envirotype-to-phenotype dynamics observed for each genotype and its evaluated growing conditions. Thus, the core of the reaction-norm not only determines the main drivers of the G×E (Millet et al., 2019; Costa-Neto et al., 2021a; Porker et al., 2020; Heinemann et al., 2022), but also explains how the environmental conditions have shaped the phenotypic correlations among experimental sites. Because of this property, here we used EPA as an intermediary step to adapt the envirotyping data (covariables) into environmental weights of similarity, while also exploring genotype responses for those factors (that is, reaction-norms) as a secondary source of genetic variation to approach genomics and the observed G×E. This analysis was also interpreted as a "reinforcement learning algorithm" capable of recycling past EPA as a precursor for future growing conditions and phenotypic responses.

To test this hypothesis, we analyzed a real-world breeding program and its intrinsic complexities observed in experimental trials (e.g., unbalanced conditions and diverse sets of locations across years). Within the CIMMYT Global Wheat Breeding Program, recent efforts have been made to describe, measure, and analyze G×E in MET tested breeding materials across major wheat growing areas around the world. This has shown that performance in CIMMYT's main breeding and testing location in Mexico is correlated with various international sites belonging to TPE (or mega-environments, in this case) that represent the world's major wheat producing areas (Rajaram et al., 1994; Braun et al., 1996; Crespo-Herrera et al., 2021). The delineation of TPE are based on climate, soil and hydrological characteristics and can also include socioeconomic factors,

such as the resource levels of farm households. There are also different ways to group trials and environments into a TPE. Data from environmental sensors and satellites can be used to develop stratified hierarchical cluster analyses of sites and thus identify homogeneous environments wherein line performances will be highly correlated (Ornellas et al., 2019; Crespo-Herrera et al., 2021). A TPE can be defined as a group of production environments that can be utilized for breeding in future years and/or in variable growing seasons, it is expected that the G×E may result from relatively static and predictable variation—for example, in soil or other conditions across the field—and dynamic, unpredictable, and often significant temporal variability—i.e., weather over different years. Thus, here we also check the merit of EPA as a data-driven approach in identifying those mega-environments and pinpoint the "essential trialing sites" within each environmental group, that is, identify the number of locations that represents the maximum diversity of growing conditions for each TPE (or across TPEs).

Based on the previous considerations the objectives of this study were to (1) demonstrate the benefits of the environmental-wide association analysis (EPA) combining historical phenotypic and envirotyping data for purposes of predictive breeding (e.g., multi-environment genomic prediction) and analytics (e.g., identifying essential trialing sites); (2) present three new multi-environment GBLUP models combining diverse EPA outcomes and compare them with benchmark approaches with and without envirotyping data; (3) discuss the limits and the potential research gaps for predicting yet-to-be-seen years in multi-environmental trials in predictive breeding; (4) discuss goals (1-3) in terms of its applications for a real world breeding program for wheat in India.

METHODS

CIMMYT historical wheat data

We used grain yield (GY, Mg ha^{-1}) as the reference trait in this study. Data were collected from 2011–2018 crop cycles of Elite Spring Wheat Yield Trials (ESWYT) nurseries carried out in India and based on a previous and extensive study done by

Crespo-Herrera et al. (2021) that aimed to define TPEs for CIMMYT wheat breeding in India (see Supplementary Figure S1). Here, we considered 370 wheat lines and 130 environments, which involved 36 locations and three TPEs: TPE 1, the optimally irrigated Northwestern Plain Zone; TPE 2, the optimally irrigated, heat-stressed Northeastern Plains Zone; and TPE 3, the drought-stressed Central-Peninsular Zone. Details about the data involved in the experimental analysis, phenotype correction criteria and genotyping are given in Crespo-Herrera et al. (2021). Due to the highly unbalanced conditions (different genotypes across years and years with different locations), this data set is ideal for testing the prediction of untested genotypes in yet-to-be-seen years in a real-world breeding program. Details about how this trialing complexity was explored in different cross-validation scenarios are presented in subsequent sections.

Envirotyping protocol

Here we considered "location" as a certain trialing site (geographical location), while "environment unit", or simply "environment", was considered as a combination of certain location × year × management. In this study, we considered only macroenvironmental variations (those variations for an entire field experiment, rather than in-plot or plant specific microenvironments). Thus, considering a historical breeding data, a certain location experienced a diverse set of macroenvironments across the years, and receiving a diverse set of genotypes evaluated at those field experiments.

We considered two ways of envirotyping: one focused on characterizing "environments" using information for a single season (Steps 1 and 2 in the protocol detailed below), while a second was dedicated to characterizing "locations", using the information across seasons (Steps 1, 2 and 3 in the protocol described below). Consequently, this resulted in two different environmental similarity matrixes (ERM). For "environments", we used the conventional terminology of **W**-matrix, with dimension of environments × envirotyping covariables. For "locations", we introduced the use of a new ERM based on the seasonal-averaged effects (**S-**matrix for locations), with

dimension of locations × seasonal envirotyping covariables. Below, we describe the four steps of those protocols.

Step 1: Raw data collection. The raw environmental data (soil and meteorological variables, which can be found in Supplementary Table S1) was collected using remote sensing databases, such as NASA POWER and in-field evaluations of soil data, as done by Crespo-Herrera et al. (2021), where more details can be found. Additionally, we computed variables with a higher biological meaning, such as growing degree days (GDD), vapor pressure deficit (VPD, kPa/day), evapotranspiration (ET0), atmospheric water balance (P-ETP) and the effect of temperature stress on radiation use efficiency, FRUE = {0,1}, as described in the *EnvRtype* package in R (Costa-Neto et al., 2021c). The cardinal temperatures for GDD and FRUE calculations were assumed the same for pre and post anthesis stages, with Tmin = 0, Topt = 27.7, and Tmax = 40, as given in Wang and Engel (1998). We also subdivided the crop lifetime into "time windows" denoting a generalization of the expected development stage according to the Zadock's scale for wheat (Zadoks et al. 1974), and following heat units for defining the windows of each stage (Rawson and Macpherson, 2000): (Stage 0 - 3) sowing to emergence, then to double-ridge (260° C.day^{-1}); (Stages 3 to 5) double-ridge to terminal spikelet (+150° C.day^{-1}); (Stages 5 to 6) terminal spikelet to heading (+350° C.day^{-1}); (Stages 6 to 7) heading to anthesis (+° C.day^{-1}); (Stages 7 to 9) anthesis to grain-filling and maturity (+500° C.day^{-1}), where the symbol "+" denotes the accumulation of the given heat units for the current stage in relation to the previous stage.

Step 2: Computing envirotyping covariables (EC) for each environment. The computation of EC for each environment was done by a combination of meteorological variable x time window x quantile. Only soil covariables were considered static for each environment (e.g., clay content); that is, with no temporal scale across the crop lifetime, where we assumed EC as the combination of soil variable x quantile. Finally, a total of 108 high-quality ECs for each one of the 130 environments (combinations of year and locations) were found, resulting in 14,040 entries of envirotyping information. The final **W**-matrix (130 × 108) was mean-centered and scaled, with $w_k \sim N(0,1)$.

Step 3: Computing ECs for each location across seasons. For each location, let's assume that for each year of yield testing (each season), a specific **W**-matrix was considered. This is a season-specific matrix of envirotyping data and may not be repeatable across the years, that is, this is a "snapshot" of the possible growing conditions a certain location may face. Each season has a diverse **W**-matrix, in terms of locations and the occurrence of environmental factors; however, the core of **W**-matrices across the years can be used to solve this and provide a "prior expectation" for a given location. It is possible to implement this only when a long-term time-series of envirotypic and phenotypic data is available for a given location. Thus, by using historical environmental data, it is possible to calculate the distribution of each combination of EC by time window across the seasons (assumed here as different years). This was implemented in terms of quantiles (10%, 50%, and 90%, Morais-Junior et al, 2018; Costa-Neto et al., 2021a) of each EC for a given location, plus each static factor. Finally, the data were mean-centered and scaled, with $s_n \sim N(0,1)$. After removing missing values and duplicated columns, the result was put into a matrix (hereinafter called an **S**-matrix) of n = 294 variables for each l = 36 locations, with **S** ($l \times n$).

Learning EPA through Partial Least Squares (PLS)

For the environmental-phenotype associations (EPA), we adopted the non-linear iterative PLS regression as an approach (NIPALS) due to its popularity, simplicity, and effectiveness for diverse data types (e.g., Vargas et al., 1999; Teixeira et al., 2011; Monteverde et al., 2019; Porker et al., 2020; Montesinos-Lopez et al., 2022). The EPA analysis was performed in two ways, hereinafter referred to as PLS 1 (univariate PLS, focused on each genotype at the time) and PLS 2 (multivariate PLS, focused on the environmental relatedness), as detailed jointly with the genomic prediction models in further sections (see Supplementary Figure S2). Additional details of the NIPALS algorithm can be found in Palermo et al. (2009) and Sanchez (2012), as well as in the Appendix section. For each model, we also computed the Variable Importance in

Projection (VIP) score. The VIP score has been used as a variable selection method, and can be computed for each covariable and latent factor as:

$$VIP_j = \sqrt{\frac{m}{\sum_{h=1}^{c} SS_h(\boldsymbol{Y})} \sum_{h=1}^{c} SS_h(\boldsymbol{Y}) l_{hj}^2} \quad \text{(Eq. 1)}$$

where $\boldsymbol{Y}$ is the matrix of responses, m is the number of variables in **X** (matrix of ECs, see Appendix); c is the number of components considered; l_{hj}^2 is the PLS weight of the j-th (j = 1, 2,..., m) variable for the h-th (h = 1, 2,..., c) at each component; and $SS_h(Y)$ is the proportion of **Y** explained by the h-th component. From the VIP, we calculated the relative VIP % by dividing each VIP by the maximum VIP for each genotype (PLS 1) or environment (PLS 2). The PLS approach was implemented in R using the *plsdepot* package (Sanchez, 2012).

Integrating EPA outcomes in Predictive Models

Next, we introduced the use of EPA as an intermediary learning step in genomic prediction. We used six models, in which the first three (M01-M03) were used as benchmark approaches and are currently used for multi-environment GBLUP, while the last three models (M04-M06) involved the inclusion of some EPA outcomes. A summary of the kernel assumptions is given in Table 1. More detail about the relationship matrix is given below according to each model's description. A detailed summary of these approaches is given in the Appendix 2 , Table A1, Genomic prediction models. In this study, we assumed M01 as simplest multi-environment GBLUP model, not considering other possible structures for variance-covariance for residuals or GxE, such as done using factor analytic with envirotyping data (e.g, Rogers et al., 2021; Rogers and Holland (2022). By doing this, we expected to measure the independent effect due to the addition of complex environment relationship structures and EPA results, as described below in the next models (M02-M06). Implications of such approaches involving EPA analysis, as well its adoption over different packages or computational tools, should be discussed in further studies

All models were fitted using the *EnvRtype*:kernel_model() functions, which is based on the kernel-optimized Hierarchical Bayesian approach implemented by the BGGE package (Granato et al., 2018). More details about this approach can be found at Costa-Neto et al. (2021b). We considered 15,000 iterations, in which the first 5,000 were used as burn-in considering a thinning of 5. The seed used was equal to *set.seed*(1112).

M01: Conventional multi-environment GBLUP

This model accounts for the main genetic effects and kernel-based G×E, assuming a block diagonal genomics by environment (Lopez-Cruz et al., 2015; Souza et al., 2017), following

$$\boldsymbol{y} = 1\boldsymbol{\mu} + \boldsymbol{Z_E u_E} + \boldsymbol{Z_G u_G} + \boldsymbol{u_{GE}} + \boldsymbol{\varepsilon} \qquad \text{(Eq.2)}$$

where $\boldsymbol{y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y_n}]'$ are the vectors of observations collected in each of the q environments with p wheat lines; $1\boldsymbol{\mu}$ is the vector of fixed intercept; $\boldsymbol{Z_G}$ and $\boldsymbol{Z_E}$ are the incidence matrix of genetic and environmental effects; $\boldsymbol{u_E}$ is the vector of random environmental effects modeled by $\boldsymbol{u_E} \sim N(0, \boldsymbol{\sigma^2_E K_E})$, in which $\boldsymbol{\sigma^2_E}$ is the variance component related to the macro-environmental variation and $\boldsymbol{K_E}$ is the kernel of macro-environmental effects, where the ERM were assumed as an identity matrix of q environments across the p genotypes $\boldsymbol{K_E} = \boldsymbol{I_q} \otimes \boldsymbol{J_p}$, that is, no prior relation was expected among environments. The genetic effects are modeled by a random main effect ($\boldsymbol{u_G}$), distributed as $\boldsymbol{u_G} \sim N(0, \boldsymbol{K_G \sigma^2_G})$, with $\boldsymbol{K_G} = \boldsymbol{J_q} \otimes \boldsymbol{G}$, and a secondary kernel for environment-specific genetic effects due to G×E, represented by ($\boldsymbol{u_{GE}}$), with $\boldsymbol{u_{GE}} \sim N(0, \boldsymbol{K_{GE} \sigma^2_{GE}})$ and modeled as multiplicative effects due to the GRM and ERM variations:

$$\boldsymbol{K_{GE}} = \boldsymbol{K_E} \odot \boldsymbol{K_G} = (\boldsymbol{I_q} \otimes \boldsymbol{J_p}) \odot (\boldsymbol{J_q} \otimes \boldsymbol{G}) \cong \boldsymbol{Z_E I_q Z_E^T} \odot \boldsymbol{Z_G G Z_G^T}$$

where $\boldsymbol{Z_E I_q Z_E^T} = \boldsymbol{Z_E Z_E^T}$, with $\bigotimes$and $\odot$ were denoted as the products of Kronecker's and Hadamard's, respectively. The matrix $\boldsymbol{J_p}$ and $\boldsymbol{J_q}$ denote the full-entry matrix of 1*s* with dimensions of *p* x *p* and *q* x *q*, respectively. Under balanced situations (every genotype at every environment), $\boldsymbol{K_{GE}}$ could simply be computed as: $\boldsymbol{K_{GE}} = \boldsymbol{I_q} \otimes \boldsymbol{G}$. Finally, the vector of residuals is assumed as a relaxed form for residuals, with $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I_n}\boldsymbol{\sigma_\varepsilon^2})$.

M02: Reaction-norm GBLUP with a linear kernel for W-matrix ($\boldsymbol{\Omega}$)

In this model (Jarquín et al., 2014), we considered the conventional use of envirotyping data (**W**-matrix) and consequently, also the linear variance-covariance matrix for environmental similarity as:

$$\boldsymbol{\Omega} = \frac{\boldsymbol{WW^T}}{\boldsymbol{tr}(\boldsymbol{WW^T})/\boldsymbol{q}} \qquad \text{(Eq. 3)}$$

with $\boldsymbol{\Omega}$ ($q \times q$), where $\boldsymbol{W}$ ($q \times m$), with m environmental covariables (m = 108) and q environments. Consequently, the new kernel for modeling macro-environmental effects is now given by $\boldsymbol{K_E} = \boldsymbol{\Omega} \bigotimes \boldsymbol{J_p}$, and the subsequent G×E structure is:

$$\boldsymbol{K_{GE}} = \boldsymbol{K_E} \odot \boldsymbol{K_G} = \left(\boldsymbol{\Omega} \bigotimes \boldsymbol{J_p}\right) \odot \left(\boldsymbol{J_q} \otimes \boldsymbol{G}\right) \cong \boldsymbol{Z_E \Omega Z_E^T} \odot \boldsymbol{Z_G G Z_G^T}$$

or simply $\boldsymbol{K_{GE}} = \boldsymbol{\Omega} \otimes \boldsymbol{G}$ under balanced conditions as previously described. Thus, now the $\boldsymbol{K_E}$ accounts for the degree of similarity among environments (given by a linear correlation), while the $\boldsymbol{K_{GE}}$ is now the genotype-environment specific effects aimed to mimic reaction-norms.

M03: Reaction-norm GBLUP with a nonlinear Gaussian kernel for W-matrix ($\boldsymbol{\gamma}$)

From models M03 to M06, we adopted the Gaussian Kernel approach (GK) as the nonlinear method for modeling the environmental similarity (given by distances, He et al., 2019 and Costa-Neto et al., 2021b). In model M03, although the nonlinear ERM ($\boldsymbol{\gamma}$) was computed using the same **W**-matrix in M02, we adopted a different notation to differentiate the linear kernel ($\boldsymbol{\Omega}$) and the nonlinear kernel ($\boldsymbol{\gamma}$), which was estimated as:

$$\boldsymbol{\gamma} = exp\left\{\frac{\|w-w'\|^2}{\theta}\right\} \quad \text{(Eq. 4)}$$

with $\boldsymbol{\gamma}$ ($q \times q$), where $\|w - w'\|^2$ is the Euclidean Distance between each element of the **W**-matrix, and θ is a scaling factor assumed as the median value of the Euclidean distance matrix. Note that no bandwidth factor was considered in this equation (assumed as 1). Thus, model M03 is composed of $\boldsymbol{K_G} = \boldsymbol{J_q} \otimes \boldsymbol{G}$, $\boldsymbol{K_E} = \left(\boldsymbol{\gamma} \otimes \boldsymbol{J_p}\right)$ and $\boldsymbol{K_{GE}} = \boldsymbol{K_E} \odot \boldsymbol{K_G} = \left(\boldsymbol{\gamma} \otimes \boldsymbol{J_p}\right) \odot \left(\boldsymbol{J_q} \otimes \boldsymbol{G}\right) \cong \boldsymbol{Z_E}\boldsymbol{\gamma}\boldsymbol{Z_E^T} \odot \boldsymbol{Z_G}\boldsymbol{G}\boldsymbol{Z_G^T}$ or simply $\boldsymbol{K_{GE}} = \boldsymbol{\gamma} \otimes \boldsymbol{G}$ under balanced conditions as previously described.

M04: Reaction-norm GBLUP with environmental weights ($\boldsymbol{\Phi}$) from EPA

Here we introduce the EPA algorithm as an intermediary step in the multi-environment GBLUP. EPA is expected to be capable of linking historical phenotypic and envirotyping data, reducing the issue of lacking future envirotyping data while estimating the actual magnitude and impact of the environmental factors in the phenotypic variation observed in the trialing network. Below, we describe the four steps for incorporating EPA outcomes from the EPA algorithm in the baseline multi-environment GBLUP .

Step 1: Computation of the long-term envirotyping-based S-matrix

Now we consider the **S**-matrix instead of the **W**-matrix. As described in the Envirotyping Protocol section, the **S** is a matrix with dimension of location x seasonal ECs, where the seasonal ECs are a combination of ECs by quantiles (10%, 50% and 90%) across seasons; thus 108 x 3 = 324 seasonal ECs.

Step 2: Computation of the empirical Φ_0 kernel of phenotypic correlations

To address the lack of information for future envirotyping data, this step aims to connect the past "envirotyping-realized" similarity into an "actual" similarity observed among environments, that is, the real phenotypic correlations among trialing sites. To

implement this, the second step of this approach now focuses on computing a core of "empirical index" for a given trait, similarly as proposed by the conventional mean-centered average yield value approaches (Finlay and Wilkinson, 1963; Eberhart and Russell, 1966).

For each location, we computed quantiles of 10%, 50% and 90% for grain yield using all phenotypic records across years and genotypes, thus attempting to capture the distribution for grain yield for a given location. Then, we extracted a pool of "empirical environmental indices" (hereafter called an $\mathbf{S_0}$-matrix), which are based on empirical phenotypic data. Thus, while $\mathbf{S}$ contains the seasonal envirotyping data, the $\mathbf{S_0}$ contains the seasonal descriptors of the environmental quality derived from the past observed phenotypes. This $\mathbf{S_0}$-matrix was then used to create an empirical location-relatedness kernel ($\boldsymbol{\Phi}_0$, with $r \times r$, where r is the number of locations), also following the nonlinear Gaussian kernel approach as:

$$\boldsymbol{\Phi}_0 = exp\left\{\frac{\left\|s_0 - s_0^{'}\right\|^2}{\theta}\right\} \qquad \text{(Eq. 5)}$$

Note that this approach can be used considering an expected cropping time for a given location, which means that a statistic for each trialing site could accommodate a time-scale variation for diverse planting dates. For example, a same location could accommodate several "expected environments" by a combination of planting dates at each location. However, for the purpose of simplicity, and because most plant breeding programs perform analysis in almost the same planting seasons across years, here we considered the seasonal variations of envirotyping data for a static location and planting date.

Step 3: Translating envirotyping data into environmental weights ($\boldsymbol{\varphi}$)

Next, the phenotype-based matrix $\boldsymbol{\Phi}_0$ is then dissected using the envirotyping-based $\mathbf{S}$-matrix with the multivariate PLS enabled by the NIPALS algorithm (PLS 2, see Appendix). This approach can be mathematically represented as:

$$\boldsymbol{\Phi}_0 = \boldsymbol{S\varphi} + \boldsymbol{\Phi}_0^* \qquad \text{(Eq. 6)}$$

where $\boldsymbol{\Phi}_0$ ($r \times r$) is an empirical (phenotype-based) relatedness; $\boldsymbol{S}$ ($r \times n$) is the matrix of envirotyping data by location averaged across seasons; $\boldsymbol{\varphi}$ ($r \times n$) is a matrix of estimated orthogonal PLS coefficients for each n envirotyping data, interpreted here as the *environmental weights* derived from the environmental-phenotype association and specific for each location; and $\boldsymbol{\Phi}_0^*$ ($r \times r$) is the residual location diversity not captured by the envirotyping covariables in the **W**-matrix.

An interesting output of this analysis is the interpretation of the environmental factors that drive the similarity among locations. For this, the locations can be grouped using some clustering analysis (e.g., hierarchical clustering based on the principal components analysis). Finally, the interpretation of the VIPs scores (see Eq. 1) for each environmental covariable is a sign of how these factors have historically shaped the diversity among locations for a given trait in the breeding program.

Step 4: Nonlinear kernel (Φ) *using the environmental weights* (φ)

The environmental weights should be a better descriptor of the environmental similarities because they highlight how the growing conditions have affected phenotypic variation and experimental quality. Consequently, we propose using them as *markers of the environmental relatedness*, which can replace the conventional direct use of envirotyping covariables as done in the previous ERM approaches. Here, by using the nonlinear Gaussian kernel over the $\boldsymbol{\varphi}$ matrix, now it is possible to compute an environmental-similarity matrix for locations as:

$$\boldsymbol{\Phi} = exp\left\{\frac{\|\varphi - \varphi'\|^2}{\theta}\right\} \qquad \text{(Eq.7)}$$

with $\boldsymbol{\Phi}$ ($r \times r$), thus, $\boldsymbol{K_E} = \left(\boldsymbol{\Phi} \otimes \boldsymbol{J_p}\right)$ and, consequently, $\boldsymbol{K_{GE}} = \left(\boldsymbol{\Phi} \otimes \boldsymbol{J_p}\right) \odot \left(\boldsymbol{J_q} \otimes \boldsymbol{G}\right) \cong \boldsymbol{Z_G G Z_G^T} \odot \boldsymbol{Z_E \Phi Z_E^T}$.

M05: Reaction-norm GBLUP with genotype-specific factors (R-matrix) from EPA

Model M05 introduces the use of a second EPA outcome. Here the focus is on studying the reaction-norm patterns from the genotypes in historical breeding trials, to finally recycling it as secondary source of genetic variation (hereinafter called **R**-matrix). This model is an update of model M04, that is, it considers the protocols previously described for computing the $\boldsymbol{\Phi}$ matrix.

Step 1: EPA analysis for computing genotype-specific coefficients ($\boldsymbol{\lambda}$)

A site-regression (SREG, Crossa and Cornelius, 1997) model is fitted for each genotype with both phenotypic and envirotypic data available, which means that a specific regression model is fitted according to the number of observations available for a given genotype. The purpose of this approach is to dissect the genetic-plus G×E effects in terms of genotype-specific coefficients of reaction norm (hereafter named $\boldsymbol{\lambda}$) using the univariate PLS (PLS 1, see the Appendix). This modified SREG model using envirotyping data (Costa-Neto et al., 2020) is given by:

$$\boldsymbol{y} - \boldsymbol{y_q} = \boldsymbol{W\lambda} + \boldsymbol{F}^* \qquad \text{(Eq. 8)}$$

where $\boldsymbol{y} - \boldsymbol{y_q}$ ($q \times 1$) is a vector of phenotypic values for each genotype centered for each test environment; $\boldsymbol{W}$ ($q \times m$) is a matrix of environmental covariables per environment (conventional envirotyping data); $\boldsymbol{\lambda}$ ($m \times 1$) is a vector of genotype-specific coefficients, assumed as the empirical reaction-norms for the given environmental covariable (reaction-norm); and $\boldsymbol{F}^*$ ($q \times 1$) is a vector of the residual SREG analysis not captured by the univariate PLS algorithm. Finally, after running this model for each p genotype, every vector containing the m genotype-specific coefficients was combined into a single matrix as $\boldsymbol{\Lambda} = \left[\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \cdots, \boldsymbol{\lambda_p}\right]^{\boldsymbol{T}}$, with $\boldsymbol{\Lambda}$ ($p \times m$).

Due to the process of scanning reaction-norms into the modified SREG model, it is also possible to use the outcomes of the PLS approach to identify what environmental factors most affect the performance of a certain genotype. The interpretation is done by analyzing the magnitude and direction (positive/negative) of the genotype-specific

reaction-norms, while the variable importance on projection (VIP) could be used to score its importance. Because of this, now the EPA analysis integrates the classical G×E dissection studies (e.g., Crossa et al., 1999; Porker et al., 2019; Costa-Neto et al., 2020) into the current multi-environment genomic-enabled prediction platforms.

Step 2: Estimation of the adaptability through a reaction-norm based kernel (R)

In this step, we recycled the past reaction-norms as markers of the major G×E variation observed in the historical field trials. Then, the $\boldsymbol{\Lambda}$ matrix, used as a genetic covariable, aimed to serve as a secondary source of environmental-specific genetic variation due to the shared reaction-norm patterns of the past evaluated genotypes. Because of the process carried out, this source of genetic variation does not follow an infinitesimal model and presented the same issues that are expected for environmental data (lack of linearity and additivity, Costa-Neto et al., 2021a,b). Next, we used a nonlinear kernel (Gaussian Kernel) to translate the $\boldsymbol{\Lambda}$ matrix into a reaction-norm similarity matrix ($\boldsymbol{R}$) as:

$$\boldsymbol{R} = exp\left\{\frac{\left\|\Lambda-\Lambda'\right\|^2}{\theta}\right\} \qquad \text{(Eq. 9)}$$

with $\boldsymbol{R}$ ($p \times p$) if all genotypes are considered.

Step 3: Estimation of the reaction-norm weighted G-matrix (H)

In practice, plant breeders don't have enough phenotypic information to compute the $\boldsymbol{R}$-matrix, especially if the goal is to predict the performance of a new genotype into a new season. To address this, we propose merging this $\boldsymbol{R}$-matrix with the conventional $\boldsymbol{G}$-matrix in an analogy of what is commonly done for combining pedigree and genomics in the so-called "single-step genomic prediction" (Martini et al., 2018). However, for this scenario, the $\boldsymbol{G}$-matrix ($p \times p$) is known for all genotyped individuals (whether or not it was previously evaluated), while the **R**-matrix ($t \times t$) only considers the empirical reaction-norm estimated for tested genotypes, that is, the training set, with t genotypes, in which $p = t + v$, and v is the number of yet-to-be tested genotypes (testing set).

Because of this, by using the equation from Martini et al. (2018) it is possible to merge $\boldsymbol{G}$ and $\boldsymbol{R}$ into a so-called "**G**-matrix weighted by reaction-norms" (hereinafter called $\boldsymbol{H}$ or **H**-matrix), or simply a Single-Step Reaction-Norm model, computed as:

$$\boldsymbol{H} = \boldsymbol{G} + \begin{bmatrix} \boldsymbol{G}_{12}\boldsymbol{G}_{22}^{-1}(\boldsymbol{R}_{22} - \boldsymbol{G}_{22})\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21} & \boldsymbol{G}_{12}\boldsymbol{G}_{22}^{-1}(\boldsymbol{R}_{22} - \boldsymbol{G}_{22}) \\ (\boldsymbol{R}_{22} - \boldsymbol{G}_{22})\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21} & (\boldsymbol{R}_{22} - \boldsymbol{G}_{22}) \end{bmatrix} \quad \text{(Eq.10)}$$

with

$$\boldsymbol{H}_{22} = \left[\tau\boldsymbol{R}^{-1} + (1-\omega)\boldsymbol{G}_{22}^{-1}\right]^{-1} \quad \text{(Eq. 11)}$$

where $\boldsymbol{H}$ ($p \times p$), τ and ω are scaling factors aimed at reducing inflations of the predictions and ensuring the convergence of the iterative approaches in the mixed models. To simplify the methodology, in this study the scaling factors were assumed as 1; however, this can be fine-tuned in further studies considering the characteristics of each germplasm and trialing network.

Finally, the kernel-based G×E term was replaced from $\boldsymbol{K_{GE}}$ to $\boldsymbol{K_H}$ as $\boldsymbol{K_{GE}} = \boldsymbol{K_H} = \left(\boldsymbol{J_q} \otimes \boldsymbol{H}\right) \cong \boldsymbol{Z_G}\boldsymbol{H}\boldsymbol{Z_G^T}$ with: $\boldsymbol{K_G} = \boldsymbol{J_q} \otimes \boldsymbol{G}$ and $\boldsymbol{K_E} = \left(\boldsymbol{\Phi} \otimes \boldsymbol{J_p}\right)$. We considered that there was no need to replace $\boldsymbol{G}$ with $\boldsymbol{H}$ in those effects, as $\boldsymbol{H}$ is a matrix dedicated to model only the G×E. Thus, this is a solution for incorporating genotype-specific sensibilities into a more parsimonious kernel-based model. The single-step procedures were done using the *AGHmatrix* package (Amadeu et al., 2016).

M06: Reaction-norm GBLUP with single-step G×E kernel from EPA

Model M06 involved the implementation of single-step weighted G×E kernel considering all the information from the matrices $\boldsymbol{G}$, $\boldsymbol{R}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\Phi}$ as previously described. The goal of this approach was to implement an "environmental learning" practice capable of merging the past trends of reaction-norm ($\boldsymbol{R}$, $\boldsymbol{\gamma}$) with expectations of reaction-norm ($\boldsymbol{G}$, $\boldsymbol{\Phi}$). In this model, the main genetic effects and environmental variations were

modeled the same way as in models M04 and M05, that is, as: $\boldsymbol{K_G} = \boldsymbol{J_q} \otimes \boldsymbol{G}$ and $\boldsymbol{K_E} = \left(\boldsymbol{\Phi} \otimes \boldsymbol{J_p}\right)$, respectively. The difference in this model is how the nonlinear G×E interaction terms are modeled. This approach is implemented in three main steps.

Step 1: Estimating past kernel-realized reaction-norms

First, we estimated the past G×E, that is, the G×E observed in the historical field trials used as training sets. We considered a realization of these past or historical G×E by combining the GRM based on the $\boldsymbol{R}$-matrix (t x t) and an ERM based on the $\boldsymbol{W}$-matrix (j x j), in which j is the number of environments in the training set. This step is preceded by the PLS approach, as done for model M05, in which we already estimated the $\boldsymbol{R}$-matrix, as well as considered a nonlinear $\mathbf{W}$-matrix of covariables ($\boldsymbol{\gamma}$), as in model M03. Using the Kronecker product between these two matrices, we achieved the past G×E (for the historical data set) by $\boldsymbol{M} = \boldsymbol{\gamma} \otimes \boldsymbol{R}$, with $\boldsymbol{M}$ ($s \times s$) and s as the size of historical data using ($s = vt$) as the training population set. We used the notation $\boldsymbol{M}$ to simplify further algebraic demonstrations.

Step 2: Estimating future kernel-realized reaction-norms

In an analogy to the method presented in the previous subsection, we created a "full rank G×E", involving all genotypes and locations, as observed in the G×E kernel in model M04. Then we used the static-effect matrix, built up from the environmental weights that each location historically faced ($\boldsymbol{\Phi}$) as a precursor of a next-year growing condition, plus a G×E kernel (here referred to as $\boldsymbol{N}$ to facilitate the mathematical demonstration) was created as: $\boldsymbol{N} = \boldsymbol{\Phi} \otimes \boldsymbol{G}$, with $\boldsymbol{N}$ ($n \times n$) as n is the number of genotypes and environments considered for analysis (training set + testing set). We used the notation $\boldsymbol{N}$ to simplify further algebraic demonstrations.

Step 3: Combining past and future kernel realizations of G×E

Finally, the end-result weighted kernel for G×E effects ($\boldsymbol{K_{GE}}$) is directly computed by combining the $\boldsymbol{M}$ and $\boldsymbol{N}$ kernels as:

If $\boldsymbol{M} = \boldsymbol{\gamma} \otimes \boldsymbol{R}$ and $\boldsymbol{N} = \boldsymbol{\Phi} \otimes \boldsymbol{G}$, then:

$$\boldsymbol{K_{GE}} = \boldsymbol{N} + \begin{bmatrix} \boldsymbol{N}_{12}\boldsymbol{N}_{22}^{-1}(\boldsymbol{M}_{22} - \boldsymbol{N}_{22})\boldsymbol{N}_{22}^{-1}\boldsymbol{N}_{21} & \boldsymbol{N}_{12}\boldsymbol{N}_{22}^{-1}(\boldsymbol{M}_{22} - \boldsymbol{N}_{22}) \\ (\boldsymbol{M}_{22} - \boldsymbol{N}_{22})\boldsymbol{N}_{22}^{-1}\boldsymbol{N}_{21} & (\boldsymbol{M}_{22} - \boldsymbol{N}_{22}) \end{bmatrix} \quad \text{(Eq. 12)}$$

with

$$\boldsymbol{K_{GE}}_{22} = \left[\tau \boldsymbol{M}^{-1} + (1 - \omega)\boldsymbol{N}_{22}^{-1}\right]^{-1} \quad \text{(Eq. 13)}$$

where $\boldsymbol{K_{GE}}$ is an *n* x *n* dimension, and τ = ω =1. In Eq. 12, we have two groups of genotype-environment combinations. Group 1 involves the genotypes already evaluated in past field trials, for which we have all possible information related to the phenotypes, reaction-norm and faced environmental growing conditions. Group 2 is based on genotypes never tested in any past field trial, and environments for which we only know the static-effect priors obtained from long-term historical envirotyping analyses. All analyses were conducted using R statistical software (R Core Team, 2022).

Prediction scenarios over the historical CIMMYT wheat trials

Four scenarios were evaluated in this study. **Scenario 1**: the first scenario considered the predictions of an entire new year using the previous year as a training set. As the data set is highly unbalanced, this also involved the prediction of new genotypes at locations that may or may not be considered in the training set. **Scenario 2** considers the same approach, albeit a different model was adjusted for each one of the three TPE previously identified by Crespo-Herrera et al. (2021) (see the Historical wheat trial data section). **Scenario 3** is the same as Scenario 1 but considers multiple years as the training set. In this scenario, we considered two consecutive years to predict a subsequent year. **Scenario 4** is the same as Scenario 2, but also considers pairs-of-

previous years as a training set for a subsequent year for each TPE individually. The goal of Scenarios 3 and 4 was to verify if it was necessary to use phenotypic data from more than one subsequent year since subsequent years are expected to have the highest genetic correlation, in addition to being more similar in environmental conditions. Thus, the comparison of Scenarios 1-2 and 3-4 may support the interpretation of how phenotypic and envirotypic information from another TPE is useful for making broad predictions for a following year, or whether the use of TPE-specific models is more suitable.

Three statistics were used to measure the quality of the prediction: (1) average predictive ability (*pa*) at "model level", as is conventionally done for testing models in genomic prediction; (2) genotype-specific predictability, based on the Spearman rank correlation between the predicted and observed values for each genotype in a new year, as suggested by each genotype (Costa-Neto et al., 2021a,b); (3) the relation between the predictive ability of the model and the proportion of the wheat population that was considered "predictable", that is, with $pa \geq 0.2$. Here we consider the combination of predictive ability and resolution as a practical measure of the accuracy and usefulness for plant breeders when making decisions for the following years.

Search for Essential Locations in Indian TPEs

The last proposal of our theory relies on the supposition that a good enviromic-based pipeline for testing cultivars and training models for predicting G×Y must also account for the search of "essential locations for phenotyping". This approach was implemented by running a grouping analysis of the locations using the environmental weights ($\boldsymbol{\varphi}$) instead of using raw envirotyping data as was done in the past. For this, we ran a principal component analysis (PCA), followed by a hierarchical clustering using the Euclidean Distance and then the Ward's grouping algorithm. These approaches were implemented using the package FactorMineR (Lê et al., 2008). Then, a number of clusters (Nc) found in this analysis was used as the number of "essential locations".

Next, we searched for the most representative locations at each cluster, here called "pivot locations". Again, the matrix of environmental weights ($\boldsymbol{\varphi}$) was used for this purpose and were dissected using a genetic algorithm based on PEVmean criteria, as implemented in the SPTGA R package (Akdemir and Sánchez, 2019). The genetic algorithm was parametrized using 100 iterations, and five solutions selected as elite parents were used to generate the next set of solutions and mutations of 80% for each solution generated. For each cluster, we analyzed the similarity between each location and its cluster-pivot location in order to find possible "replacements" for those locations. This approach was done by selecting those locations with a similarity equal to or higher than 95% with each cluster's pivot location. The clusters found in this study were also interpreted using the TPE characterization done by Crespo-Herrera et al. (2021).

RESULTS

Variable importance for modeling environmental similarities across years

This section presents the results of the environment-wide characterization focused on understanding the envirotype-phenotype association (EPA) driving similarity among trialing sites (locations). The results are presented (and interpreted) in two steps: (1) in terms of variable importance in projection (VIP) and (2) the use of weights for grouping environments and finding essential locations for METs in India.

Figure 1 shows the general characterization of the 36 locations in India using the seasonal envirotyping covariables and EPA analysis across years. The relative values of VIP % derived from the multivariate PLS (PLS 2, Figure 1A) involving the long-term envirotyping for locations and an empirical environmental index (from phenotypes). The number of latent vectors (LV) was equal to 8, which was selected based on the number of LV capable of explaining at least 90% of the phenotypic-based environmental similarity. For computing the most important variables, we assumed a threshold of 95% (here relative to VIP % equal to or higher than 85%). More detail is given in the supplementary Figure S3. For identifying key development stages, we ran the same analysis for each panel (development stage or soil properties) to find the average VIP for each panel (Figure 1A).

From the environmental weights ($\boldsymbol{\varphi}$ matrix), it was possible to group the locations in five groups (Figure 1B). These groups were found by using the matrix for a principal component analysis (PCA), followed by a hierarchical grouping using Ward's algorithm. Then we were able to pinpoint the ECs with higher importance (VIP > 85%) as a descriptor of the environmental profile that each group has (Figure 1C). However, our results suggest that every environmental factor has a minimum VIP% of ~30% (Figure 1A), which means that not one piece of environmental information is irrelevant, rather some of them are responsible for the main expected similarities between the groups of environments (Figure 1C). As a great number of environmental factors have greater importance (60-85%, green colors, Fig. 1A), all ECs were used to identify the group environments. The signatures of the remaining variables is detailed in Supplementary Figure S4.

With regards to development states, the stage from heading to anthesis (stages 3 to 5) is key for differentiating the quality of the growing environments for wheat in India. At this stage, a great number of variables has the highest impact (VIP >=85%), such as the minimum quantile of evapotranspiration (ETP at q10), median value of air humidity (RH2M q50), minimum quantile of global solar radiation (SRAD at q10) and rainfall precipitation (PRECTOT q10). Other important combinations of stages/variables include the vapor pressure deficit (VPD) at the early vegetative stages (e.g., during double ridge appearance), as the impact of temperature on radiation use efficiency (FRUE) at anthesis (stages 6 to 7) are key factors contributing to increasing the diversity of growing conditions.

Locations that represent the diversity of growing conditions in India

Then, we used the number of clusters (5) as an indication of the effective number of locations that represents the environmental diversity of the locations in CIMMYT's experimental network for wheat in India. Next, we used a genetic algorithm based on PEV mean criteria to analyze the matrix and identify the five locations most likely to be "pivotal-locations" for the phenotyping network for India. In this analysis, we identified

locations ID: Group 1 = 22231, Group 2 = 22257, Group 3 = 22260, Group 4 = 22268 and Group 5 = 22276 (Figure 2). This means that these locations are essential for conducting adequate phenotyping that can capture the diversity of growing conditions, thus lowering the number of field trials.

As occurs in any real-world breeding program, due to some external factors (logistic, political issues, budget), some favorable locations are not available for conducting future field trials. To avoid any theoretical gaps in the methodological analysis, we proposed a score of the locations in relation to the pivot location of each group identified in the clustering analysis (Figure 2). As a threshold, we adopted a similarity equal to or higher than 95%. We observed that the following locations (gray colors) are not necessary for the experimental network: group 1 (22275, 22251, 22205, 22256); group 2 (22214); and group 5 (22284, 22206, 22261). Location 22276 (group 2) seems to be essential, and no other location can replace it. For Groups 4 (pivot 22257) and 5 (pivot 22231), it is easy to replace this location with any other, giving the opportunity to use other locations in this group for screening secondary growing conditions, such as fertilizer or irrigation levels and biotic stresses.

Based on the TPE to which each one of those pivot locations belongs, there is a concordance with the study conducted by Crespo-Herrera et al. (2021). However, variations within each TPE were detected, which increased the resolution to find key locations that represent the variability of all MET in India. For example, the diversity of TPE 1 is represented by two pivot locations, 22260 (group 1) and 22231 (group 5), where it seems that the patterns the ECs ETP_q10 [Stages 3 to 5], FRUE q10 [Stages 6 to 7] and windspeed (WS2M_10 and q10 at stage 7 to 9) were key for splitting TPE 1 into two distinct groups (Figure 1C). The diversity of TPE 2 is represented by the pivot location 22257 (group 4). Another within-TPE division was observed for TPE 3, which is now represented by groups 22276 (group 2) and 22268 (group 3). Within TPE 3, the locations that represent the diversity are differentiated by ECs for a specific development stage and with a major effect on the atmospheric water balance, mostly related to precipitation, radiation and air humidity during Stages 3 to 5.

Hence, the design of field trials in India must follow the pivots (and the environments related to these pivots), as described in Figure 2, where it should be possible to better represent the G×E for the purposes of data analysis and training prediction. It is interesting to note that TPE 1 is the easiest to design, as we have two main pivots and a larger number of highly similar locations as options. In TPE 3, the same situations occur, but one essential location (22276) seems to be indispensable to better capture the diversity of TPE 3. It seems that TPE 2 is the easiest to design due to the higher similarity among the locations that are part of it. Thus, the characterization done by Crespo-Herrera et al. (2021) proved to be very successful for gathering the locations within TPE 2; however, there is a remaining diversity in TPEs 1 and 3 that is now better understood due to the current study. The ideal MET must consider the five pivot locations in each year. However, this is not the situation observed in the Indian MET considered in this study (Figure 2B). The implication of this condition is discussed in the diagnosis of the joint predictive ability trends across years in the next sections.

Plasticity patterns reveled by the univariate PLS 1 algorithm

The reaction-norms were computed for each genotype using the univariate version of the PLS algorithm (PLS 1). To run this algorithm, we first assumed the ideal number of latent vectors (LV) as those capable of explaining at least 90% of the genotype-specific G+GE variability. After a basic study and a data control analysis, we identified 61 ECs suitable for this analysis (Figure 3), which means that ECs with missing values in some environment or with difficult to interpret biological meaning (e.g., reaction-norm for wind speed) were removed in this step.

A major proportion of the genotypes were well fitted ($R^2 \geq 0.90$) with the number of latent vectors ranging from 3 to 10 (>90% of the germplasm), and a few were outliers, such as LV from 15 to 30 (2% of the germplasm). Then we collected the variable importance in projection (VIP) statistic and converted it into relative VIP % by dividing the VIPs for each genotype by its maximum VIP value (that is, the most important

environmental variable for a given genotype). The distribution of the % VIPs considering all genotypes is given in Figure 3.

Soil properties such as CEC (soil cation exchange capacity), WWP (soil water wilting point) and CLYPPT (clay content) seem to be essential for modeling the adaptability of the genotypes in all years considered (Figure 3A). For the meteorological variables (divided by development stages), we observed a wide range of VIPs (sometimes ranging from 0% to 100%), which indicates huge variability in the germplasm in terms of reaction-norm. Taking the average VIP for each variable and looking at those higher than 80% (that is, of higher importance), this variability is also reflected in differences among years and crop development stages. Some variables exhibit a consistent higher VIP% across years for most of the germplasm (Figure 3B), such as the photoperiod (N) in the early vegetative stages, such as 0 to 3 and 3 to 5. Factors related to the air temperature such as maximum temperature (T2M_MAX) in stages 5 to 6, 6 to 7 and 7 to 9, in degree of importance, respectively, seem to play an important role in the plasticity of the wheat germplasm in India. The temperature range (T2M_RANGE) is also important for stages 7 to 9. Finally, a related covariable is the effect of temperature on radiation use efficiency (FRUE), which is highly important for reproductive stages 5 to 6 and 6 to 7 but has a minor effect during the vegetative stages. Another important EC is the variables of vapor pressure deficit (VPD) during reproductive stages (5 to 6), during terminal spikelet formation (3 to 5) and grain filling to maturity (7 to 8).

The diversity of genotype-specific coefficients ($\boldsymbol{\Lambda}$ matrix, illustrated in Figure 4A) highlights the problem of selecting environmental covariables that explain G×E. There is huge variability in signal (positive/negative reaction-norms) and magnitude across different years and germplasm (as each year has a different set of genotypes). A few genotypes demonstrate “outliers” in their reaction-norm responses for some specific factors (extreme blue or red colors). On the other hand, most of the reaction-norms range from -0.5 to 0.5 (orange to green), but their diversity across diverse development stages and years suggests the difficulty of selecting a few covariables to really explain the G×E across years. Because of this, we considered the matrix of all genotype-specific

coefficients ($\boldsymbol{\Lambda}$) to create the **R**-matrix (Figure 4B). Thus, we observed that this **R**-matrix could be a potential secondary source of genetic relationship, which in this case accounts for the shared reaction norm patterns observed in past field trials. Consequently, this empirical source of genetic variation could help to model G×E variation when combined to the conventional **G-**matrix

This empirical relationship can also be useful to understand the lack of relatedness between the set of genotypes (wheat lines) in a given year and the whole breeding germplasm– in this case, the Indian wheat germplasm from CIMMYT (Figure 4C). The set of genotypes evaluated in 2015, for instance, is highly related (blue-purple colors) to the core germplasm, which can suggest, for example, that this germplasm is a good training set for predictions across years. On the other hand, the set of genotypes in 2011 is poorly related to the core germplasm (black, red colors), which is a sign of the lack of genetic relatedness and the differential genetic progress observed in the Indian TPEs that have shaped (and diverged) the past wheat lines from the more recently developed germplasm. For prediction purposes, we considered the **R**-matrix for pairs of years (or three years, according to the prediction scenario) for running our genomic prediction analysis.

Comparison between linear and nonlinear environmental relationships for G×Y predictions

The average values of predictive ability, measured by the Spearman correlation, are illustrated for each model and scenario in Figures 5 and 6. As expected, the baseline model without enviromics (M01) was the worst model for any scenario. The average predictive ability for this model (horizontal red lines) was very close to 0 (horizontal black line); below 0 for the analysis considering all TPEs together (Scenarios 1 and 3); and above zero for those models fitted for each TPE independently (Scenarios 2 and 4). This result suggests that the definition of TPEs, followed by fitting independent models for each TPE, is a good strategy when no enviromic data are considered. TPE 2 was the most predictable without enviromic data, but even in the best scenario, a lower

predictive ability (r ≅ 0.20) was observed, and in most cases, r ≅ 0.00 or negative. In Figure 6B, we present a fair comparison between how frequently the alternative models outperform M01.

In this study, our goal was to predict a new year, but before going any further, it is essential to discuss models M02 (conventional linear reaction-norm, Jarquín et al., 2014) and M03 (nonlinear reaction-norm, Costa-Neto et al., 2021b). On average, these models outperformed the M01 models in every scenario (Figure 6A), with gains up to +218% (Scenario 1, M03) and +287% (Scenario 3, M02). For any TPE and scenario, the M02 and M03 seem to always outperform the M01 in most years, except for Scenario 1 – TPE 3 (M02) and Scenario 3 – TPE 3 where the use of M01 fitted using TPE-specific environments seems to lead to better results (Figure 6B). Both models are based on the same **W**-matrix of covariables (the same used to compute the genotype-specific coefficients in the PLS 1 step). However, the difference is that one model accounted for a linear approach for the ERM (correlation based) and the second model considered the nonlinear approach (Gaussian process for the similarities). Except for Scenario 3 (Figure 5), in all other scenarios the M03 outperformed M02, both in terms of higher mean (dots) and median (horizontal colorful lines).

Figure 6 indicates that in Scenario 1 (using two years of data in a single model for all TPEs), the gains compared to M01 observed for M03 ranged from +90% (TPE 2) to +480% (TPE 1), while for model M02, they ranged from +20% (TPE 2) to +200% (TPE 3). A similar trend was observed for Scenario 2, in which in TPE 2, the M02 was outperformed by M01 (- 15% reduction), while for the same TPE, the M03 outperformed M01 by +37%. For Scenarios 3 and 4 (using more than 2 years), the same trend was observed for those three models, except for TPE 3 in Scenario 3, where M03 was the worst model. Thus, in general, the use of a nonlinear kernel for ERM (M03) outperformed the conventional way of using environmental information in genomic prediction (M02).

Accuracy gains when adding EPA as an intermediary step in GBLUP

Looking at only the benchmark models (M01, M02 and M03), the results suggest that TPE 1 was the TPE that most benefitted by the inclusion of enviromics, irrespective of the kernel method or envirotyping protocol used. However, when we analyzed the proposed enviromic-aided strategies (M04, M05 and M06), the gains were outstanding for most scenarios and TPEs. For example, in Scenario 1 (Figure 5), while M03 achieved a gain equal to +218% (r =0.18) over M01 (r=-0.05), the replacement of the conventional **W**-matrix by the environmental weights from the S-matrix (M04) achieved a gain of up to +524% (r=0.58). Higher gains from M04 were also observed in Scenario 2 (+82%, r = 0.41), Scenario 3 (+281%, r=0.25) and Scenario 4 (+128%, r = 0.33). Considering the predictions for each TPE (Figure 6A), model M04 was very effective for predicting the TPE 3 in most scenarios, with gains due to M01 up to +540% (Scenario 1, r =0.54), +386% (Scenario 2, r=0.56) and +18% (Scenario 4, r = 0.18); however, there was a reduction of -6% for Scenario 3 (r=0.16). A negative result was also observed for TPE 2 in Scenario 2 (+12%, lower than other models such as M03) and Scenario 3 (-11%). Thus, model M04 aggregates benefits and increases accuracy over M03 in most cases; nevertheless, for some scenarios/TPE, we observed instability in the predictions due to a lack of correlation relative to the "genetic causes of G×E", that is, the environmental weights alone generally help to increase the prediction, but they are not enough to ensure stability in the predictions across all scenarios and TPEs.

The inclusion of the reaction-norm based matrix (**R**-matrix, M05) helped solve these issues in some scenarios and TPEs (Figures 5-6), but it does not always add accuracy in relation to the standard use of the environmental weights. For example, the comparison of the observed global predictive ability (Figure 5) for models M04 and M05 showed almost the same gains for the scenarios that included two-year data (Scenarios 1 and 2), but higher gains for scenarios that included more than one year (Scenarios 3 and 4). In this last scenario, the gains over the M01 ranged from +328% (r = 0.28) and +141% (r=0.27), respectively. The predictions for each TPE (Figure 6A) were also interesting, achieving the best performance in terms of predictive ability in Scenario 2 for TPE 1 (r= 0.38, +145% over M01) and TPE 3 (r=54, +371% over M01). Another

interesting comparison can be observed in TPEs 2 and 3 in Scenario 3, with gains ranging from +10% to +88% compared to -6% to 25% observed for M04. For TPE 1, however, it seems that M04 and M05 were always better choices for modeling in any scenario in comparison with M01 (Figure 6B), while M06 seemed to present almost the same performance despite being a more complex modeling approach.

Finally, model M06 is the combination of the reaction-norm matrix and environmental weights. In most cases, the models seemed to achieve similar performance as their predecessors (M04 and M05), but when looking at the variability of the predictive abilities (size of the boxplots, that is, 25% to 75% quantiles), these models suggested more consistent predictions across all scenarios (Figure 6B). Model M06 achieved its highest performance in Scenario 1, with gains ranging from +130% (TPE 2) to +460% (TPE 1) and +540% (TPE 3). For TPE 3 in Scenario 3 (Fig. 6A), this model was the best model (r=0.40, +44% over M01), with the potential to achieve predictive ability values close to r = 0.65.

Resolution of the G×E prediction at the genotype level

The resolution of the G×E prediction was given by the genotype-specific Spearman correlation between observed and predicted yield values for a future year. Figure 7 presents these results considering all years (from 2012 to 2018; 2011 is not considered because it was not predicted). For all models, the differences in the prediction scenarios affected the resolution of the G×E prediction. The resolution of M01 (no enviromics) ranged from 3% (Scenario 4) to 35% (Scenario 2). For predicting a new year based on the previous one, this resolution was equal to 11%. This value means that the genomic prediction model is capable of adequately predicting the G×E variations for 11% of the germplasm (the remaining 89% is unpredictable). The inclusion of enviromics was key for increasing the resolution. For Scenario 1, the use of a nonlinear kernel for **W**-matrix (M03) pushed the resolution from 11% (M01) and 25% (M02) to almost 50% of the germplasm. This same trend was observed for Scenario 4. However, when a TPE-specific model was fitted (Scenario 2), those three benchmark models had almost the same proportion of the germplasm predicted (~45%). It should

be noted that the resolution of the M03 was still the highest one (proportion of blue and purple colors). In Scenario 3, M02 exhibited the highest resolution, perhaps due to the fact that the linear kernel adequately captured the phenotype-envirotype association driving the G×E pattern population set.

The highest resolution was achieved by model M06 in prediction Scenario 1. This means that in this scenario, we observed the highest predictive ability gain (+559%, r = 0.56, with r_{max}=0.69) and highest resolution (~95% of the germplasm). This resolution was drastically reduced in the other scenarios, especially 2 and 3 (inclusion of more than one year), which suggests that the increase of phenotypic data does not act as a source for increasing the resolution of G×E. The separation of the models for each TPE using two-year data (Scenario 2) seems to be a satisfactory strategy for most models, but here models M04 and M05 outperformed model M06. This suggests that a good adaptability pedigree within each TPE and a good measure of the within-TPE weights must be useful for training the models.

DISCUSSION

In this study, we present results addressing the following topics: (1) how the enviromics linking EPA studies could leverage the current TPE characterization priors used to fit TPE specific GBLUP models, while borrowing information within and across TPEs; (2) how nonlinear kernel methods seem to outperform linear methods in terms of accuracy; (3) how environment-wide characterizations revealed the differential effect of the seasonal variations in magnitude, direction and impact of the environmental factors in the trialing networks across different years, which affects any reaction-norm estimation; (4) how EPA studies could help the development of reinforcement algorithms for "learning" how the macro-edaphoclimatic variations affect the environmental quality for a given trait across historical data sets and ; (5) how the current cultivar testing pipelines can accommodate these EPA approaches into a breeder-friendly manner, without adding costs to the current budgets, and as such, add

value in the decision-making processes of selection and design of more representative and cheaper trialing networks.

EPA enables environmental characterization and plasticity modeling

In this study we introduced a pipeline for characterizing the similarity among locations using EPA, envirotyping data and historical yield data. In the summary, our environment-wide characterization suggests that: (1) soil properties and elevation are less important for describing the variation across years (average VIP % = 50%, with maximum average higher importance equal to 60%), which is reasonable when considering that the main difference observed in environmental conditions is due to seasonal variations in meteorological variables; (2) the temporal variation of meteorological factors is key for understanding the similarity among environments, in which some variables are more important than others due to the crop-specific developmental stages; and (3) due to the use of diverse statistics for capturing the distribution of these factors, we observed that median values are not always the best statistics for describing the distribution and the weight of some environmental factors. Hence, we posit that the process of selecting covariables is not the ideal situation for modeling the effects of seasonality across years since the envirotype-phenotype patterns are not static across years due to seasonal variations in climatic factors and genetic variations between the different genotypes evaluated in each cycle. Thus, although the EPA approach was useful in revealing some "environmental signatures" that describe the major factors affecting the diversity of growing conditions within TPEs, the use of environmental-wide factors seems to be a more feasible and conservative way of modeling the similarities among locations in the future.

The EPA analysis of genotypes (reaction-norm modeling) also suggests that the selection of covariables as part of prediction models could be a contradictory strategy. While the selection of covariables could increase the accuracy of previous years, which can be useful for exploratory purposes, it may not be feasible for supporting prediction models in real-world breeding programs, especially when the goal of these models is the prediction of future years (new G×Y) where the plasticity of the yet-to-be-seen genotypes is unknown. This is a consequence of the huge time-scale differences in VIP%

across years (seasons), germplasm (different genotypes evaluated in different seasons) and across development stages. Because of this, for further analysis, we advocate using all envirotyping covariables to create the environmental relationship matrices – and consequently, the kernel-based reaction norms in genomic prediction. For modeling reaction-norms, the PLS approach seems to oversimplify the diversity of reaction-norms; consequently, we suggest that in future studies, nonlinear approaches such as GAM (Heinemann et al., 2022) be used as an alternative approach. Despite this, the use of PLS-based genotype-specific coefficients of reaction-norm were successful in recovering the main trends of plasticity and leverage of the genomic prediction for future years in the historical data sets. This will be discussed in further sections and has been observed in other applications described by Montesinos-Lopez et al. (2022).

Linear kernel methods do not capture the nonlinearity among field growing conditions

The statistical modeling approaches for the analyses of plant breeding in multi-environment trials continue evolving as more data with more complex structure are being collected in plant breeding programs (Crossa et al., 2021, Teixeira et al., 2011). For example, diverse sets of methods are available for dealing with multi-dimensional data, such as the nonlinear approaches that are now being applied for modeling environmental relatedness using large-scale envirotyping data (Washburn et al., 2021; Rogers et al., 2021; Westhues et al., 2021; Costa-Neto et al., 2021a,b). Here were compared the conventional multi-environment GBLUP (M01, no enviromics) and two reaction-norm GBLUPs, the first using envirotyping data on a conventional linear kernel (M02, linear W-matrix, Jarquin et al., 2014) and the second using these data on a nonlinear Gaussian kernel (M03, nonlinear W-matrix, Costa-Neto et al., 2020b). M02 and M03 assumed that *we know the future envirotyping data*, but model M03 (nonlinear kernel) outperformed model M02 (linear kernel) in every scenario tested. This result agrees with what was achieved in wheat in Australia (He et al., 2019) and tropical maize in Brazil (Costa-Neto et al., 2021b), in which both the use of nonlinear kernels (Gaussian

kernel and/or Deep Kernel) were superior to the linear kernel for modeling environmental relationships from ECs.

From the biological point of view, it seems reasonable to affirm that environmental covariables are non-additive between each other (Costa-Neto et al., 2021c). Environmental covariables also have higher collinearity and a lack of orthogonality (Heinemann et al., 2022). Because of this, some studies have added the step of "variable selection" (e.g., Millet et al., 2019; Westhues et al., 2021; Mu et al., 2020), which could help in to overcome this issue but could cost a loss of information when trying to predict a yet-to-be-seen GxE, as will be discussed in further sections of this paper. Here we approached these issues by building non-parametric, nonlinear environmental kernels, that take into account (leveraged) some phenotypic data as prior information to adjust bandwidth factors (Costa-Neto et al., 2021b) and that are able to learn hidden nonlinearities underlying the variations between the observed macro-environmental influence (from envirotyping) and the actual phenotypic variation (the resulted GxE). This "environmental learning" approach is a supervised method that goes in the same direction of other approaches used for this purpose, such as convolutional neural networks (Washburn et al., 2021) and gradient boosting machine (Westhues et al., 2021), but with the benefit of dimensionality reduction as the historical GxE and reaction-norm patterns are summarized in two simple matrices ($\boldsymbol{\gamma}$ and $\boldsymbol{\Phi}$).

We also observed that nonlinear kernels are better at handling outliers in the environmental covariables and are also more conservative in building similarity relations (not covariances). For instance, the Gaussian Kernel for environmental similarity led to a diagonal matrix (diag(GK) = 1), with off-diagonal effects equal to the degree of similarity (from 0 to 1) among environments. Considering the lack of similarity (identity matrix) is the same as assuming that every envirotyping data does not add value in weighting those similarities. In this sense, a good envirotyping protocol must be designed and conducted to avoid misspecification of those similarities, while differentiating the effect of different timescales (e.g., time-windows during crop lifetime, phenology) according to the crop phenology. For example, there is no sense in using monthly temperature values to model the growing conditions faced by an annual crop

such as maize, due to the fact that during 30-day intervals, the crop goes through a wide number of development stages, and for each one of them, the crop will not face the same environment due to the differential needs and sensibility to inputs and stress factors, as evidenced by Rogers and Holland (2022). On the other hand, the use of high-resolution data, such as an hour-scale, may not aggregate value and moreover, may even be a source of difficulty that will limit the accuracy of the reaction-norm models (Jarquin et al., 2020).

However, our results show that even when we maintain the same envirotyping protocol, the kernel structure choice is key for the successful prediction of G×E. In fact, the lower accuracy results obtained from linear envirotyping-based ERM in this study, and in other publications (e.g., Millet et al., 2019; Rogers et al., 2021; Rogers and Holland, 2022) can be attributed to the fact that they fail to reproduce the true quality of the environment by overestimating or underestimating its quality. For this reason, data analysts should dedicate efforts towards developing a diverse set of "environmental markers" that can be globally used for a certain species or that are germplasm specific. These markers must be designed and tested to account for actual plant ecophysiology interactions, which can be computed by crop growth models or using frequencies of occurrence of environmental typologies (Costa-Neto et al., 2021c), which in both ways can be leveraged by using environmental-phenotype association when historical databases are available for it, especially when we want to predict future growing conditions based on past priors.

TPE-specific models do not ensure higher accuracy in G×Y predictions

TPE characterization is a key for reconciling breeding goals and expectations with the actual structure available to support the screening for genotypes as candidate cultivars. This also provides information to check the representativeness of the current experimental network – and, sometimes, highlights key aspects to optimize it. One of these key aspects is implied directly in the design of strategies for crossing and selecting more adapted genotypes for key growing conditions, which leverages the genetic gains and ensures the efficiency of the breeding program, allowing breeders to focus on other

breeding goals – that is, secondary traits, such as nutritional quality. Thus, it is a sustainable way to measure the quality of the germplasm for specific growing conditions, and consequently, to better deal with future extreme events, such as climate-change panels around the world. Thus, the envirotyping protocols are now integrated into breeding and post-breeding pipelines as a "thermometer of the germplasm adaptation", by showing the potential phenotypic landscapes (Messina et al., 2018; Costa-Neto et al., 2021a) and for providing evidence about how the environment could affect past genetic gains (Heinemann et al., 2019; Cooper et al., 2021), as well as the actions that must be taken to achieve the breeding goals in upcoming years (Elli et al., 2020). In this context, envirotyping data, historical data or simulated data from crop growth models have been used to clearly define the limits of each TPE in breeding programs and, as a result, this prior information can be used to control (or exploit) the G×E patterns within homogeneous growing conditions (Windhausen et al., 2012; Crespo-Herrera et al., 2021).

The G×E interaction is a key property of the experimental network, and its magnitude and nature depend on the diversity of the experimental network (Costa-Neto et al., 2021c;). This was successful in reducing the crossover G×E interaction when targeting cultivars and borrowing information across countries for training genomic prediction (Crespo-Herrera et al., 2021). However, for the prediction of new G×Y, our results suggest that models without the division of TPEs favor the EPA study, as the amount of screened growing conditions for each genotype ensures better estimation of the reaction-norms. In addition, the increased number of observed environments also provides a better estimation of the environmental weights, which favors the enviromic-aided models. The use of a particular model for each TPE favored the classic conventional multi-environment GBLUP (M01) (as expected, models M02 and M04 in most cases). However, the accuracy gains achieved by the models accounting for enviromic-aided pedigrees were much higher, for it was observed that accuracy gains were nearly 1,000%; there were also higher resolution gains.

Since our data set is highly unbalanced across years, we failed in testing to see whether replication of certain genotypes across years, for a specific TPE, can be useful to optimize cross-TPE predictions over different years. Thus, it is possible that using TPE-specific phenotypes, while accounting for replications of genotypes across years, could be a good balance between the optimization of independent phenotyping efforts within each TPE and an accurate Envirotype-to-Phenotype Association (EPA) analysis, which is the basis of the enviromics-aided pedigrees. In addition, as observed in other studies in maize (Costa-Neto et al., 2021b), it is possible that not every phenotype data point is useful to aggregate information in genomic prediction – in fact, sometimes this is one source of difficulty that causes predictive ability losses. Then, the use of multi-year data, under unbalanced conditions, might not contribute at all.

For some years, the lack of similarity present in the $\mathbf{R}$ matrix also partially reflects the lack of accuracy. In the years 2012 and 2016, even when considering more phenotyping data (across years and TPEs), the observed predictive ability was low for most of the models.

Association studies between envirotyping data and phenotypic variation helps for dealing with the uncertainty of future G×Y

The use of historical yield data accomplished with long-term envirotyping is the basis for understanding how the environmental factors (meteorological, soil, biotic factors), across the plants' lifetime, affects the end-result phenotype of interest. As such, the field of the Envirotype-Phenotype Association (EPA) can approach this uncertainty, by reverting to past studies on factorial regression (e.g., Hardwick and Wood, 1972; Denis, 1988; Costa-Neto et al., 2021a) and partial least squares (Vargas et al., 1999; Porker et al., 2020) to provide a more realistic descriptor of the environmental impacts on the experimental network. As each experimental network is composed of genotypes and environments, two types of analysis were proposed in this study: (1) the search for descriptor adaptability by estimating the empirical reaction-norms of the genotypes

tested in the past; (2) measuring the weights of each environmental factor in driving the relatedness and diversity among field trials (or locations across years).

The first approach focused on identifying a possible "population structure" for reaction norms, that is, to summarize the impact of the historical plasticity patterns shared by the genotypes, which can be used as an additional genetic relationship matrix (GRM) for multi-environment predictions. The second is a new way to implement a weighted environmental relationship matrix (ERM), which is capable of accommodating past and future environmental covariables for predicting the expected "phenotypic similarity" for a given trait. Then, for the latter, different weights of each environmental factor should be achieved for different traits of the same genotype – which means that we can directly relate the quality of the growing conditions with the trait expression, translating the theoretical putative environmental influence on actual impact on the trait performance of the genotypes. Finally, a third "G×E pedigree" is born by combining the previous approaches, which result in a G×E kernel considering reaction-norms (genotype-specific coefficients), genetic relationships (from pedigrees or molecular markers), field-trial (or location)-specific weights for each environmental factor and observed covariables in past field trials. This last option is a more parsimonious manner of integrating different data sources capable of explaining the G×E variation for a given trait, at a given experimental network or breeding region. For this last application, it seems that a successful approach that follows a similar philosophy is the use of self-organizing maps (SOM) to identify genotype-specific responses for key environmental types combining historical yield and envirotyping data (Bustos-Korts et al., 2022).

Here, it was demonstrated that the conventional multi-environment GBLUP (M01, no enviromics) has poor resolution in explaining genotype-specific G×E variations for future years. As these models only accounts for genotypic covariates (G-matrix), and not any other environmental information, this result is not a surprise. Conversely, our results suggest that the approaches (M04, M05 and M06) are cost-effective and biologically accurate ways to correct this issue, with the benefit of better exploring the cross-TPE and cross-season phenotypic data for training the GP models. Furthermore, the models M04-M06 aims to capture the historical background of seasonality variations

by modeling the plasticity (reaction-norm matrix) of relatives while measures the actual weights of the envirotyping information in the phenotypic variation. Thus, this is an evolution of the conventional way to use raw environmental data to shape the similarity among locations.

A key point of our study was the development of the EPA-weighted location similarity matrix ($\boldsymbol{\Phi}$), used by itself (M04) or combined with other enviromic-aided approaches. The ERM derived from this matrix is modeled by nonlinear GK, bringing all the benefits of this approach, as already discussed in the last section. Perhaps an interesting point of this matrix is that it can always be updated (every year) and can also include new testing sites. For example, if we don't have any phenotypic data for a new location (that is, we don't have matrix $\boldsymbol{\Phi}_0$ considering this new location, Eq. 11), then we can still use the long-term envirotyping data for a new *j* location, assuming the weights for the *m* environmental covariables of this location as $\boldsymbol{\varphi_j} = 1_m^T$, while the observed *r* location carries its specific weights computed by Eq. 11 using the data of the past experimental network; then $\boldsymbol{\varphi} = \left[\boldsymbol{\varphi_r}, \boldsymbol{\varphi_j}\right]$. This is a feasible and practical approach capable of using the outputs of the EPA analysis to unlock the avenue for predicting the expected phenotypic correlation for a new testing location. Finally, here we used a Hierarchical Bayesian approach, and the results demonstrated that the use of the kernels ($\boldsymbol{K_G}$, $\boldsymbol{K_E}$ and $\boldsymbol{K_{GE}}$) can be easily adapted for other computational platforms, such as convolutional neural networks (CNN) or assembly methods such as gradient boosting machine. Thus, a generic representation of a multi-environment genomic prediction could be given by $\boldsymbol{y} = f(\boldsymbol{K_G}, \boldsymbol{K_E}, \boldsymbol{K_{GE}}) + \boldsymbol{\varepsilon}$,in other modeling forms, just considering markers (SNP) and supervised EPA-based outcomes, such as: $\boldsymbol{y} = f(SNP, \boldsymbol{\varphi}, \boldsymbol{\Lambda}) + \boldsymbol{\varepsilon}$.

Environmental-Phenotype Association act as a reinforcement algorithm in recycling information from historical trialing networks

Finally, we can suggest a pipeline that combines our results with other approaches. For example, from the same multi-environment trial data we can follow two pathways. The first pathway is the "genotype analysis" based on the process of PLS 1 (univariate PLS for modeling the reaction-norms). This process has led to different purposes, but mostly to the estimation of the genotype-specific coefficients that can be part of cultivar targeting (e.g., Porker et al., 2020; Costa-Neto et al., 2021a), which could also be used as a strategy for reducing the dimensionality in genome-wide studies across multiple environments. For the latter, the use of the "reaction-norm coefficients" as a "trait" for genome-wide association studies (Li et al., 2021; Mu et al., 2022) cold help to understand the interplays between genomics and phenotypic plasticity. Here we introduce the reaction-norm matrix (**R** matrix) as containing "markers" of the plasticity patterns observed in past evaluated genotypes. Thus, instead of using a large data base of phenotypes from genotype x environment combinations, we can add the same information by recycling the past reaction-norms into an intuitive and synthetic matrix of "plasticity markers" (phenotype ~environment variation associations).

The second pathway is the "environmental analysis", which is based on the PLS 2 algorithm (multivariate PLS for modeling environmental weights). This approach is useful in creating more accurate ERM, which could be used to improve the accuracy of GP but also for identifying key environments and optimizing multi-environment trial efforts, which consequently could also extend the range of applications of the predictive analytic pipelines due to the possibility of better predicting the quality of some future environment. This last application can be used, for example, not only in the first multi-environment trial evaluations in breeding pipelines, but also in post-breeding stages aimed to support the decision about which locations are more likely to be successful as seed production fields, which is a key logistic aspect for multiplying seeds of the candidate cultivars.

Nonetheless, it is important to highlight two key aspects. First, every reaction-norm is just a sample of the possible reaction-norms that a certain genotype could

experience due to its potential genetic-defined plasticity (Costa-Neto et al., 2021c), that is, a "snapshot" of its potential phenotypic plasticity as this reaction-norm coefficient is dependent on the diversity level of the MET and how it can represent the actual TPE. Because of this, some GP/GWAS models accounting for reaction-norms could have noise related to the limiting environmental diversity that genotype experienced in the past. Thus, in the step of reaction-norm computation, the use of large historical data sets is key to reduce this possible bias. Secondly, but not less important, to work as reinforcement learning, this approach should be conducted after every year of trialing, where each new EPA is inserted into the data base and reaction-norms/environmental weights are fine-tuned.

CONCLUSIONS

The effect of climate seasonality and the differences in the genetic constitution of the populations evaluated in different years makes difficult the prediction of G×E without any additional information. In this study we proposed a novel framework capable to recycle the historical data and use it to leverage prediction accuracy of the conventional GBLUP methods (with and without reaction-norms) for a new year. To be able to capture the essential patterns of the G×E signal, the key was to use the partial least square algorithm as a first step, that is, as an "environmental learning way" for modeling reaction-norms from the past evaluated genotypes, while computes how the environmental factors impacts on the phenotypic correlation across multiple locations. The first provided to us an empirical "reaction-norm matrix" (**R**-matrix), who was used as secondary source of genetic variation, while the latter provided the "environmental weight" matrix aimed to replace the direct use of envirotyping data for unsee years. We call this approach as "Environment-Phenotype Associations (EPA)" since it captures the essentials of G×E information from historical breeding data, to be able to use this refined information in the prediction models.

Then, in a second step, both outcomes were parsimoniously integrated in the multi-environment GBLUP models with which the predictions of unseen year or

environments were finally performed. This led to a biological enrichment of the genetic and environmental relationship matrices (GRM, ERM), and consequently the G×E structures, without increasing model complexity and dimension. The G×E signal was better accounted for, while the noise contained in both environmental and marker data was at least partially removed. New models computing environmental weights and envirotyping data to model ERM, while considering a mix of G-matrix and R-matrix for modeling GRM, seems to increase the ability to explain the past while increases accuracy for the future. Using CIMMYT's multi-year wheat data, our approach outperformed the conventional methods considering only genomic information under a multi-environment GBLUP context, as well its expansion using envirotyping data (reaction-norm GBLUP). During this process, as observed by other authors, we observed that the use of nonlinear kernels for modeling environmental relationship matrices outperform the state-of-art considering linear covariances. Thus, all EPA outcomes were also modeled as nonlinear kernels. Additionally, we observed that because of the seasonality, any strategy for variable selection using envirotyping data may lead to a loss of information and a prediction bias. Conversely, our framework takes advantage of all envirotyping data possible, while captures linear and nonlinear patterns since it was implemented under a mix of linear and nonlinear kernels. This also resulted in the possibility of identifying which locations are the most important for designing future METs. Finally, we encourage more empirical evaluations using our proposed framework to provide more empirical information about its advantages and disadvantages.

Data availability

Data and codes for EPA analysis and for creating the matrices and genomic predictions are available at https://github.com/gcostaneto/EPA-PLS. Additional R codes not necessary to perform the analysis are available upon request from the first author.

Funding

We are thankful for the financial support provided by the Bill & Melinda Gates Foundation [INV-003439, BMGF/FCDO, Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods (AG2MW)], the USAID projects [USAID Amend. No. 9 MTO 069033, USAID-

CIMMYT Wheat/AGGMW, AGG-Maize Supplementary Project, AGG (Stress Tolerant Maize for Africa], and the CIMMYT CRP (maize and wheat). We are also thankful for the financial support provided by the Foundation for Research Levy on Agricultural Products (FFL) and the Agricultural Agreement Research Fund (JA) through the Research Council of Norway for grants 301835 (Sustainable Management of Rust Diseases in Wheat) and 320090 (Phenotyping for Healthier and more Productive Wheat Crops). Finally, we highly appreciate the time invested by the reviewer and the AE for reading, correcting, and improving the readability of the article.

References

Aastveit, A. H., & Martens, H. (1986). ANOVA interactions interpreted by partial least squares regression. *Biometrics*, 829-844.

Akdemir, D., & Isidro-Sánchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. *Scientific reports*, *9*(1), 1446. https://doi.org/10.1038/s41598-018-38081-6

Amadeu, RR, et al. (2016). AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. The Plant Genome 9, 4.

Souza, B. M., Cuevas, J., Couto, E. G., de, O., Pérez-Rodríguez, P., Jarquín, D., et al. (2017). Genomic-enabled prediction in maize using kernel models with genotype × environment interaction. G3 7, 1995–2014. doi: 10.1534/g3.117.042341

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J. (2015). Breeding schemes for the implementation of genomic selection in wheat (*Triticum spp.*) *Plant Science* 242: 23-26.

Braun, H. J., Rajaram, S. and Van Ginkel, M. (1996). CIMMYT's approach to breeding wheat for wide adaptation. Euphytica. 92:175 – 183

Bustos-Korts, D., M. P. Boer, J. Layton, A. Gehringer, T. Tang *et al.*, 2022 Identification of environment types and adaptation zones with self-organizing maps; applications to sunflower multi-environment data in Europe. Theoretical and Applied Genetics. 135:2059 – 2082

Cooper M, Powell O, Voss-Fels KP, Messina CD, Gho C, Podlich DW, *et al.* (2021). Modelling selection response in plant-breeding programs using crop models as mechanistic gene-to-phenotype (CGM-G2P) multi-trait link functions. In Silico Plants 3.

Cornelius, P.L., J. Crossa, M. Seyedsadr (1996). Statistical tests and estimators for multiplicative models for cultivar trials. In Kang, M.S., and Gauch, H.G., Jr., (Eds.), Genotype-by-Environment Interaction. Boca Raton:CRC Press, pp. 199-234.

Costa-Neto, G., J. Crossa, and R. Fritsche-Neto(2021a) Enviromic Assembly Increases Accuracy and Reduces Costs of the Genomic Prediction for Yield Plasticity in Maize. Frontiers in Plant Science 12:.

Costa-Neto, G., R. Fritsche-Neto, and J. Crossa, (2021b) Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. Heredity (Edinb) 126: 92–106.

Costa-Neto, G., G. Galli, H. F. Carvalho, J. Crossa, and R. Fritsche-Neto, 2021c EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. G3 Genes|Genomes|Genetics.

Costa-Neto, G. M. F., O. P. Morais Júnior, A. B. Heinemann, A. P. de Castro, and J. B. Duarte, (2020) A novel GIS-based tool to reveal spatial trends in reaction norm: upland rice case study. *Euphytica* 216(3). DOI:10.1007/s10681-020-2573-4

Crespo-Herrera LA, Crossa J, Huerta-Espino J, Mondal S, Velu G, Juliana P, Vargas M, Pérez-Rodríguez P, Joshi AK, Braun HJ and Singh RP (2021). Target population of environments for wheat breeding in India: Definition, prediction and genetic gains. Frontiers in Plant Science 12: 638520.

Crossa J. and P. L. Cornelius (1997). Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances. *Crop Science* 37:406-415.

Crossa, J., M. Vargas, F. A. van Eeuwijk, C. Jiang, G. O. Edmeades et al., (1999). Interpreting genotype by environment interaction in tropical maize using linked molecular markers

and environmental covariates. *Theor Appl Genet* 99: 611–625. https://doi.org/10.1007/s001220051276

Crossa, J., Yang, R.-C., and Cornelius, P.L.. (2004). Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. J. Agric. Biol. Environ. Stat. 9: 362– 380. doi: 10.1198/108571104X4423.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O.A., Jarquín, D., de Los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R.K. 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci*., 22(11):961-975.

Crossa J, Fritsche-Neto R, Montesinos-Lopez OA, Costa-Neto G, Dreisigacker S, Montesinos-Lopez A and Bentley AR (2021) The Modern Plant Breeding Triangle: Optimizing the Use of Genomics, Phenomics, and Enviromics Data. Front. Plant Sci. 12:651480. doi: 10.3389/fpls.2021.651480

Crossa, J., O. Montesinos-López, P. Pérez-Rodríguez, G. Costa-Neto, R. Fritsche-Neto *et al.*, (2022) Genome and Environment Based Prediction Models and Methods of Complex Traits Incorporating Genotype × Environment Interaction, pp. 245–283 in *Genomic Prediction of Complex Traits*, edited by N. Ahmadi and J. Bartholomé. Humana Press Springer, New York

Cuevas, J., Crossa,J., Soberanis, V.,Pérez-Elizalde, S., Pérez-Rodríguez, P., de los Campos, G., Montesinos-López, O. A., Burgueño, J. (2016). Genomic prediction of genotype × environment interaction kernel regression models. *The Plant Genome* 9(3):1:20.

Cuevas, J., Crossa,J., Montesinos-López, O. A., Burgueño, J., Pérez-Rodríguez, P., de los Campos, G. (2017). Bayesian Genomic prediction with genotype × environment kernel models. *G3: Genes|Genomes|Genetics, 7(1): 41-53*

Cuevas, J., Granato, I., Fritsche-Neto, R., Montesinos-Lopez, O. A., Burgueño, J., Bandeira e Sousa, M., Crossa, J. (2018). Genomic-Enabled Prediction Kernel Models with Random Intercepts for Multi-environment Trials. Genes, Genomes and Genetics, 8(4): 1347-1365.

Cuevas, J., Montesinos-López, O.A., Juliana, P., Guzmán, C., Pérez-Rodríguez, P., González-Bucio, J., Burgueño, J., Montesinos-López, A., and Crossa, J. (2019). Deep Kernel for Genomic and Near Infrared Predictions in Multi-environment Breeding Trials. G3-Genes Genomes Genetics. 9(9):2913-2924.

Denis J-B (1988) Two-way analysis using covariates. Statistics 19:123–132

de los Campos, G., Pérez-Rodríguez, P., Bogard, M. *et al.* A data-driven simulation platform to predict cultivars' performances under uncertain weather conditions. *Nat Commun* **11,** 4876 (2020). https://doi.org/10.1038/s41467-020-18480-y

Eberhart S.A., Russell W.A. (1966). Stability parameters for comparing varieties. Crop Sci., 6: 36–40.

Elli EF, Huth N, Sentelhas PC, et al (2020) Global sensitivity-based modelling approach to identify suitable Eucalyptus traits for adaptation to climate variability and change. In Silico Plants **2**, 1. https://doi.org/10.1093/insilicoplants/diaa003

Finlay, K.W. and Wilkinson, G.N. (1963). The Analysis of Adaptation in a Plant-Breeding Programme. Australian Journal of Agricultural Research, 14, 742-754. http://dx.doi.org/10.1071/AR9630742

Gauch, H.G. (1988) Model Selection and Validation for Yield Trials with Interaction. Biometrics, 44, 705-715. http://dx.doi.org/10.2307/2531585.

Granato, I., J. Cuevas, F. Luna-Vázquez, J. Crossa, O. Montesinos-López et al., (2018). BGGE: A new package for genomic-enabled prediction incorporating genotype × environment interaction models. G3 Genes, Genomes, Genet. 8: 3039–3047. https://doi.org/10.1534/g3.118.200435.

Hardwick, R C, and Wood, J T. (1972). Regression methods for studying genotype-environment interactions. *Heredity*, **28**, 209–222.

He, S., R. Thistlethwaite, K. Forrest, F. Shi, M. J. Hayden et al., 2019 Extension of a haplotype-based genomic prediction model to manage multi-environment wheat data using environmental covariates. Theoretical and Applied Genetics 132: 3143–3154.

Heinemann AB, Ramirez-Villegas J, Rebolledo MC, Costa Neto GMF and Castro AP (2019) Upland rice breeding led to increased drought sensitivity in Brazil. Field Crops Research 231: 57-67

Heinemann, A. B., G. Costa-Neto, R. Fritsche-Neto, D. H. da Matta, and I. K. Fernandes, 2022 Enviromic prediction is useful to define the limits of climate adaptation: A case study of common bean in Brazil. Field Crops Research 286: 108628.

Helland, I. (1988), ''On the structure of partial least squares regression,'' Communications in Statistics, Simulation and Computation, 17(2), 581- 607.

Heslot, N., D. Akdemir, M. E. Sorrells, and J.-L. Jannink, 2014 Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theoretical and Applied Genetics 127: 463–480.

Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J et al. (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607

Jarquin, D., Howard, R., Crossa, J., Beyene, Y., Gowda, M., Martini, J. W. R., et al. (2020). Genomic prediction enhanced sparse testing for multi-environment trials. G3 Genes|Genomes|Genet. 10, 2725–2739. doi: 10.1534/g3.120. 401349

Lê, S., Josse, J. & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software. 25(1). pp. 1-18.

Li, X., Guo, T., Wang, J., Bekele, W.A., Sukumaran, S., Vanous, A.E., McNellie, J.P., Cortes, L.T., Lopes, M.S., Lamkey, K.R. et al. (2021). An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops Mol. Plant, 14 (2021), pp. 874-887.

Martini, J. W. R., M. F. Schrauf, C. A. Garcia-Baccino, E. C. G. Pimentel, S. Munilla et al., 2018 The effect of the H -1 scaling factors τ and ω on the structure of H in the single-step procedure. Genetics Selection Evolution 50:.

Messina, C. D., Technow, F., Tang, T., Totir, R., Gho, C., and Cooper, M. (2018). Leveraging biological insight and environmental variation to improve phenotypic prediction: integrating crop growth models (CGM) with whole genome prediction (WGP). *Eur. J. Agron.* 100, 151–162. doi: 10.1016/j.eja.2018.01.007

Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, Charcosset A, Welcker C, van Eeuwijk F and Tardieu F (2019). Genomic prediction of maize yield across European environmental conditions. Nature Genetics 51: 952-956.

Montesinos-López, O. A., A. Montesinos-López, Kismiantini, A. Roman-Gallardo, K. Gardner et al., 2022 Partial Least Squares Enhances Genomic Prediction of New Environments. Frontiers in Genetics 13:920689

Monteverde E, Gutierrez L, Blanco P, Pérez de Vida F, Rosas JE, Bonnecarrère V, Quero G and McCouch S (2019). Integrating molecular markers and environmental covariates to interpret genotype by environment interaction in rice (Oryza sativa L.) grown in subtropical areas. G3: Genes, Genomes, Genetics 9: 1519-1531.

Morais Júnior, O. P., Duarte, J. B., Breseghello, F., Coelho, A. S. G., and Magalhães Júnior, A. M. (2018). Single-step reaction norm models for genomic prediction in multienvironment recurrent selection trials. *Crop Sci*. 58, 592–607. doi:10.2135/cropsci2017.06.0366.

Morisse M, Wells DM, Millet EJ, Lillemo M, Fahrner S, Cellini F, Looten P, Muller O, Herrera JM, Bentley AR, Janni M. (2021). A European perspective on opportunities and demands for field-based crop phenotyping. *Field Crops Research* https://doi.org/10.1016/j.fcr.2021.108371.

Mu, Q., T. Guo, X. Li, and J. Yu, 2022 Phenotypic plasticity in plant height shaped by interaction between genetic loci and diurnal temperature range. New Phytologist 233: 1768–1779.

Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink JL, Singh RP, Autrique E, de los Campos G. Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 (Bethesda). (2015) 5(4):569-82. doi: 10.1534/g3.114.016097. PMID: 25660166; PMCID: PMC4390573.

Ornellas, L., Kruseman, G., and Crossa, J. Satellite Data and Supervised Learning to Prevent Impact of Drought on Crop Production: Meteorological Drought. 2019. IntechOpen, The Shard, 32 London Bridge Street, London SE1 9SG, United Kingdom. DOI: 10.5772/intechopen.85471.

Palermo, G., Piraino, P., & Zucht, H. D. (2009). Performance of PLS regression coefficients in selecting variables for each response of a multivariate PLS for omics-type data. *Advances and applications in bioinformatics and chemistry : AABC*, *2*, 57–70. https://doi.org/10.2147/aabc.s3619

Pérez-Rodríguez, P., J. Crossa, K. Bondalapati, G. De Meyer, F. Pita, and G.D.L. Campos. 2015. A pedigree-based reaction norm model for prediction of cotton yield in multienvironment trials. Crop Sci. 55:1143–1151. doi:10.2135/ cropsci2014.08.0577

Porker K, Coventry S, Fettell NA, Cozzolino D and Eglinton J (2020) Using a novel PLS approach for envirotyping of barley phenology and adaptation. Field Crops Research 246: 1-11

Rawson, H. M., and H. G. Macpherson, 2000 Irrigated wheat : managing your crop (H. M. Rawson & H. G. Macpherson, Eds.). Food and Agriculture Organization of the United Nations, Rome.

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. Austria. ISBN 3-900051-07-0. URL http://www.R-project.org/

Rajaram, S., Ginkel, M., Fischer, R. A. (1994). CIMMYT's wheat breeding Mega-environments (ME). In Proceedings of the 8th International Wheat Genetic Symposium. Beijing, China. 1101 - 1106 p

Rogers AR, Dunne JC, Romay C, Bohn M, Buckler ES, Ciampitti IA, Edwards J, Ertl D, Flint-Garcia S, Gore MA, Graham C, Hirsch CN, Hood E, Hooker DC, Knoll J, Lee EC, Lorenz A, Lynch JP, McKay J, Moose SP, Murray SC, Nelson R, Rocheford T, Schnable JC, Schnable PS, Sekhon R, Singh M, Smith M, Springer N, Thelen K, Thomison P, Thompson A, Tuinstra M, Wallace J, Wisser RJ, Xu W, Gilmour AR, Kaeppler SM, De Leon N and Holland JB (2021). The importance of dominance and genotype-by-environment

interactions on grain yield variation in a large-scale public cooperative maize experiment. G3: Genes, Genomes, Genetics 11: jkaa050.

Rogers, A. R., and J. B. Holland, 2022 Environment-specific genomic prediction ability in maize using environmental covariates depends on environmental similarity to training data. G3: Genes, Genomes, Genetics 12:.

Resende RT, Piepho HP, Rosa GJM, Silva-Junior OB, E Silva FF, de Resende MDV, Grattapaglia D. (2021). Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. Theor Appl Genet. 134(1): 95-112. doi: 10.1007/s00122-020-03684-z.

Sanchez, G. (2012). plsdepot: Partial Least Squares (PLS) Data Analysis Methods. R package version 0.1.17. https://CRAN.R-project.org/package=plsdepot, DOI: 10.1166/jnn.2012.6776

Teixeira, J., S. Cardoso, M. Bonazzola, J. Cole, A. Del Genio, C. DeMott, C. Franklin, C. Hannay, C. Jakob, Y. Jiao, J. Karlsson, H. Kitagawa, M. Köhler, A. Kuwano-Yoshida, C. LeDrian, J. Li, A. Lock, M.J. Miller, P. Marquet, J. Martins, C.R. Mechoso, E.van Meijgaard, I. Meinke, P.M.A. Miranda, D. Mironov, R. Neggers, H.L. Pan, D.A. Randall, P.J. Rasch, B. Rockel, W.B. Rossow, B. Ritter, A.P. Siebesma, P.M.M. Soares, F.J. Turk, P.A. Vaillancourt, A. Von Engeln, and M. Zhao, (2011). Tropical and sub-tropical cloud transitions in weather and climate prediction models: The GCSS/WGNE Pacific Cross-section Intercomparison (GPCI). *J. Climate*, **24**, 5223-5256, doi:10.1175/2011JCLI3672.1.

Vargas, M., J. Crossa, K. Sayre, M. Reynolds, M. E. Ramírez et al., (1998). Interpreting genotype · environment interaction in wheat by partial least squares regression. Crop Sci. 38: 679–687. https://doi.org/ 10.2135/cropsci1998.0011183X003800030010x

Vargas, M., J. Crossa, F. A. van Eeuwijk, M. Ramírez, and K. Sayre, (1999). Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype ×environment interaction.

Wang, E., and T. Engel, 1998 Simulation of Phenological Development of Wheat Crops. Agricultural Systems 58: 24.

Washburn JD, Cimen E, Ramstein G, Reeves T, O'Briant P, McLean G, Cooper M, Hammer G, Buckler ES. (2021). Predicting phenotypes from genetic, environment, management, and historical data using CNNs. Theor Appl Genet. Dec;134(12):3997-4011. doi: 10.1007/s00122-021-03943-7. Epub 2021 Aug 27. PMID: 34448888.

Westhues, C. C., Simianer, H., & Beissinger, T. M. (2021). learnMET: an R package to apply machine learning methods for genomic prediction using multi-environment trial data. *bioRxiv*.

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink JL and Sorrells ME (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3: Genes, Genomes, Genetics 2: 1427-1436.

Yates, F. and Cochran, W.G. (1938). The Analysis of Groups of Experiments. Journal of Agricultural Science, 28, 556-580. http://dx.doi.org/10.1017/S0021859600050978

Zadoks, J. C., T. T. CHANGt, C. F. Konzak, and J. D. Fryer, 1974 A decimal code for the growth stages of cereals.

Table 1. Modeling assumptions for each kernel ($\boldsymbol{K_G}, \boldsymbol{K_E}$ and $\boldsymbol{K_{GE}}$) and model (M01-M06). Details about each genetic ($\boldsymbol{G}, \boldsymbol{H}, \boldsymbol{R}$) and environmental relationships ($\boldsymbol{\Omega}, \boldsymbol{\gamma}, \boldsymbol{\Phi}$) are given in the following sections. $\otimes$ denotes the Kronecker product.

Kernel assumption	**Model**					
	M01	**M02**	**M03**	**M04**	**M05**	**M06**
$\boldsymbol{K_G} = (\boldsymbol{J_q} \otimes \boldsymbol{G})$	x	x	x	x	x	x
$\boldsymbol{K_E} = (\boldsymbol{I}_q \otimes \boldsymbol{J_p})$	x					
$\boldsymbol{K_E} = (\boldsymbol{\Omega} \otimes \boldsymbol{J_p})$		x				
$\boldsymbol{K_E} = (\boldsymbol{\gamma} \otimes \boldsymbol{J_p})$			x			
$\boldsymbol{K_E} = (\boldsymbol{\Phi} \otimes \boldsymbol{J_p})$				x	x	x
$\boldsymbol{K_{GE}} = (\boldsymbol{I_q} \otimes \boldsymbol{G})$	x					
$\boldsymbol{K_{GE}} = (\boldsymbol{\Omega} \otimes \boldsymbol{G})$		x				
$\boldsymbol{K_{GE}} = (\boldsymbol{\gamma} \otimes \boldsymbol{G})$			x			
$\boldsymbol{K_{GE}} = (\boldsymbol{\Phi} \otimes \boldsymbol{G})$				x		
$\boldsymbol{K_{GE}} = \boldsymbol{K_H} = (\boldsymbol{J_q} \otimes \boldsymbol{H})$					x	
$\boldsymbol{K_{GE}} = f(\boldsymbol{M} = \boldsymbol{\gamma} \otimes \boldsymbol{R}, \boldsymbol{N} = \boldsymbol{\Phi} \otimes \boldsymbol{G})$ [Eq. 12,13]						x

Figure 1. Environmental characterization of the CIMMYT's Indian locations using EPA analysis covering the seasonal effects across 36 locations over the years between 2011 and 2018. (A) Summary of the relative variable importance on projection (VIP) in defining the environmental similarity across locations, considering the main environmental variables and development stages (panels) after quality control. A horizontal solid line indicates the threshold of VIP % = 85%, which corresponds to the absolute value relative to the 95% of probability in the distribution of computed VIP values. **(B)** Principal component analysis and clusters found by the Ward Algorithm using the environmental weight as covariables; **(C)** Environmental signature of each group found considering the environmental weights ECs with VIP > 85%. Details about the signatures for all ECs are given in Supplementary Figure S2. Environmental weights equal to 1.0 (horizontal solid line) denote that the given EC is not weighted by EPA. Each dot represents the actual weight of the respective EC for a given a location within the group. The smooth line represents the overall trends considering the key drivers (red dots found in the label A).

Figure 2. Diagnosis of the environmental diversity observed within each cluster of locations found in this study for CIMMYT's wheat trials in India. (A) Relative similarity between each location and its pivot-locations within each cluster. Each pivot was selected by the clustering analysis of the environmental weights from the EPA analysis using the multivariate PLS 2 algorithm. **(B)** General overview of the representativeness of the multi-environment trial (MET) network in India for each season (year). Values above (n=3, n=4, etc.) indicate the number of pivot locations considered in each year. Red colors denote the proportion of essential locations (pivot) in each year. Blue colors denote the proportion of environments somehow related to some pivot (similarity >= 95% with the pivot), using as threshold 95%. Gray colors denote the proportion of environments considered to represent the diversity of the MET.

Figure 3. Distribution of the relative variable importance in projection (VIP, %) from the PLS 1 analysis for each genotype considering each envirotyping covariable (environmental factor x development stage) and years between 2011 and 2018. (**A**) detail of VIP distribution considering all genotypes and detailing for each year and development stage. (**B**) general overview considering all years. The vertical red line denotes the VIP = 80%. Blue denote those variables with higher importance (VIP > 80%) for the given year/development stage combination.

Figure 4. Characterization of the reaction-norms the subsequent similarity of the plasticity patterns for CIMMYT's wheat germplasm in India. (A) panel of genotype-specific coefficients for each genotype (columns) and envirotyping covariables (rows) across from the EPA analysis over the G+GE effects across years; **(B)** genetic similarity (gaussian kernel) of the empirical plasticity computed from the reaction-norm coefficients treated as "plasticity markers". **(C)** patterns of similarity between the germplasm tested in a given year (from 2011 to 2018) and the all germplasm (n=360).

Figure 5. Predictive ability for grain yield considering each statistical model and prediction scenario of G×Y between 2011 and 2018 in India. (**A**) Scenario 1, considering all TPEs together and a pair of years (the previous year as the training set, the following year as the testing set). (**B**) Scenario 2, considering all TPEs together and three sets of years (pairs of subsequent previous years as the training set, a single following year as the testing set). (**C**) Scenario 3, for each TPE, considering a pair of years (the previous year as the training set, the following year as the testing set). (**D**) Scenario 4, for each TPE, considering three sets of years (pairs of subsequent previous years as the training set, a single following year as the testing set). The white points within each boxplot represent the average value (mean), while the horizontal lines represent the median (quantile 50%). Vertical lines denote the reference of not predictability (null predictability = 0, dashed black line) and the reference of the average predictability in terms of the Spearman rank correlation of the baseline model (M01, no enviromics, red line). Percentage values above each boxplot represent the average gain/loss in predictive ability in relation to the baseline M01

Figure 6. Summary of the TPE-specific predictions considering the years between 2011 and 2018 in India. (**A**) Average predictive ability (PA) for grain yield for each statistical model and prediction scenario. Vertical lines denote the reference of not predictability (null predictability = 0, dashed black line), average predictive value for each panel and TPE, considering all models (solid black lines), and the reference of the average predictability of baseline model (M01, red line). (**B**) head-to-head relative frequency of the PA superiority of each model (M02-M06) over the baseline GBLUP (M01) across years, with "better than M01" (green colors) indicating how many times each one of the tested models (M02-M06) outperformed M01 and "worst than M01" (red color) relating on how many times the M01 was the winner model.

Figure 7. Resolution of the genomic prediction models for each scenario tested. (**A**) heatmap of the spearman's rank correlation for each genotype in prediction Scenario 1. (**B**) Resolution in terms of the distribution frequency of predictability for Scenario 1. (**C**) Heatmap of Spearman's rank correlation for each genotype in prediction Scenario 2. (**D**) Resolution in terms of the distribution frequency of predictability for Scenario 2. (**E**) Heatmap of Spearman's rank correlation for each genotype in prediction Scenario 3. (**F**) Resolution in terms of the distribution frequency of predictability for Scenario 3. (**G**) Heatmap of Spearman's rank correlation for each genotype in prediction Scenario 4. (**H**) Resolution in terms of the distribution frequency of predictability for Scenario 4.

Appendix 1: PLS for EPA

PLS regression is a technique used to combine the benefits of conventional ordinary least squares (OLS) and principal component analysis (Aastveit and Martens, 1986). Thus, it is useful for both predictive and exploratory analyses. Another important use of PLS analyses relies on the integration of environmental information in systems

biology (Teixeira et al., 2011) and G×E analysis, this latter as an alternative to deal with the higher and natural collinearity of environmental factors and explain possible environmental causes of adaptation (Vargas et al., 1999; Monteverde et al., 2019; Porker et al., 2020). In fact, the first mention of "enviromics" used this technique to support "reverse engineering" of the genomic features related to the dynamics within the cell environment (Teixeira et al., 2011). Due to its popularity and simplicity, we chose this approach to perform the so-called Envirotype-Phenotype Association (EPA) in two ways: (PLS 1) to estimate "adaptability descriptors" of the genotypes, that is, genotype-specific coefficients for each genotype and environmental factor; and (PLS 2) to extract weights of environmental diversity, that is, associating environmental factors and the observed phenotypic-based similarity between environments. For both purposes, a generic representation of the PLS model is given below and its particular use in supporting the development of "enviromic-aided pedigrees" will be discussed in the next sections.

In the PLS algorithm, a matrix of predictors (**X,** q environments × m variables) is decomposed into three matrices: (1) a matrix of scores (**T**) (referred to as X-score), (2) a matrix of loadings ($\boldsymbol{P}'$, X-loading, $m \times a$) and (3) the residuals of this decomposition (**C**); thus:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}' + \boldsymbol{C} \qquad \text{(Eq. 1, Appendix)}$$

Simultaneously, the matrix of responses ($\boldsymbol{Y}$) is decomposed into the three matrices of Y-scores (**U** matrix), Y-loadings ($\mathbf{Q}'$ matrix), and the Y-residuals (denoted here as $\boldsymbol{F}$); thus:

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}' + \boldsymbol{F} \qquad \text{(Eq. 2, Appendix)}$$

To estimate adaptability descriptors (PLS 1), we adopted the univariate PLS approach, considering the vector of phenotypes for each genotype (**Y** assumed as a vector of environmental-centered values for each genotype at each q environment × 1) of all observations across all test environments. In a certain way, this is a site-regression

approach (SREG) (Crossa and Cornelius, 1997), where the main genetic effect plus genotype-specific G×E variation (G+GE) were attributed to the **y** vector. Here we are interested in modeling the combination of broad genetic effect (model intercept) plus a GE in terms of reaction-norms (Costa-Neto et al., 2020). To implement this, each model accounts for an independent single-genotype variation analysis over a matrix of environmental descriptors (***W***-matrix for environments, with q environments × k covariables). More details are given in the description of M05 in Genomics Prediction section.

On the other hand, for extracting the weights of environmental diversity, we considered the multivariate PLS approach (**Y** as a matrix, with dimension $q \times q$, here q = number of locations), associating the long-term environmental descriptors for a given location (here a W-matrix for locations, with q locations × k covariables), and a matrix of phenotypic-based environmental similarity (S_0-matrix) for each location. More details are given in the description of M05 in Genomics Prediction section.

In summary, the PLS algorithm aims to minimize the observed norm of ***F***, while looking to maximize the correlation between ***X*** and ***Y*** by the inner relation ***U***=***TD***, where ***D*** is a diagonal matrix. Consequently, the X-scores are orthogonal, assuring the control of collinearity among the original ***X*** predictors. This process results in the estimation of linear combinations of the original variables, computed from a matrix of weights (***L***), with a dimension of $a \times b$, where a is the number of rows in the ***X*** matrix and b is the number of components (latent vectors) considered, with ***T***=***XL***. Finally, the linear coefficients to relate **X** and **Y** (referred here as **B**) are computed by using the iterative relation:

$$(1): \boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}^{\boldsymbol{T}} + \boldsymbol{F}$$

$$(2): \boldsymbol{Y} = \boldsymbol{T}\boldsymbol{D}\boldsymbol{Q}^{\boldsymbol{T}} + (\boldsymbol{P}\boldsymbol{Q}^{\boldsymbol{T}} + \boldsymbol{F})$$

$$(3): \boldsymbol{Y} = \boldsymbol{T}\boldsymbol{K}^{\boldsymbol{T}} + \boldsymbol{F}^{*}$$

$$(4): \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{L}^{*}\boldsymbol{K}^{\boldsymbol{T}} + \boldsymbol{F}^{*}$$

(5): $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{F}^*$

Thus, in summary,

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{F}^* \quad \text{(Eq. 4)}$$

where the matrix of coefficients is computed by $\boldsymbol{B} = \boldsymbol{LK}^{\boldsymbol{T}}$, and $\boldsymbol{F}^*$ is the residual variance in $\boldsymbol{Y}$ not captured by the core of latent vectors (hereafter called LV). Here we adopted the nonlinear iterative partial least squares (NIPALS) algorithm to sequentially extract the PLS components. Details on the NIPALS algorithm can be found in Palermo et al. (2009) and Sanchez (2012).

Appendix 2: Genomic Prediction Models

In the Appendix Table A1 we presented a summary of all genomic prediction models in terms of basic data input (genomics, enviromics), its adoption or not of EPA outcomes, which relationship matrices were considered and the kernel assumptions.

Appendix Table A1: Detailed model assumptions considered in this study.

	Model					
	M01	**M02**	**M03**	**M04**	**M05**	**M06**
Basic Data Input						
Marker/ SNP	x	x	x	x	x	x
Envirotyping		x	x	x	x	x
Environmental-Phenotype Associations (EPA)						
EPA: PLS1 - Reaction-norms ($\boldsymbol{\Lambda}$)					x	x
EPA: PLS2 - Environmental Weights ($\boldsymbol{\varphi}$)				x	x	x
Relationship Matrices						
Genomic Relationship Matrix (**G**)	x	x	x	x	x	x
Genotype-Specific Reaction-Norm (**R**)					x	x
Identity environmental relationship (ERM, $\boldsymbol{I}_q$)	x					
Linear ERM using **W**-matrix ($\boldsymbol{\Omega}$)		x				

Nonlinear ERM using **W**-matrix ($\boldsymbol{\gamma}$)			x			
Nonlinear ERM using env. weights ($\boldsymbol{\Phi}$)				x	x	
G-matrix weighted by reaction-norms ($\boldsymbol{H}$)					x	
Past G×E (historical data set) by $\boldsymbol{M} = \boldsymbol{\gamma} \otimes \boldsymbol{R}$						x
Future (expectation) of G×E by $\boldsymbol{N} = \boldsymbol{\Phi} \otimes \boldsymbol{G}$						x
Kernels assumptions for $\boldsymbol{K_G}, \boldsymbol{K_E}$ and $\boldsymbol{K_{GE}}$						
$\boldsymbol{K_G} = (\boldsymbol{J_q} \otimes \boldsymbol{G})$	x	x	x	x	x	x
$\boldsymbol{K_E} = (\boldsymbol{I}_q \otimes \boldsymbol{J_p})$	x					
$\boldsymbol{K_E} = (\boldsymbol{\Omega} \otimes \boldsymbol{J_p})$		x				
$\boldsymbol{K_E} = (\boldsymbol{\gamma} \otimes \boldsymbol{J_p})$			x			
$\boldsymbol{K_E} = (\boldsymbol{\Phi} \otimes \boldsymbol{J_p})$				x	x	x
$\boldsymbol{K_{GE}} = (\boldsymbol{I_q} \otimes \boldsymbol{G})$	x					
$\boldsymbol{K_{GE}} = (\boldsymbol{\Omega} \otimes \boldsymbol{G})$		x				
$\boldsymbol{K_{GE}} = (\boldsymbol{\gamma} \otimes \boldsymbol{G})$			x			
$\boldsymbol{K_{GE}} = (\boldsymbol{\Phi} \otimes \boldsymbol{G})$				x		
$\boldsymbol{K_{GE}} = \boldsymbol{K_H} = (\boldsymbol{J_q} \otimes \boldsymbol{H})$					x	
$\boldsymbol{K_{GE}} = f(\boldsymbol{M} = \boldsymbol{\gamma} \otimes \boldsymbol{R}, \boldsymbol{N} = \boldsymbol{\Phi} \otimes \boldsymbol{G})$ [Eq 12]						x

Figure 1

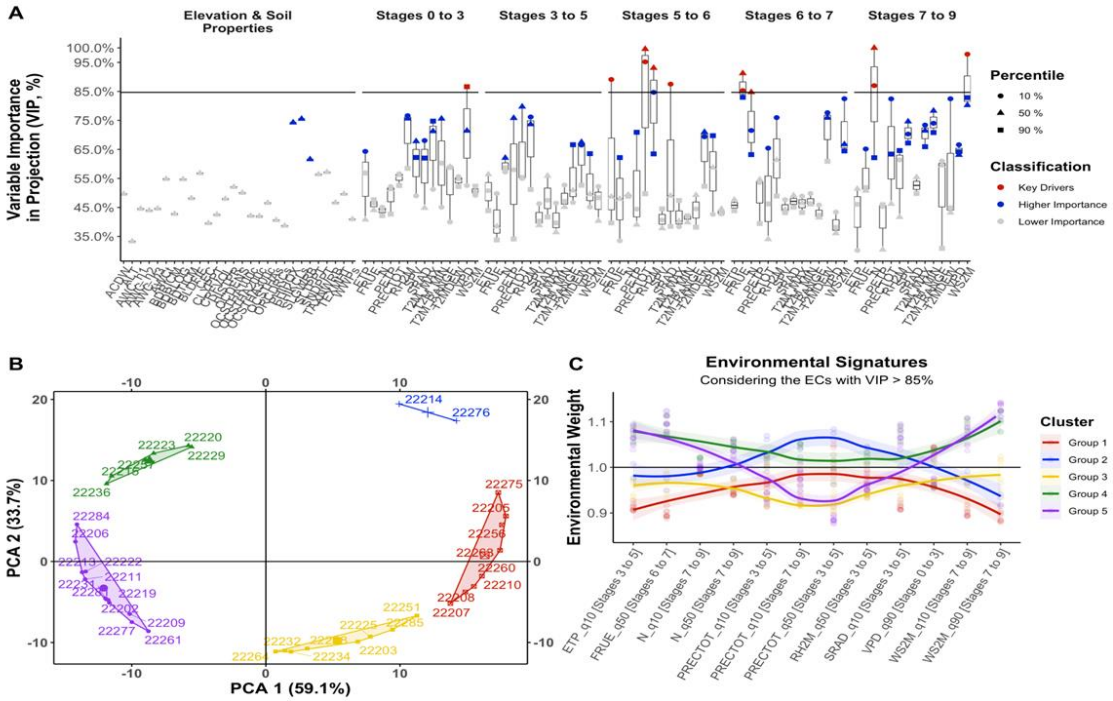

Figure 1
165x93 mm (.56 x DPI)

Figure 2

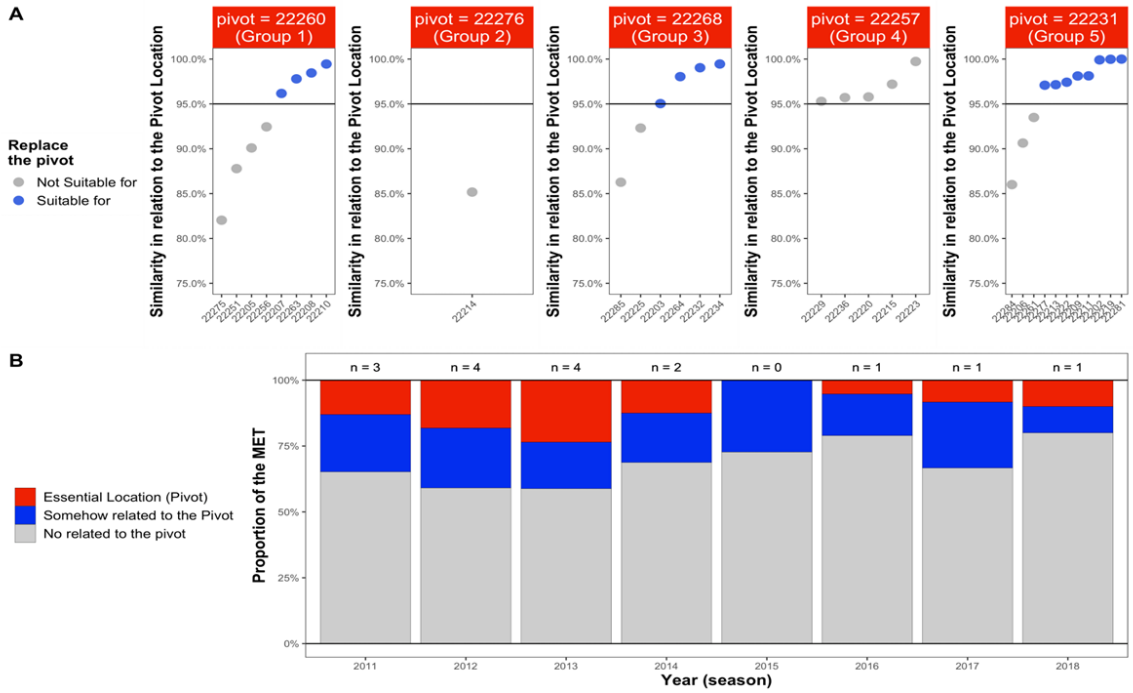

Figure 2
165x93 mm (.56 x DPI)

Figure 3

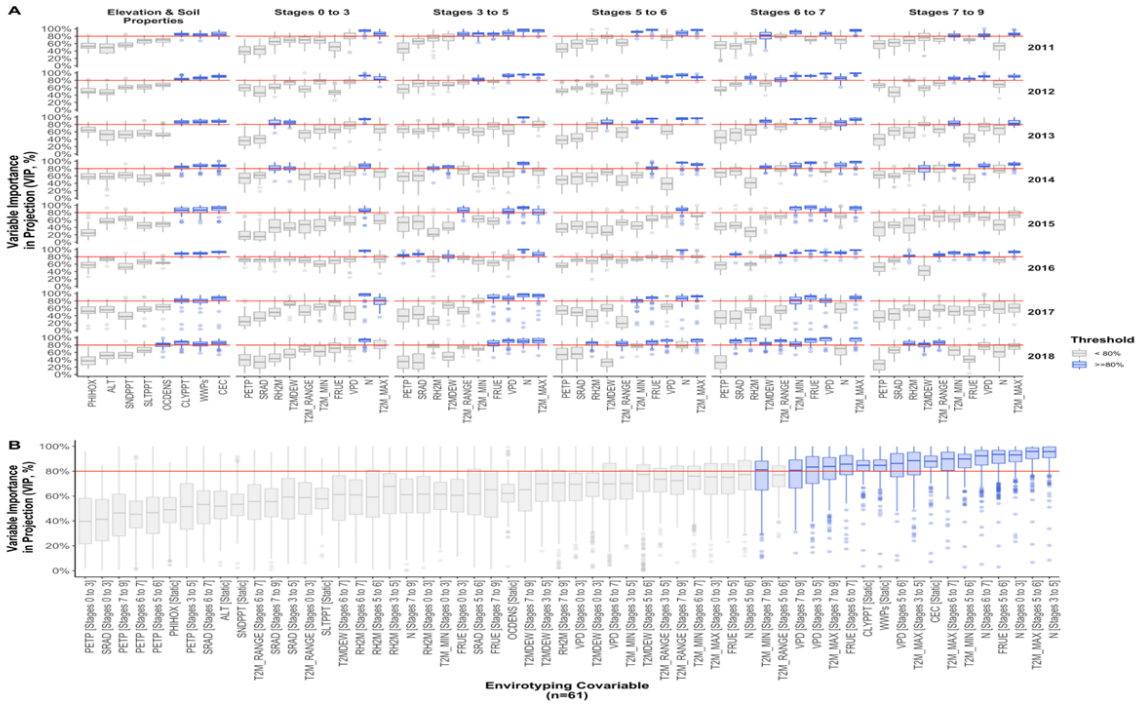

Figure 3
165x93 mm (.56 x DPI)

Figure 4

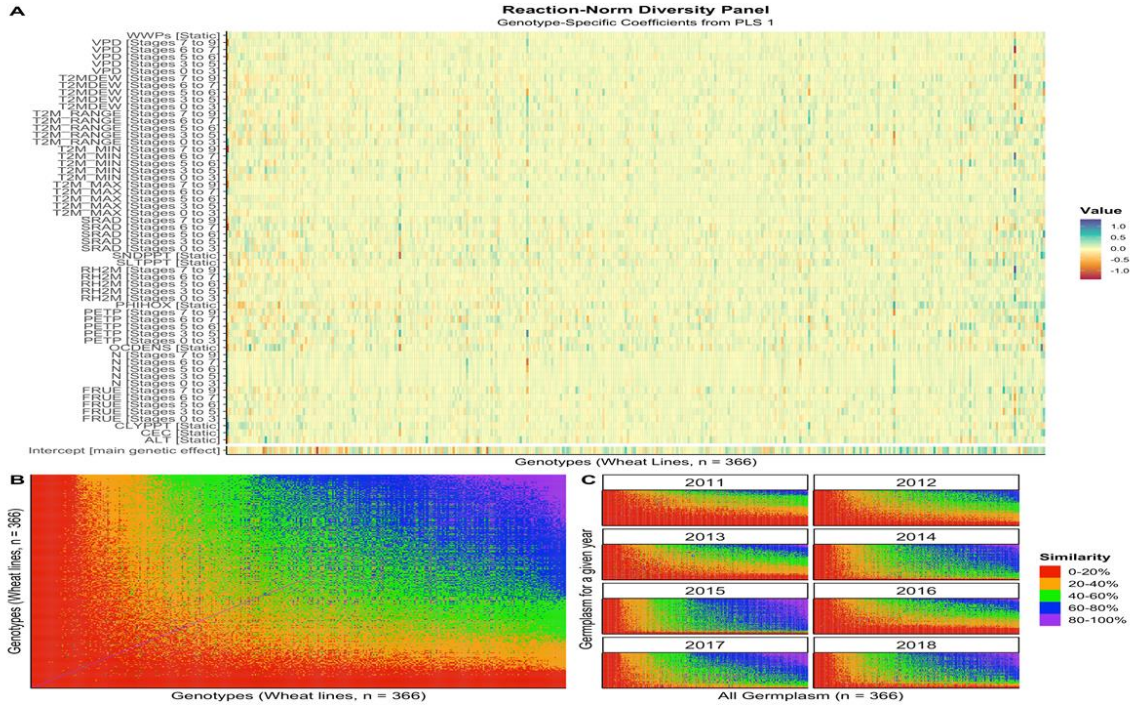

Figure 4
165x93 mm (.56 x DPI)

Figure 5

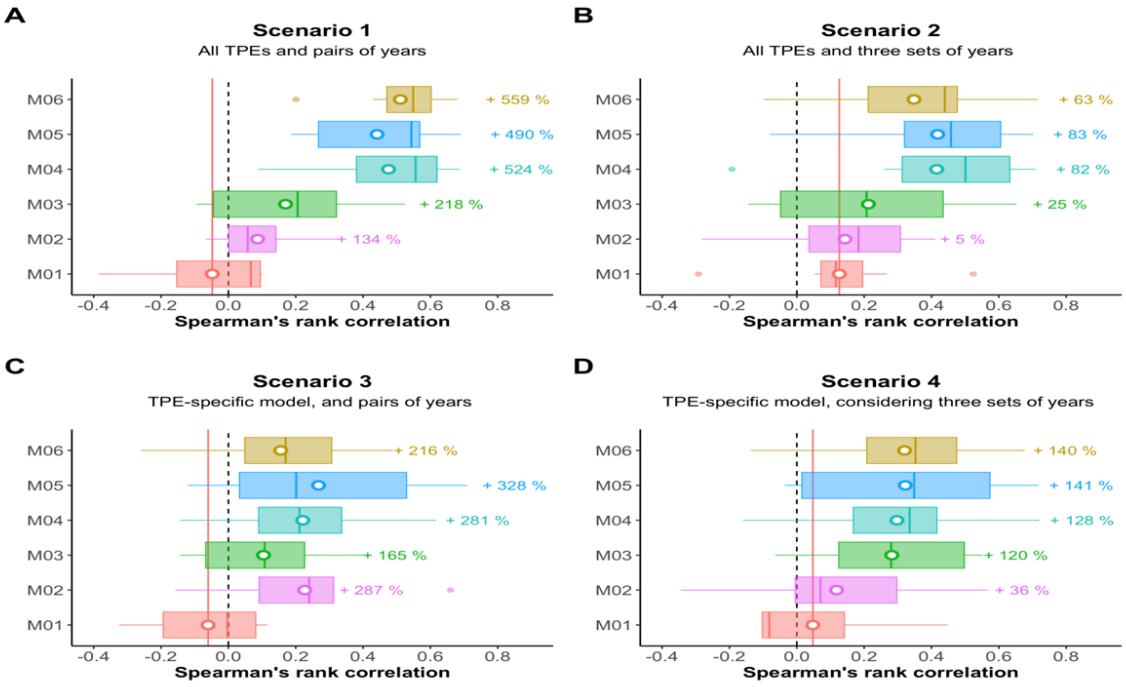

Figure 5
165x93 mm (.56 x DPI)

Figure 6

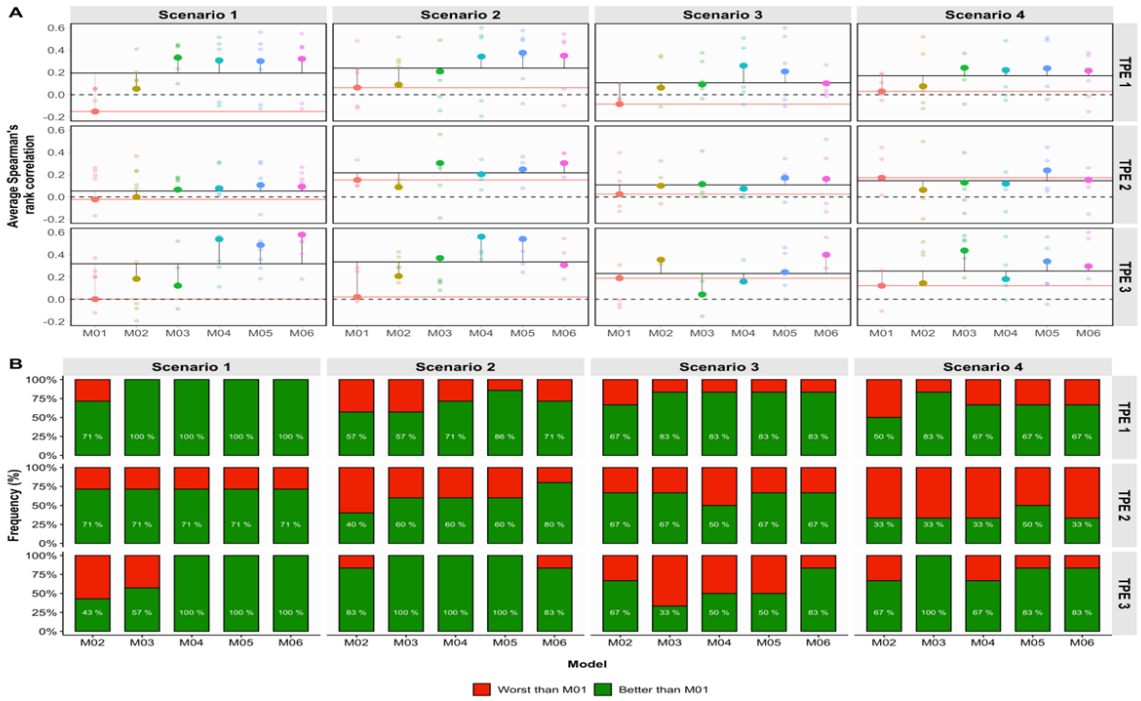

Figure 6
165x93 mm (.56 x DPI)

Figure 7

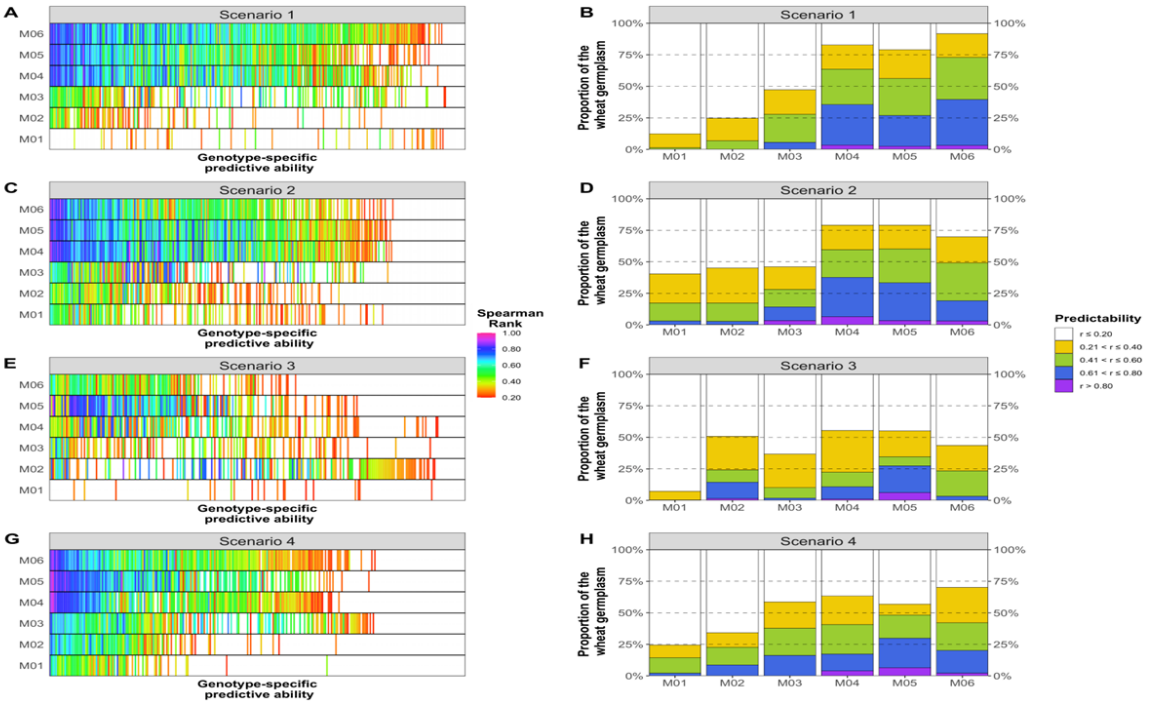

Figure 7
165x93 mm (.56 x DPI)