



OPEN ACCESS

EDITED BY

Leif Skot,
Aberystwyth University, United Kingdom

REVIEWED BY

Gurjeet Singh,
Texas A and M University, United States
Jérôme Bartholomé,
CIRAD, France

*CORRESPONDENCE

Roberto Fritsche-Neto
✉ rfneto@agcenter.lsu.edu

RECEIVED 02 July 2024

ACCEPTED 07 October 2024

PUBLISHED 23 October 2024

CITATION

Prado M, Famoso A, Guidry K and
Fritsche-Neto R (2024) Optimizing multi-
environment trials in the Southern US Rice
belt via smart-climate-soil prediction-based
models and economic importance.
Front. Plant Sci. 15:1458701.
doi: 10.3389/fpls.2024.1458701

COPYRIGHT

© 2024 Prado, Famoso, Guidry and
Fritsche-Neto. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Optimizing multi-environment trials in the Southern US Rice belt via smart-climate-soil prediction-based models and economic importance

Melina Prado¹, Adam Famoso², Kurt Guidry²
and Roberto Fritsche-Neto^{2*}

¹Department of Genetics, "Luiz de Queiroz" College of Agriculture/University of São Paulo, Piracicaba, Brazil, ²H. Rouse Caffey Rice Research Station, Louisiana State University Agricultural Center, Rayne, LA, United States

Rice breeding programs globally have worked to release increasingly productive and climate-smart cultivars, but the genetic gains have been limited for some reasons. One is the capacity for field phenotyping, which presents elevated costs and an unclear approach to defining the number and allocation of multi-environmental trials (MET). To address this challenge, we used soil information and ten years of historical weather data from the USA rice belt, which was translated into rice response based on the rice cardinal temperatures and crop stages. Next, we eliminated those highly correlated Environmental Covariates (ECs) (>0.95) and applied a supervised algorithm for feature selection using two years of data (2021-22) and 25 genotypes evaluated for grain yield in 18 representative locations in the Southern USA. To test the trials' optimization, we performed the joint analysis using prediction-based models in four different scenarios: i) considering trials as non-related, ii) including the environmental relationship matrix calculated from ECs, iii) within clusters; iv) sampling one location per cluster. Finally, we weigh the trial's allocation considering the counties' economic importance and the environmental group to which they belong. Our findings show that eight ECs explained 58% of grain yield variation across sites and 53% of the observed genotype-by-environment interaction. Moreover, it is possible to reduce 28% the number of locations without significant loss in accuracy. Furthermore, the US Rice belt comprises four clusters, with economic importance varying from 13 to 45%. These results will help us better allocate trials in advance and reduce costs without penalizing accuracy.

KEYWORDS

target population of environments, market segments, genotype x environment, envirotyping, supervised learning

1 Introduction

Among the core objectives of rice breeding programs is the release of cultivars with improved yield, nutritional capacity, resistance to pests/diseases, and climate-smart (Hickey et al., 2019). However, some aspects hold back the genetic gains of breeding programs. One is the high costs associated with phenotyping (Furbank and Tester, 2011; Araus and Cairns, 2014), which can cause uncertainty in the number and allocation of trials throughout the locations to increase further. Another is unclear approaches to align multi-environmental trials (MET) with the target population of environments (TPE), especially in a climate change scenario that constantly challenges the definition of a TPE (Cooper and Messina, 2021).

Design of TPEs generally considers historical data from the soil, climate, hydrological aspects, management strategies, and sometimes socioeconomic data from locations where the crops are frequently produced, that is, it represents a set of characteristics that constitute future growing seasons, years and environments (Crespo-Herrera et al., 2021; Cooper et al., 2023). To avoid the yield gaps between the expected yield potential in the TPE and the real on-farm yield that farmers achieve, it is important to characterize better and select the locations used for testing/selection and to understand how they are related (mega-environments) by a process called enviromic or envirotyping. Specifically, in this process, the environmental covariates are collected and processed in appropriate MET groupings, which are analyzed concerning the alignment with TPE and, thus, used to capture genotype reaction norms models (Cooper and Messina, 2021; Costa-Neto et al., 2023; Callister et al., 2024).

The concepts of environmental characterization, the genotype by environment interaction (GxE), and the target population of environments have been mentioned for many years in corn breeding in the USA, with this information presenting itself as a valuable resource for the decision-making process in breeding programs (Boer et al., 2007; Gaffney et al., 2015). However, little attention has been given to the American rice belt's environmental characterization and TPE delineation. Although the United States of America (USA) has a small rice production compared to Asian countries, the country is responsible for 5% of all rice exports in the world and has tripled its imports since 2001/02, showing a clear increase in crop demand (USDA, 2023).

The United States has four major rice-producing regions produced through the irrigated rice crop system: the Southern USA Rice Belt representing 85.6% of the USA Rice belt, with the Arkansas Grand Prairie, Mississippi Delta (Arkansas, Mississippi, Missouri, and Northeast Louisiana), and Golf Coast (Texas and Southwest Louisiana); and the Western USA Rice belt (14.4%), with only the Sacramento Valley California. These regions produce different types of rice classified by the United States, largely defined by grain market classes, with approximately 75% of the country producing long grain, 24% producing medium grain, and just 1% producing short grain. These last two types are produced partially by the state of Arkansas and mainly by the state of California (USDA, 2023).

Among all the breeding programs in the country, the Louisiana State University AgCenter is a centennial rice breeding program that utilizes a wide network of locations to conduct its METs across the US Rice belt. Because of its extensive field phenotypic evaluations, some questions were raised, such as: Is it necessary for so many locations? Does our MET match the TPE? Are we allocating our field trials properly? If we reduce the number of trials, we reduce the cost, but what happens with accuracy? In this context, several studies have indicated that the environmental covariates inclusion can enhance prediction accuracy (Moura-Bueno et al., 2021; Neyhart et al., 2022; Rogers and Holland, 2022; Montesinos-López et al., 2023). For instance, including environmental covariates for designing optimized training sets for genomic prediction can improve the response to selection per dollar invested by up to 145% compared to the model without environmental data (Gevartosky et al., 2023).

Therefore, we use historical data from the LSU Rice Breeding Program as a training set to address these questions and optimize the allocation of rice multi-environment trials (MET) in the USA Rice Belt via smart-climate prediction models based on historical weather and yield data, economic importance, artificial intelligence, and mixed model equations.

2 Materials and methods

2.1 Plant materials and trials

The experimental material consisted of 25 rice genotypes tested in two years, 2021 and 2022, and phenotyped for rice grain yield (kg ha⁻¹). There were 25 genotypes, 21 were of the long, and 4 were of the medium grain types. Louisiana State University, University of Arkansas, Horizon Ag, and Nutrien each provided five new genotypes. Also, five checks were included in the trials that were conducted at 19 different locations in the Mississippi Delta (Arkansas, Mississippi, Missouri and Louisiana) and Golf Coast (Texas and Southwest Louisiana). Since these were advanced trials, we had 25 genotypes tested in 2022, but 9 genotypes tested in 2021 (9 genotypes were common between 2021 and 2022.). The experimental design of these trials was a Randomized Complete Block Design, with 3 or more blocks. A figure showing how many times the genotypes were phenotyped per location and per year is available as [Supplementary Material \(Supplementary Figure S1\)](#). We define this main dataset as “LSU” for convenience.

2.2 Single trial analysis

We performed a two-stage analysis using linear mixed models to estimate grain yield using BLUEs (Best Linear Unbiased Estimators) for each individual trial (location × year), similar to the method used by Jarquín et al. (2014). The first model was calculated using the SpATS package in the R environment (version 4.3, <https://www.r-project.org/>):

$$y_{LSU_{ijmkn}} = \mu + G_i + R_j + L_m + C_n + \varepsilon_{ijmkn} \quad (1)$$

With $yLSU_{ijmkn}$ being the phenotypic values of grain yield and equal to an overall mean μ , plus a fixed effect from genotypes (G_i ; $i = 1, \dots, I$), a random effect from replicates (R_j ; $j = 1, \dots, J$), a random effect from rows (L_m ; $m = 1, \dots, M$), a random effect from columns (C_n ; $n = 1, \dots, N$) and a random term describing residuals (ε_{ijmkn} ; $k = 1, \dots, r_{ijmkn}$). The random terms were assumed to be independent and identically distributed, where $R_j \sim N(0, \sigma_r^2)$, $L_m \sim N(0, \sigma_l^2)$, $C_n \sim N(0, \sigma_c^2)$, $e \varepsilon_{ijmkn} \sim N(0, \sigma_e^2)$. We utilized an auxiliary function in the SpATS model for spatial correction, which models the spatial heterogeneity effect using a two-dimensional penalized tensor-product of B-spline basis functions. A table containing broad heritability per trial (combinations of years by locations) is provided as [Supplementary Material \(Supplementary Table S1\)](#).

2.3 Multi trial analysis

The second step of the analysis was to predict the genotypic values using BLUP and estimate heritability in a multi-trial analysis. This time we used *sommer* package to perform linear mixed model calculations in the R environment ([Covarrubias-Pazarán, 2016](#)):

$$yLSU_{xijk} = \mu + Y_x + G_i + E_j + GE_{ij} + \varepsilon_{xijk} \quad (2)$$

With $yLSU_{xijk}$ being the grain yield BLUES from model (1) and equal to an overall mean μ , plus a fixed effect from year (Y_x ; $i = 1, \dots, X$), a random effect from genotypes (G_i ; $i = 1, \dots, I$), a fixed effect from environment/location nested in year (E_j ; $i = 1, \dots, J$), a random effect from the interaction between genotype and environment and a random term describing residuals (ε_{xijk} ; $k = 1, \dots, r_{xij}$). The random terms were assumed to be independent and identically distributed, where $G_i \sim N(0, \sigma_g^2)$ and $\varepsilon_{xijk} \sim N(0, \sigma_e^2)$. As the design matrix of the genotype by environment interaction is the Hadamard product (\odot) of $[Z_g Z'_g]$ and $[Z_e Z'_e]$, and Z_g and Z_e are the incidence matrix of genotype and environment, $G E_{ij} \sim N(0, [Z_g Z'_g] \odot [Z_e Z'_e] \sigma_{ge}^2)$. The same model, but with all random effects, was used to assess the variance partitioning among the different variance components of year, location, genotype, and genotype-by-environment effects. The variance components for the random model are presented as [Supplementary Material \(Supplementary Table S2\)](#).

2.4 Environmental covariates

The “envRtype” package was used to obtain environmental data ([Costa-Neto et al., 2021](#)). The package uses data from NASA’s orbital sensors along with location, geographic coordinates, and time range data to extract environmental data related to the experimentation locations. After obtaining the data, we tuned the environmental covariates (EC) with the cardinal limits for temperature on the phenology development of rice ([Table 1](#)). The resulting centralized and scaled matrix had 114 covariates (19 location covariates x 6 phenological stages) for 19 experimentation sites.

TABLE 1 Rice phenological stages with corresponding abbreviations and day intervals.

Phenological stage	Abbreviation	Interval (days)
Emergency - Maximum Tillering	EM_MAX.TIL	0 - 44
Maximum Tillering - Panicle Initiation	MAX.TIL_PAN.INIT	45 - 59
Panicle Initiation - Pre Flowering	PAN.INIT_PRE.FLW	60 - 74
Pre Flowering - Flowering	PRE.FLW_FLW	75 - 89
Flowering - Post Flowering	FLW_POST.FLW	90 - 104
Post Flowering - Maturity	POST.FLW_MAT	105 - 148

The *SoilType* package was used to generate the soil covariates matrix ([Fritsche-Neto, 2023](#)). The package uses GPS coordinates to capture information about the soil of the locations closest to the experimentation site through the World Soil Database (WoSIS) ([Batjes et al., 2017](#)). Based on this database, the package also calculates several chemical and physical soil covariates. For the following steps, we joined the location and soil matrices into a single environment covariates matrix, which had 125 covariates for 18 locations, since it was not possible to generate environmental information for one of the 19 locations. This matrix underwent quality control using the *caret* package to ensure that only covariates with less than 95% correlation were maintained and to reduce collinearity between covariates ([Kuhn, 2008](#)). After this control, our scaled and centered W matrix remained with only 67 covariates.

2.5 Feature selection and clustering

We used the Recursive Feature Elimination (RFE) algorithm and the random forest learning method of the *caret* R package to select the most important predictors or covariates ([Kuhn, 2008](#)). The validation method was the repeated cross-validation with 5 folds and 5 replicates. Then, models using the W matrix composed with only the predictors selected in the RFE, we calculated the enviromic-based kernel for similarity among environments (Ω), or “environmental relationship matrix” ([Costa-Neto et al., 2021](#)):

$$\Omega = \frac{WW'}{\text{tr}(WW')/\text{nrow}(W)} \quad (3)$$

Where W is the environmental covariate matrix generated in the last steps, $\text{tr}()$ is a trace matrix and $\text{nrow}()$ is the number of rows. To group the experimental locations based on the features selected in the previous step, we clustered the locations using the *factoextra* package and k-means method ([Kassambara and Mundt, 2020](#)). Then, we decided how many clusters would be used to separate the locations with the help of the “Within Cluster of Squares” method.

2.6 Multi-environment trial optimization

To explore the benefits of multi environment trial optimization through environmental covariates, we estimated the BLUPs for each genotype in four different scenarios:

I - The Multi-Environment model

The Multi-Environment model (MET) is exactly the complete model (2) previously used in the multi-trial analysis. In this work, we did not focus on the stability and specific adaptation of the varieties. Therefore, we included the GxE effect in the model only to test heritability across scenarios, as this was our primary interest:

$$y_{LSU \cdot xijk} = \mu + Y_x + G_i + E_j + GE_{ij} + \varepsilon_{xijk} \quad (4)$$

II - The Multi-environment model with Environmental Covariates

The Multi-Environment model with Environmental Covariates (MET_EC) has the same effects as the previous model, but with the addition of the enviromic-based kernel (Ω) based on the W matrix previously described:

$$y_{LSU \cdot xijk} = \mu + Y_x + G_i + E_j + GE_{ij} + \varepsilon_{xijk} \quad (5)$$

Where $GE_{ij} \sim N(0, [Z_g Z'_g] \circ \Omega \sigma_{ge}^2)$.

III - The Within Cluster Multi-Environment model

The Within Cluster Multi-Environment model (WC_MET) is the same as the MET model, but the genotype BLUPs were calculated for each of the clusters (c) individually ($c = 1, \dots, C$), with C representing the five LSU clusters with different numbers of trials each one:

$$y_{LSU \cdot xijk}^{(c)} = \mu + Y_x + G_i + E_j^{(c)} + GE_{ij} + \varepsilon_{xijk} \quad (6)$$

IV - The Optimized Multi-Environment model

Finally, the Optimized Multi-Environment model (OP_MET) is the same as the MET model, but calculated individually for each subset of locations for each replicate. Where l is a subset of five randomly chosen locations from each of the five clusters ($l = 1, \dots, L$) and h is one of the ten replicates ($h = 1, \dots, H$) performed to avoid bias by choosing only one location from each cluster:

$$y_{LSU \cdot xijk}^{(l,h)} = \mu + Y_x + G_i + E_j^{(l,h)} + GE_{ij} + \varepsilon_{xijk} \quad (7)$$

2.7 Cluster effect on rice grain yield

We calculated clusters adjusted means based on rice grain yield performance of each location to observe the effect of environmental clusters on productivity based on the following model:

$$y_{LSU \cdot xijk} = \mu + C_j + G_i + GC_{ij} + \varepsilon_{ijk} \quad (8)$$

With $y_{LSU \cdot xijk}$ being the grain yield BLUEs from model (1) and equal to an overall mean μ , plus a fixed effect from cluster (C_j ; $j = 1, \dots, J$), a random effect from genotypes (G_i ; $i = 1, \dots, I$), a random effect from the interaction between genotype and cluster (GC_{ij}) and a random term describing

residuals (ε_{xijk} ; $k = 1, \dots, r_{xij}$). The random terms were assumed to be independent and identically distributed, where $G_i \sim N(0, \sigma_g^2)$, $C_j \sim N(0, [Z_g Z'_g] \circ [Z_c Z'_c] \sigma_{gc}^2)$ and $\varepsilon_{xijk} \sim N(0, \sigma_\varepsilon^2)$, and Z_g and Z_c are the incidence matrix of genotype and cluster.

2.8 Accuracy per unit of dollar invested

To obtain accuracy for each of the four scenarios described above, we calculated broad-sense heritability (H^2) using the Cullis method (Cullis et al., 2006):

$$H^2 = 1 - \frac{v_{\Delta}^{-BLUP}}{2 \cdot \sigma_g^2} \quad (9)$$

Where the term v_{Δ}^{-BLUP} stands for the average standard error of the genotypic BLUPs. The next step was to generate the heritability value per dollar unit invested in a similar way that [Gevartosky et al. \(2023\)](#) calculated. To achieve this, we calculated the total cost for each of the four scenarios based on the rates paid by LSU. For MET, MET_EC, and WC_MET: the calculation is the product of 25 genotypes x 3 replicates x \$25.00/plot phenotyping x 18 locations x 2 years = \$67,500.00. Since all LSU locations were grouped into 5 clusters, the cost for the OPT_MET scenario was calculated as follows: 25 genotypes x 3 replicates x \$25/plot phenotyping x 5 locations x 2 years = \$18,750.00.

2.9 Cost reduction simulation in the optimized scenario

We demonstrated cost reduction through the optimization of LSU's multi-environment trials in two distinct ways: the reduction for LSU's advanced trials (described in item 2.1) and the reduction for the entire LSU experimental network, which includes not only the advanced trials but also other breeding stages in all LSU trial locations for ten years. A table containing the number of trials conducted by LSU for these 18 locations over 10 years is presented as [Supplementary Material \(Supplementary Table S3\)](#). We considered this second option because the number of trials varies across the 18 locations, with some locations tested in less years, and optimizing these locations would impact the entire LSU network. Using this list from the complete LSU experimentation network, we also calculated the percentage of trials per LSU cluster.

For the first approach, where only one trial per combination of year and location is conducted, the reduction in trials is calculated by dividing the number of optimized locations by the total number of locations (5/18), resulting in a virtual reduction of 27.7%. To simulate a cost reduction for the entire LSU experimental network, we calculated an average reduction across ten iterations. This is a simulated value, as each location has a different number of trials, and in a real-world scenario trials would also be redistributed according to the Target Populations of Environments. In each iteration of OPT_MET, the number of trials from the five selected locations (one location per cluster) is summed and then compared to the total number of clusters to determine the percentage reduction.

2.10 Delimitation and economic characterization of the target population of environments

Through the previous analyses, we were able to predict the most important environmental covariates for rice grain yield by using the LSU MET as the training set. Based on that, we used those covariates to perform a K-means cluster analysis to delimit the target population of environments (TPE) for the Southern USA Rice Belt, and just for Louisiana. The whole USA rice belt is composed of 80 counties from seven States and represents 98% of USA rice production (Farm Service Agency - USDA), with 71 counties representing the Southern USA Rice Belt (85.6% of USA rice production). Finally, based on the economic importance (rice production of each County), we estimated the mega-environment (TPE) importance and, consequently, the proportion of trials that should be allocated in that market segment. To assess how much the complete LSU network covers the American TPEs (clusters), we assigned each LSU experimentation site to one of the four American clusters. This assignment was based on the minimum Euclidean distance between the scaled environmental profiles of the LSU sites and the centroids of the four American clusters. To demonstrate the environmental characterization of the regions within the Southern US Rice Belt, we used a PCA biplot based on the W matrix, with locations separated by clusters. This plot includes a standard PCA analysis, with the addition of loading plots to show the influence of each covariate on the principal components.

2.11 Advantages of the environmental covariate matrix

We compared the environmental covariates matrix (W) with the yield-based GxE matrix (W_{yield}), as it is more conventional for studying the relationship between environments and for analyzing the stability and specific adaptation of varieties (Yan et al., 2023). To accomplish this, we performed the same analyses as we did for the W matrix: calculated the yield-based environmental relationship matrix (Ω_{yield}), clustered the locations using this matrix, calculated the percentage of trials for each cluster, and determined the clusters' adjusted means.

3 Results

3.1 Defining the Southern USA rice mega-environments

The W matrix shows the profile of LSU locations based on the environmental covariates selected by a supervised artificial intelligence algorithm (Figure 1A). Specifically, using the Recursive Feature Elimination method with the Random Forest machine learning algorithm, these eight covariates alone account for 58% of the total variation in rice yield across the LSU experimental network. The eight most critical covariates were

both temperature and soil-related. Among the temperature-related variables, there was *The dew/frost point temperature at 2 meters above the surface of the earth* (T2MDEW) in the phenological stage between Flowering and Post Flowering (FLW_POST.FLW); the T2MDEW in the phenological stage between Pre Flowering and Flowering (PRE.FLW_FLW); the T2MDEW in the phenological stage between Panicle Initiation and Pre Flowering (PAN.INIT_PRE.FLW); *The minimum hourly air (dry bulb) temperature at 2 meters above the surface of the earth in the period of interest* (T2M_MIN) in the stage between Maximum Tillering and Panicle Initiation (MAX.TIL_PAN.INIT); the *Growing Degree-Days* (GDD) at the PRE.FLW_FLW stage; And *The minimum and maximum hourly air (dry bulb) temperature range at 2 meters above the earth's surface in the period of interest* (T2M_RANGE) in MAX.TIL_PAN.INIT. Regarding the soil-related covariables, the chemical soil feature *Calcium carbonate total equivalent in g/kg* (TCEQ); and the physical soil feature *Total Silt in g/100g* (SILT). Therefore, the only phenological stages that do not produce environmental variability in rice yield are the Emergency to Maximum Tillering stage (EM_MAX.TIL), between 0 to 44 days, and Post Flowering to Maturity (POST.FLW_MAT), between 105 to 148 days.

From the Wmatrix, we performed the clustering and thus identified the mega-environments present in the LSU rice breeding program (Figure 1C). This k-means clustering analysis revealed five distinct mega-environments among the eighteen experimentation sites, with the first two Principal Components (PCs) explaining 79.3% of the variation (PC1 accounting for 56.3% and PC2 for 23%). The map at the top displays the non-optimized experimentation locations, while the map optimized by the k-means method appears at the bottom (Figure 1D). With this optimization, the number of locations was reduced by 27.7%, covering only the states of Louisiana and Texas. Furthermore, we display the percentage of trials per cluster (Figure 1E), allowing us to observe the distribution of trials across mega-environments. The mega-environments represent from as little as 0.7% of total trials (Cluster 4) to as much as 81.5% of total trials (Cluster 2). Finally, we present the adjusted means for each cluster (Figure 1F), with Clusters 4 and 5 having the highest averages (comprising 0.7% and 10.1% of trials, respectively), with approximately 10,000 and 9,500 kg per hectare in grain yield. Included within these two clusters are the locations of Rice Research Station - Late, Rice Research Station South, Mamou, Morata, Bay City, Nutrien Rice Breeding Station (El Campo), and Wintermann Rice Research Station.

3.2 Multi-environment trials optimization

The heritability graph reveals distinct outcomes for the four scenarios analyzed (Figure 2). The MET scenario has a heritability of 0.965; the MET_EC scenario has a heritability of 0.893; the OPT_MET scenario has a mean heritability of 0.892 and standard deviation of 0.015; and the WC_MET scenario has a mean of 0.840 and standard deviation of 0.067. A table containing the heritability of all scenarios, as well as the variance components of the random effects, is provided as [Supplementary Material \(Supplementary](#)

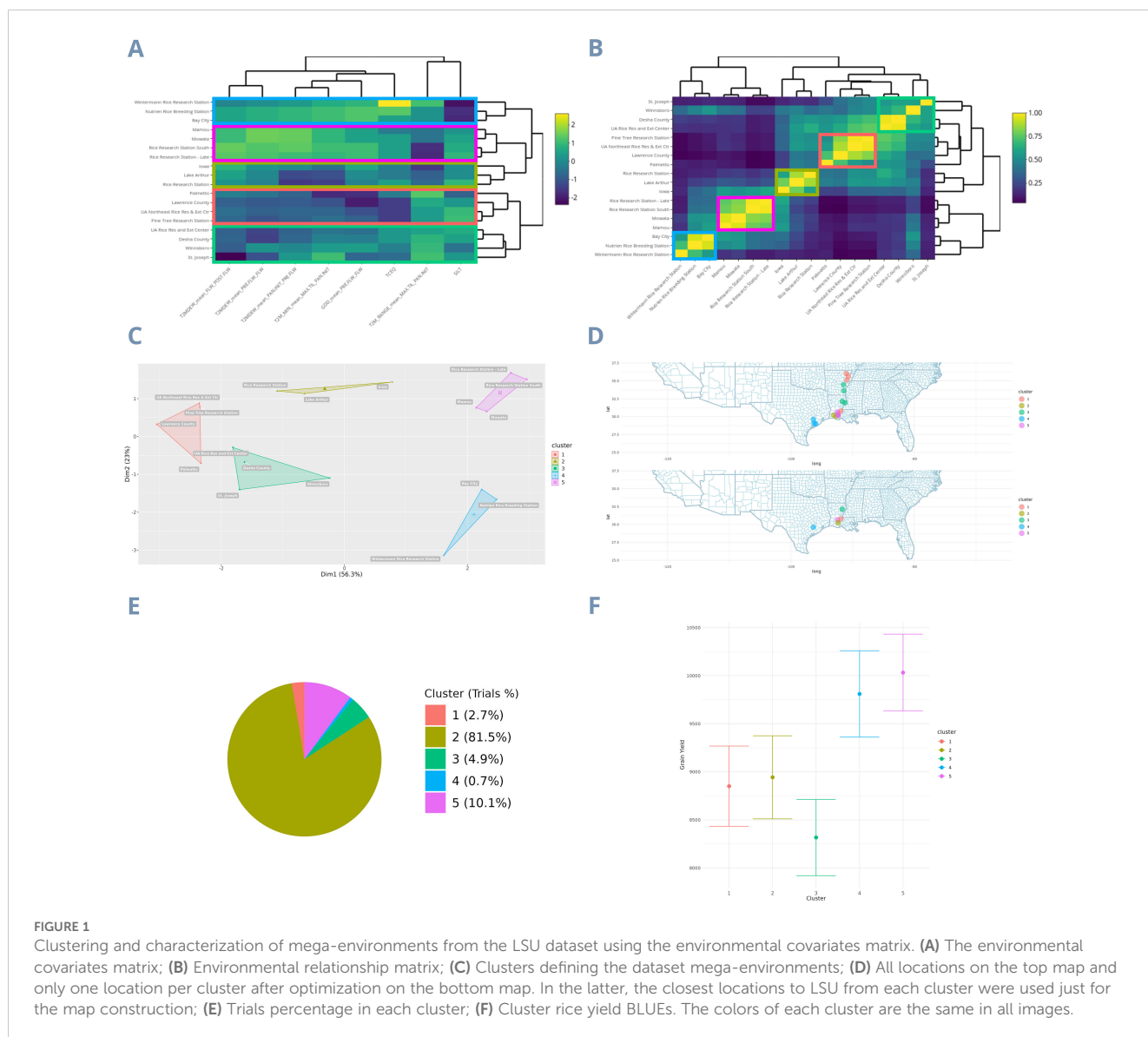


Table S4). When we divided the heritability by the total cost associated with each scenario, the ratio between heritability and cost in the OPT_MET scenario was at least four times higher than that of any other scenario, underscoring its significant economic value. Moreover, from the 829 trials conducted by Louisiana State University in ten years (whole experimentation network), only an average of 187.6 trials remained in the optimized scenario, representing just 22.6% of the original trial number, with a consequent virtual reduction of 77.4% in the number of trials. When considering only the advanced trials, a virtual 27.7% reduction in locations is achieved.

3.3 Delimitation of the USA rice belt target population of environments

The subsequent analyses delineate the USA TPE using the eight predicted covariates that best explain rice yield. We performed the

clustering and distribution of trials across all 80 counties, representing 98% of US rice production in Arkansas, California, Florida, Louisiana, Mississippi, Missouri, and Texas. The Ω matrix highlights a distinct group unrelated to the other counties and comprises those in California (Figure 3A). Consequently, the clustering of the U.S. dataset revealed significant variability between the California cluster and the rest of the American counties, resulting in the latter being considered a single cluster. This created a new dataset that excluded the 9 California counties and utilized 85.6% of the remaining USA production (Southern USA Rice Belt). (Figure 3B). The findings from this dataset indicated that the revised W, which includes the 71 remaining counties, is more homogeneous and is divided into four mega-environments (Figure 4A). The economic importance (percentage of production) of the Southern US Rice belt TPEs (clusters) 1 to 4 is 15.3%, 27.6%, 44.4%, and 12.7%, respectively. Meanwhile, the trial percentage of the whole LSU experimentation network in each of these clusters is 38.9%, 50.0%, 5.6%, and 5.6%, respectively. (Figures 4B, C). Lastly, as Supplementary Material, a new

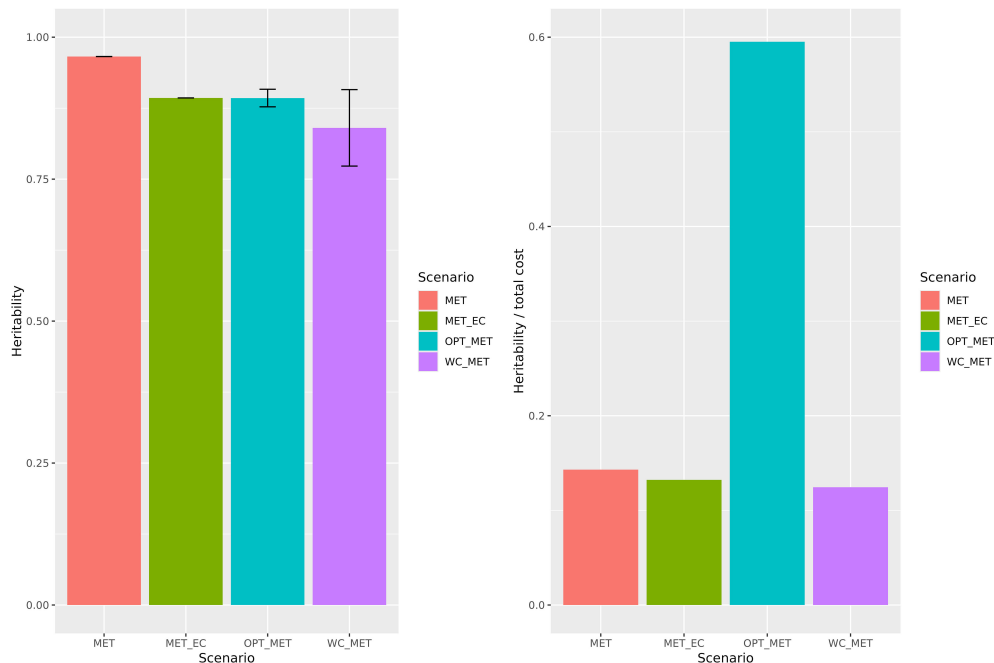


FIGURE 2

Broad-sense heritability on the left and heritability per unit of dollar invested on the right in the different scenarios for multi-environment trials optimization testing. MET - The multi-environment model; MET_EC - The multi-environment model with environmental covariates; WC_MET - The within cluster multi-environment model; OPT_MET - The optimized multi-environment model.

dataset comprising 19 Louisiana parishes was divided into two mega-environment, representing Louisiana's northern and southern regions, with 27.7% and 72.3% of trials per cluster, respectively (Supplementary Figure S2).

3.4 Environmental characterization of clusters

LSU's environments of evaluation were divided into five clusters, and these clusters are characterized based on the evaluation of a PCA biplot analysis (Figure 5A):

- Cluster 1 is related to lower covariates values, such as minimum temperatures (MAX.TIL_PAN.INIT), dew

point temperatures (from PAN.INIT to POST.FLW), and TCEQ.

- Cluster 2 exhibits higher silt contents and lower minimum temperatures (MAX.TIL_PAN.INIT), dew point temperatures (T2MDEW), and TCEQ.
- Cluster 3 is characterized by higher temperature ranges (MAX.TIL_PAN.INIT), while lower values for all the other covariates.
- Cluster 4 has higher minimum temperatures (MAX.TIL_PAN.INIT) and TCEQ, lower dew/frost temperatures (PRE.FLW to POST.FLW), and SILT content.
- Cluster 5 exhibits higher TCEQ content and minimum temperature (MAX.TIL_PAN.INIT), while SILT and dew/frost point temperatures have lower values.

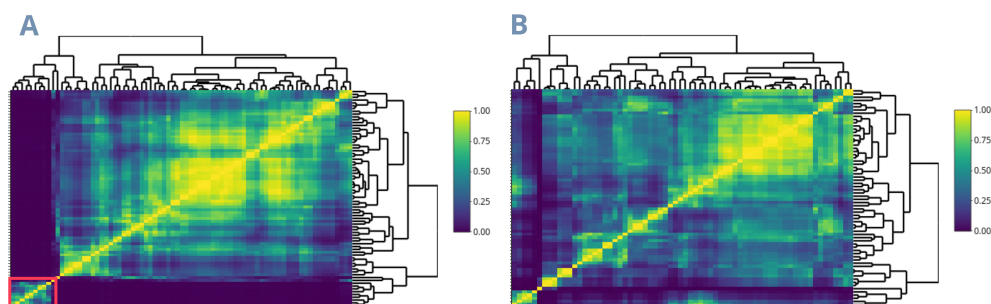
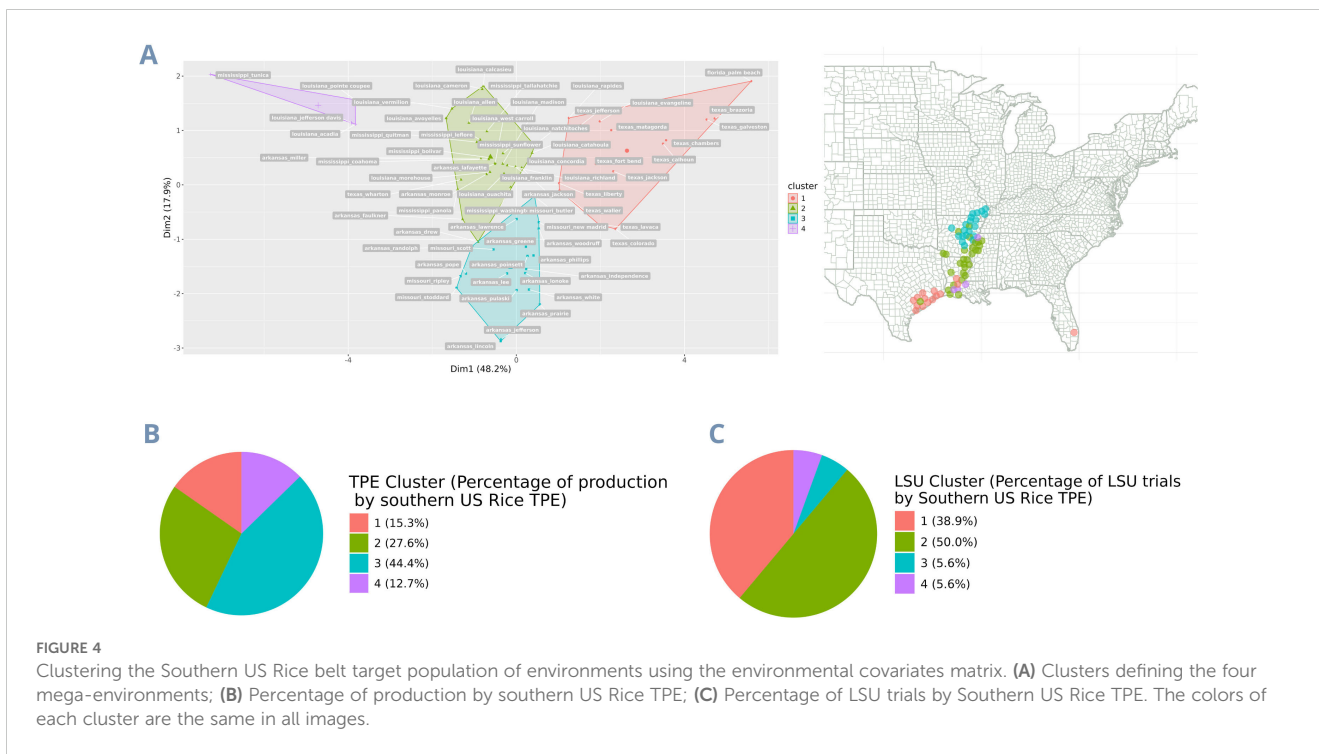


FIGURE 3

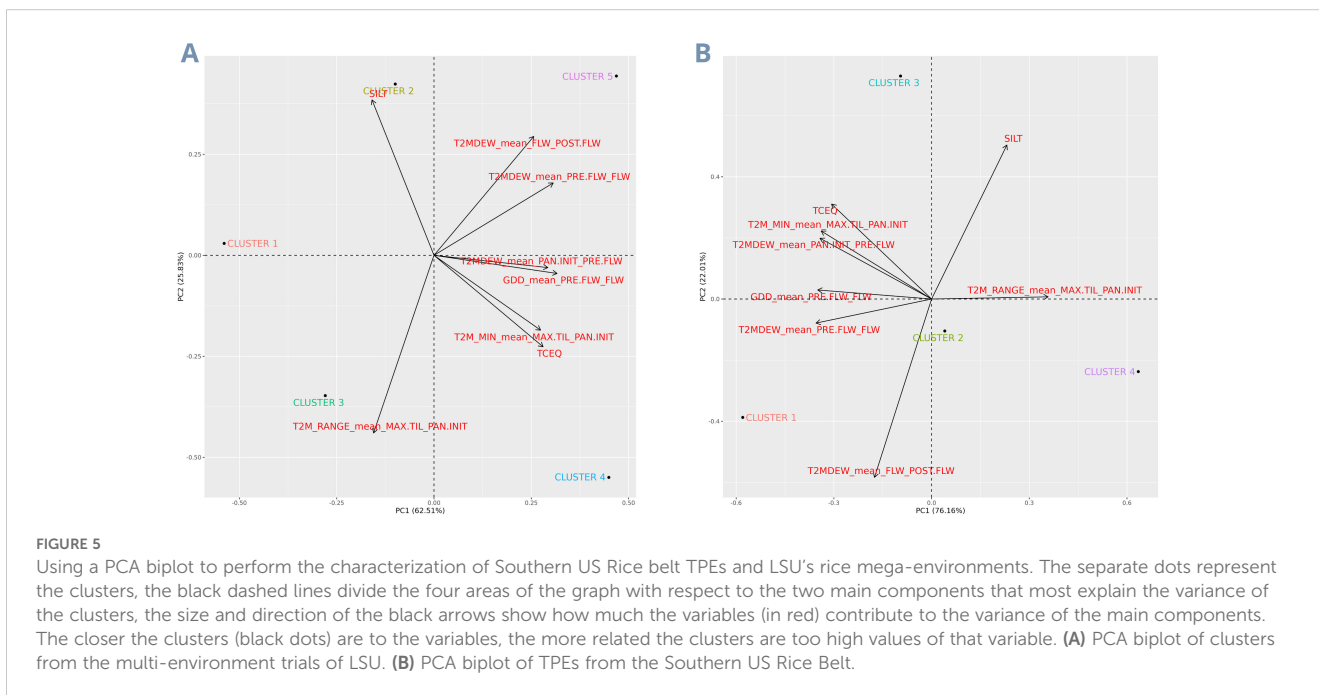
Environmental relationship matrices based on environmental covariates for the entire US Rice Belt dataset (A) and for all states in the US Rice Belt except California (B). The red square represents the sites in California, which are environmentally different from the others.



Additionally, we provide the environmental characterization of the US Rice Belt TPEs (**Figure 5B**):

- Cluster 1 exhibits higher dew point temperatures from PRE.FLW to POST.FLW and low levels of SILT in the soil.
- Cluster 2 is characterized by having intermediate environmental conditions compared to the other three clusters.

- Cluster 3 is characterized by high levels of TCEQ in the soil and high minimum temperatures at MAX.TIL_PAN.INIT, as well as high dew point temperatures at PAN.INIT_PRE.FLW.
- Cluster 4 is defined by high temperature ranges at MAX.TIL_PAN.INIT.



3.5 Exploring the environmental covariates matrix

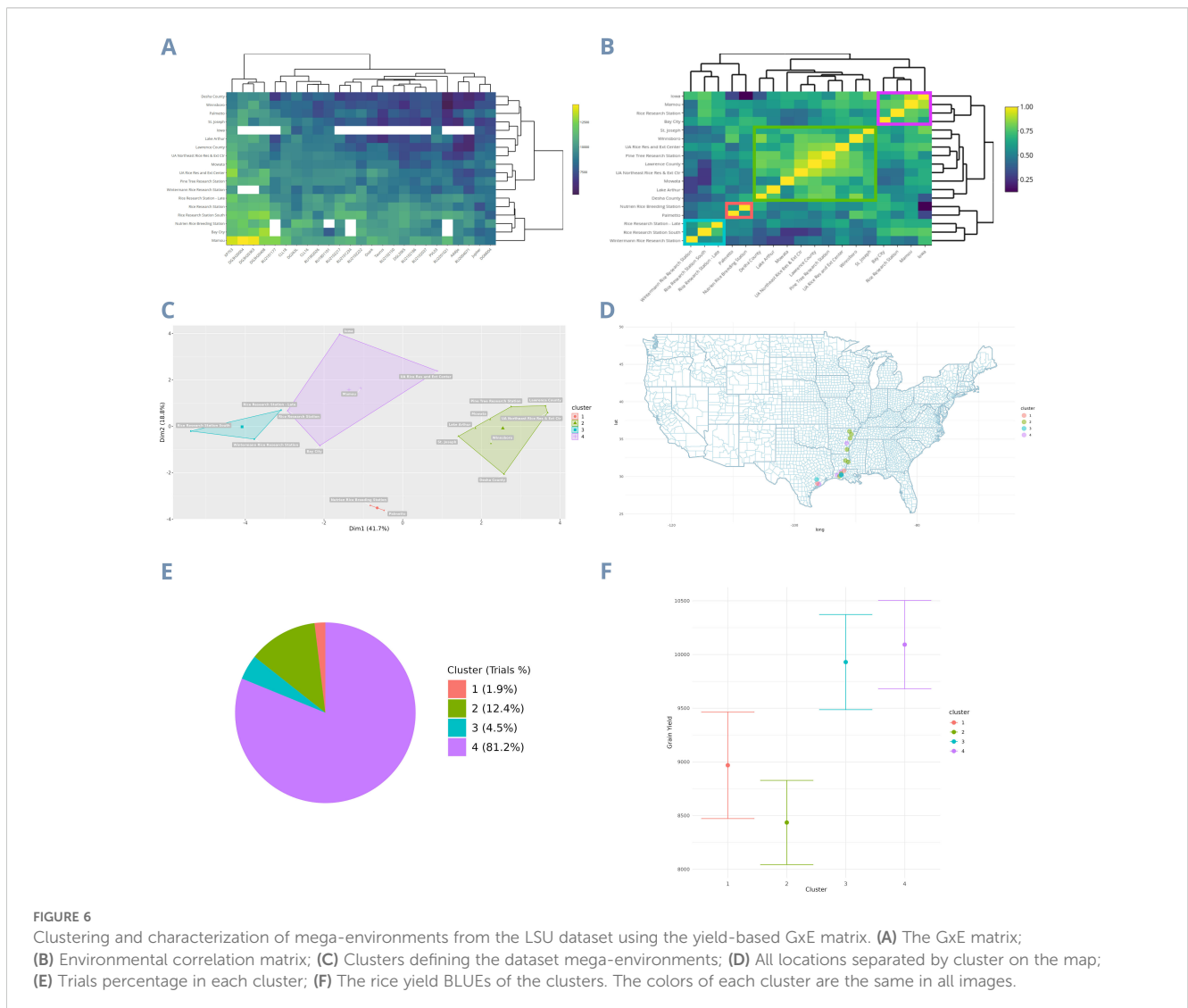
To determine how much the environmental relationship matrix based on covariates could independently explain the yield-based environmental relationship matrix, we calculated a Pearson correlation between the two. The Ω matrix (Figure 1B) has a 52.59% correlation with the Ω_{yield} matrix (Figure 6B), meaning that the eight environmental covariates alone explain 52.59% of the traditional GxE matrix. Additionally, the W_{yield} matrix, unlike the W , represents a yield gradient (Figure 6A). Consequently, the W_{yield} does not exhibit clearly separated kinship blocks (Figure 6B), and the four clusters appear closer to each other, while the variation within clusters increases. Furthermore, the first two principal components accounted for a smaller proportion of the total variation, with the first two PCs explaining 60.5% of the total variation (PC1 at 41.7% and PC2 at 18.8%) (Figure 6C). Regarding the percentage of trials per cluster, clusters 4 and 2 have the highest trial allocations, with 81.2% and 12.4% respectively (Figure 6E). The clusters with the highest yield adjusted means were 4 and 3, which

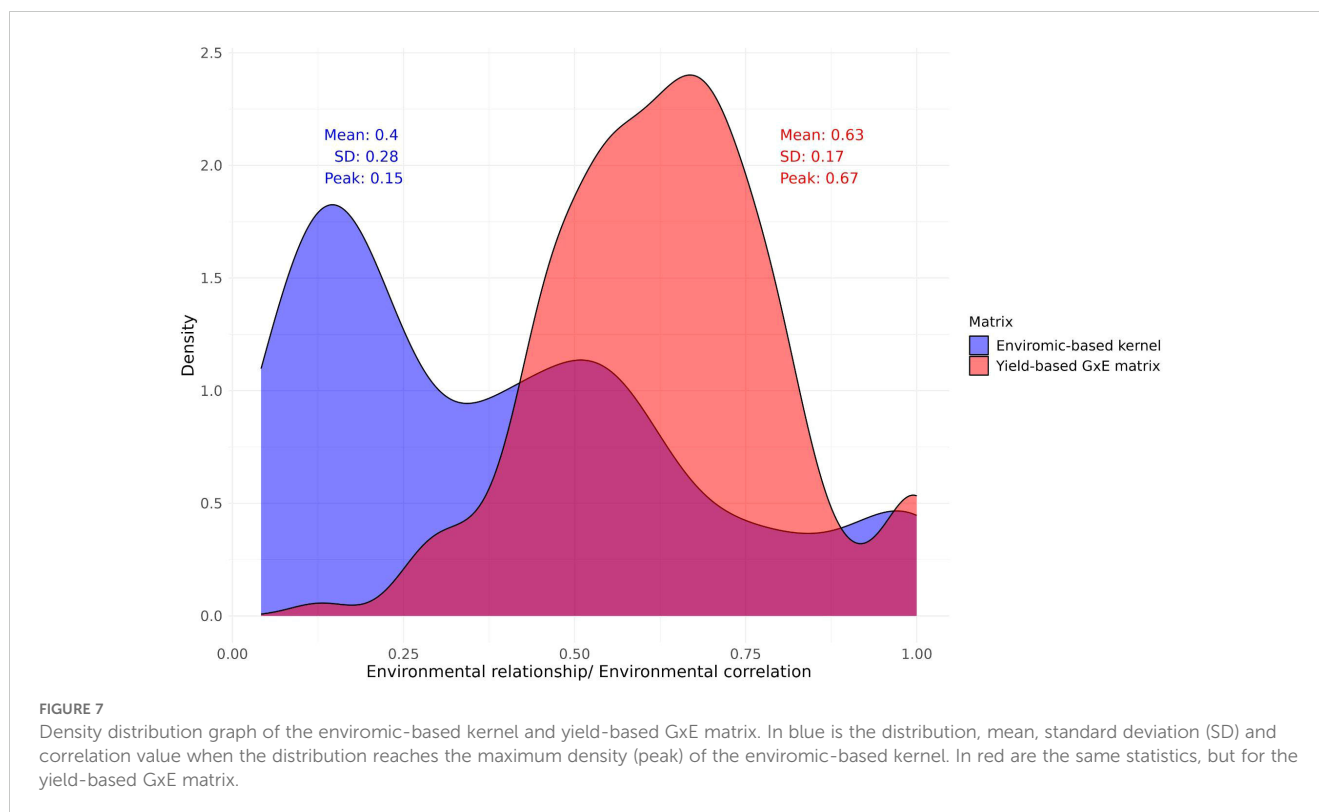
include the locations Iowa, Mamou, UA Rice Research and Extension Center, Rice Research Station - Late, Bay City, Rice Research Station, Wintermann Rice Research Station, and Rice Research Station - South (Figure 6F).

To further show the difference between Ω and Ω_{yield} , we plotted a density distribution graph (Figure 7). The Ω distribution has a high density where environments have lower relationship (Peak in 0.15) and a higher standard deviation (0.28), highlighting the environmental heterogeneity contained in the MET. On the other hand, the Ω_{yield} has a high density of more correlated environments (Peak in 0.67) and a lower standard deviation (0.17), masking the real MET environmental heterogeneity.

4 Discussion

MET are necessary to study the phenotypic plasticity of the genotypes when subjected to environmental fluctuations under different crops' phenological stages. The change in the genotypes rank can result from more static and predictable environmental





covariates, such as SILT and TCEQ as observed in our study, but also from more unpredictable ones, such as temperature in our study (Crespo-Herrera et al., 2021). The evaluation of quantitative traits that are highly prone to demonstrate phenotypic plasticity (crossover GxE interactions) must be done so that selection sites effectively meet TPE needs. The assumption is that the locations chosen to allocate the trials will efficiently represent the sets of environmental covariates found in the TPE to make it possible to outline a strategy on how to handle GxE in the selection of superior genotypes in different stages of a breeding program (Cooper and Delacy, 1994). In this context, disregarding the GxE interaction can lead to a reduced selection response. This is particularly true when programs conduct early generation selection, as they test many lines in few environments (Cooper et al., 1995). Therefore, information on the characterization and design of TPEs can benefit breeding programs by guiding the allocation of trials and considering the alignment of selection environments and TPEs.

Our study identified eight covariates that explain 58% of the variation in rice yield across the Southern USA Rice Belt. Among these, temperature was the environmental covariate explaining the largest portion of variation. Although it is widely known that rice flowering is regulated by temperature, the underlying regulatory mechanisms are not yet fully understood. Generally, higher minimum temperatures accelerate crop flowering, which can reduce biomass and grain yield. Conversely, late flowering may lead to increased biomass accumulation but can simultaneously reduce grain filling (Srikanth and Schmid, 2011). Additionally, high temperatures can deteriorate rice quality due to imbalances between protein content and starch in the grains (Liu et al., 2021). On the other hand, cold stress can reduce rice yield during any phenological

stage (da Cruz et al., 2013). In our study, minimum temperature was particularly important between maximum tillering and panicle initiation, with higher minimum temperatures associated with clusters of higher yields. Moreover, some studies highlight the importance of assessing air relative humidity when studying the response of rice to temperature (Stuerz and Asch, 2019). Accordingly, the dew/frost temperature, which depends on both air humidity and temperature, was a key covariate in explaining the different selection environments in our research. Besides the environmental factors mentioned above, two soil covariates, namely the amounts of calcium and silt, also significantly contributed to the yield variation observed in the trials.

Using these covariates, we grouped the locations into five clusters with environmental similarities. The productivity of these groups ranged from 8,500 to 10,000 kg per hectare for clusters 3, 1, 2, 4, and 5, respectively. The two clusters with the highest average productivity were associated with high minimum temperatures, dew point temperatures, and high TCEQ contents in the soil. However, the TPE with the greatest economic importance is cluster 3 (Figure 4B), comprising 44.4% of the entire rice production in the Southern USA Rice belt. This cluster contains locations primarily in Arkansas, with a few in Missouri (in blue). Based on the classifications through the Euclidean distances of the LSU experimentation sites with the centroids of the TPE clusters, it is noticeable that only 5.6% of the trials from LSU's complete network represent this cluster (Figure 4C). The results of the delimitation of TPEs demonstrate how trial allocation could be optimized for more efficient resource utilization. Additionally, the results highlight the potential for improvement in US varieties, as environments with greater economic importance (Cluster 1 and 3 in

LSU) (Figure 1C) have lower yield and less favorable environmental covariates (Figure 5A), such as lower minimum temperatures and dew/frost points. Furthermore, when covariates are deemed highly significant for a trait to the extent that they become a long-term breeding target, this would justify the establishment of fine-grained research facilities for breeding programs (Cooper and Messina, 2021; Crespo-Herrera et al., 2021). There's the possibility to conduct coarse-grained phenotyping, such as environments with and without a specific stress, or fine-grained phenotyping, encompassing a broad range of this environmental continuum and allowing for detailed study of genotype reaction-norm models (Cooper and Messina, 2021).

Furthermore, our study shows, in practice, other ways to use enviroomics to perform a good allocation of resources. We tested both the heritability increase with the addition of covariates in MET joint analysis and the optimization of MET by using a reduced and efficient number of sites. Although the OPT_MET scenario did not produce the highest heritability among all scenarios, it was possible to maintain a high level per unit of dollar invested. Indeed, this value was nearly three times that of the other scenarios (Figure 2). Testing numerous lines in many environments for years can lead to a good genotype performance recommendation. However, this strategy assumes that the same genotypes will be used in the future and that there is no budget limit, which is inaccurate for a breeding program. In this sense, we reduced the number of tested locations in the advanced trials by 27.7%, while maintaining high accuracy (0.892). This is also advantageous because, although we reduce the number of tested locations, we still assess the genotypes in all the mega-environments that were previously being tested, but with greater cost-effectiveness.

Reducing the number of tested locations will lead to lower breeding costs, particularly in one of the most expensive stages of the process, phenotyping. As LSU pays \$25 per phenotyped plot, reducing 27.7% of the locations means we would save \$48,750.00 [\$67,500.00 (MET) - \$18,750.00 (OPT_MET)]. This amount could be reallocated to an earlier stage of the breeding program, where more lines are typically tested in fewer environments. For instance, if we consider reallocating this amount to a stage where lines are tested in just one environment and one year, we could phenotype an additional 1,950 plots in phenotypic selection (\$48,750.00/\$25 per plot), which would allow us to phenotype 650 new genotypes (1,950 plots/3 replicates). If it is not necessary to phenotype these new lines in a program that implements genomic selection, this amount could instead be reallocated to genotype an additional 8,125 lines (\$48,750.00/\$6 per line for genotyping). If the number of tested genotypes increases while the number of selected genotypes remains the same, the selection intensity significantly increases, leading to greater genetic gain per cycle. Considering the entire LSU network, the reduction in trials could reach approximately 77.4%, this resource reallocation scenario would be even more advantageous.

In recent years, breeders have shown promising results when considering environmental covariates in data modeling. Gevartosky et al. (2023) improved the response to selection by 145% when they considered environmental covariates in the design of optimized training sets for genomic prediction. In the same way, Costa-Neto et al. (2023) achieved a G×E variance decrease from 22% to 15%

when environmental covariates were considered, showing a more effective G×E effect capture. Conversely, the heritability of the MET_EC scenario (0.893) was lower than that of the MET scenario (0.965), even when the Ω matrix was used in the first scenario. We believe this may have occurred because the main environmental effect already provided sufficient information to produce the highest heritability among all scenarios. This could be due to the minimal variation between locations within clusters compared to the significant variation between clusters (Figure 1C). Furthermore, when we included the Ω matrix for the genotype-by-environment interaction effect (MET_EC), we increased the G×E variance component and decreased the G variance component, slightly decreasing the heritability.

Spindel and McCouch (2016) showed that the more correlated the environments and genotypes are in the modeling, the higher the accuracy. Therefore, it was expected that the WC_MET scenario would perform similarly to the other scenarios, but its average was the lowest (0.840 ± 0.067). However, WC_MET also had the highest standard deviation, with the lowest heritability being from cluster 4 (0.757). This cluster likely has more environmental heterogeneity than the others, where there is a location that is environmentally distant from the other two cluster locations (Wintermann Rice Research Station). One option to improve the accuracy of this mega-environment would be to split it into one or more clusters. However, this would increase the number of tested locations to 6, leading to a significant rise in costs. Such decisions need to be made by each breeding program according to their respective budgets, as a program might choose to slightly compromise accuracy to reduce the number of clusters and thereby lower costs at a specific breeding stage within their programs.

Besides proving that environmental covariates can help better trial allocation while maintaining accuracy, we wanted to highlight the benefits of delineating mega-environments by comparing W and W_{yield} matrices. The traditional way to represent G×E and the correlation between environments is through the use of the W_{yield} matrix (Cooper and Delacy, 1994; Senguttuvel et al., 2021; Yan et al., 2023), as the Additive Main Effects and Multiplicative Interaction (AMMI) model (Gauch et al., 1997). Using this methodology, it is possible to discover which environment produces the highest yield and which variety is better for each environment, as long as they are tested in those same environments. However, when the W matrix is used, it is possible to know which genotypes perform well for each environmental covariate or set of covariates due to environmental stratification. Just having information from an entire environment means the breeder cannot expand this information to new environments. The advantage of this is that environmental information is broad and free (Costa-Neto et al., 2021), while the traditional G×E matrix is highly dependent on the genotypes and environment combinations used. Therefore, with historical weather data and genotypes reaction norms, it is possible to recommend the best and most stable varieties for each target region and even work to discover potential new producing regions (Cooper and Messina, 2021; Costa-Neto et al., 2023; Araújo et al., 2024; Callister et al., 2024).

As climate varies significantly from year to year, a collection of environments cannot be deemed a TPE based solely on one or a few

years of data; delimitation must be conducted based on repeatable GxE patterns (Singh et al., 2006). This fact poses further challenges to establishing TPEs using yield-based environment relationship matrices. In the case of W matrix map (Figure 1D), the clusters are regionally separated. In contrast, for the W_{yield} , the clusters are mixed throughout the map (Figure 6D), suggesting a confounding effect in the last matrix and consequently, a possible change of genotype ranks between environments. The fact that the W_{yield} matrix PCs explain less variation than the W matrix PCs, shows further that more confounding effects were included in the first methodology. It is possible to predict this by the W_{yield} matrix structure, a yield gradient with more homogeneous correlation between environments (Figures 6B, 7). Consequently, the clusters are closer to each other (Figure 6C), while the variation within the clusters is bigger.

5 Conclusion

Conclusively, our study enabled us to virtually reduce 27.7% of the trials (locations) while maintaining almost the same accuracy. Furthermore, we demonstrated how this trial reallocation will allow for better utilization of our resources, as we could better represent all TPEs within the USA Rice belt according to their economic importance. Additionally, we identified which environmental covariates have the greatest impact on rice productivity in the considered TPEs and that they explain 58% of all variation in rice yield in the USA. With this information, it is possible to establish fine-grained phenotyping and expand production to potential new areas. These findings can be invaluable information in assisting rice breeding efforts in the USA and aiding breeders in optimizing trial allocation.

Data availability statement

The codes used for the analyses described in this manuscript are publicly available on GitHub: [GitHub - MET_Optimization] (https://github.com/MelinaPrado/MET_Optimization.git). This repository includes the project structure, scripts, and data necessary to reproduce the presented results.

Author contributions

MP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. AF: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Writing – review & editing. KG: Funding acquisition, Methodology, Project administration, Resources, Writing – review & editing. RF-N: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration,

Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Louisiana Rice Research Board. Improving Rice Genetics Gains via Prediction-based Models. 2023. GDM Seeds. Incorporation of new technologies into breeding pipelines.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1458701/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Number of phenotypes evaluated per year and location for each of the 25 genotypes included in LSU's advanced trials.

SUPPLEMENTARY FIGURE 2

Clustering and characterization of Louisiana TPEs using the environmental covariates matrix. (A) The environmental covariates matrix; (B) Environmental relationship matrix; (C) Clusters defining the dataset mega-environments; (D) All locations separated by cluster on the map; (E) Trials percentage in each cluster. The colors of each cluster are the same in all images.

SUPPLEMENTARY TABLE 1

Broad heritability per trial (combination of year and location).

SUPPLEMENTARY TABLE 2

Variance components partitioning among year, location, genotype, and genotype-by-environment effects using a random model.

SUPPLEMENTARY TABLE 3

Number of trials per experimental site at LSU over ten years at various stages of the breeding program.

SUPPLEMENTARY TABLE 4

A table containing the heritability of all scenarios (MET, MET_EC, WC_MET and OPT_MET), as well as the variance components of the random effects.

References

- Araújo, M. S., Chaves, S. F. S., Dias, L. A. S., Ferreira, F. M., Pereira, G. R., Bezerra, A. R. G., et al. (2024). GIS-FA: An approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting. *TAG. Theor. Appl. Genet. Theoretische Und Angewandte Genetik* 137, 80. doi: 10.1007/s00122-024-04579-z
- Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008
- Batjes, N. H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., and Mendes de Jesus, J. (2017). WoSIS: Providing standardised soil profile data for the world. *Earth Sys. Sci. Data* 9, 1–14. doi: 10.5194/essd-9-1-2017
- Boer, M. P., Wright, D., Feng, L., Podlich, D. W., Luo, L., Cooper, M., et al. (2007). A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics* 177, 1801–1813. doi: 10.1534/genetics.107.071068
- Callister, A. N., Costa-Neto, G., Bradshaw, B. P., Elms, S., Crossa, J., and Brawner, J. T. (2024). Enviromic prediction enables the characterization and mapping of Eucalyptus globulus Labill breeding zones. *Tree Genet. Genomes* 20, 3. doi: 10.1007/s11295-023-01636-4
- Cooper, M., and Delacy, I. H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *TAG. Theor. Appl. Genet. Theoretische Und Angewandte Genetik* 88, 561–572. doi: 10.1007/BF01240919
- Cooper, M., and Messina, C. D. (2021). Can we harness “Enviromics” to accelerate crop improvement by integrating breeding and agronomy? *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.735143
- Cooper, M., Powell, O., Ghosh, C., Tang, T., and Messina, C. (2023). Extending the breeder’s equation to take aim at the target population of environments. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1129591
- Cooper, M., Woodruff, D. R., Eisemann, R. L., Brennan, P. S., and DeLacy, I. H. (1995). A selection strategy to accommodate genotype-by-environment interaction for grain yield of wheat: Managed-environments for selection among genotypes. *Theor. Appl. Genet.* 90, 492–502. doi: 10.1007/BF00221995
- Costa-Neto, G., Galli, G., Carvalho, H. F., Crossa, J., and Fritsche-Neto, R. (2021). EnvRtype: A software to interplay enviromics and quantitative genomics in agriculture. *G3 Genes[Genomes]Genetics* 11, jkab040. doi: 10.1093/g3journal/jkab040
- Costa-Neto, G., Matta, D. H., Fernandes, I. K., Stone, L. F., and Heinemann, A. B. (2023). Environmental clusters defining breeding zones for tropical irrigated rice in Brazil. *Agron. J.* 1–25, 931–955. doi: 10.1002/agi2.21481
- Covarrubias-Pazarán, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One* 11, e0156744. doi: 10.1371/journal.pone.0156744
- Crespo-Herrera, L. A., Crossa, J., Huerta-Espino, J., Mondal, S., Velu, G., Juliana, P., et al. (2021). Target population of environments for wheat breeding in India: definition, prediction and genetic gains. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.638520
- Cullis, B. R., Smith, A. B., and Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *J. Agricult. Biol. Environ. Stat* 11, 381–393. doi: 10.1198/108571106X154443
- da Cruz, R. P., Sperotto, R. A., Cargnelutti, D., Adamski, J. M., de Freitas Terra, T., and Fett, J. P. (2013). Avoiding damage and achieving cold tolerance in rice plants. *Food Energy Secur.* 2, 96–119. doi: 10.1002/fes3.25
- Fritsche-Neto, R. (2023). SoilType: An R package to interplay soil characterization in plant science. *Agron. J.* 1–7, 848–854. doi: 10.1002/agi2.21383
- Furbank, R. T., and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005
- Gaffney, J., Schussler, J., Löffler, C., Cai, W., Paszkiewicz, S., Messina, C., et al. (2015). Industry-scale evaluation of maize hybrids selected for increased yield in drought-stress conditions of the US corn belt. *Crop Sci.* 55, 1608–1618. doi: 10.2135/cropsci2014.09.0654
- Gauch, Hugh, G., and Zobel, R. W. (1997). Identifying mega-environments and targeting genotypes. *Crop Sci.* 37, 311–326. doi: 10.2135/cropsci1997.0011183X003700020002x
- Gevertosky, R., Carvalho, H. F., Costa-Neto, G., Montesinos-López, O. A., Crossa, J., and Fritsche-Neto, R. (2023). Enviromic-based kernels may optimize resource allocation with multi-trait multi-environment genomic prediction for tropical Maize. *BMC Plant Biol.* 23, 10. doi: 10.1186/s12870-022-03975-1
- Hickey, L. T., Hafeez, N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754. doi: 10.1038/s41587-019-0152-9
- Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., et al. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15, 740. doi: 10.1186/1471-2164-15-740
- Kassambara, A., and Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses (1.0.7)*. Available online at: <https://cran.r-project.org/web/packages/factoextra/index.html> (accessed 01 July 2024).
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi: 10.18637/jss.v028.i05
- Liu, W., Yin, T., Zhao, Y., Wang, X., Wang, K., Shen, Y., et al. (2021). Effects of high temperature on rice grain development and quality formation based on proteomics comparative analysis under field warming. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.746180
- Montesinos-López, O. A., Crespo-Herrera, L., Saint Pierre, C., Bentley, A. R., de la Rosa-Santamaria, R., Ascencio-Laguna, J. A., et al. (2023). Do feature selection methods for selecting environmental covariables enhance genomic prediction accuracy? *Front. Genet.* 14. doi: 10.3389/fgene.2023.1209275
- Moura-Bueno, J. M., Dalmolin, R. S. D., Horst-Heinen, T. Z., Grunwald, S., and ten Caten, A. (2021). Environmental covariates improve the spectral predictions of organic carbon in subtropical soils in southern Brazil. *Geoderma* 393, 114981. doi: 10.1016/j.geoderma.2021.114981
- Neyhart, J. L., Silverstein, K. A. T., and Smith, K. P. (2022). Accurate predictions of barley phenotypes using genomewide markers and environmental covariates. *Crop Sci.* 62, 1821–1833. doi: 10.1002/csc2.20782
- Rogers, A. R., and Holland, J. B. (2022). Environment-specific genomic prediction ability in maize using environmental covariates depends on environmental similarity to training data. *G3 Genes[Genomes]Genetics* 12, jkab440. doi: 10.1093/g3journal/jkab440
- Senguttuvel, P., Sravanraju, N., Jaldhani, V., Divya, B., Beulah, P., Nagaraju, P., et al. (2021). Evaluation of genotype by environment interaction and adaptability in lowland irrigated rice hybrids for grain yield under high temperature. *Sci. Rep.* 11, 1–13. doi: 10.1038/s41598-021-95264-4
- Singh, M., Grando, S., and Ceccarelli, S. (2006). Measures of repeatability of genotype by location interactions using data from barley trials in northern Syria. *Exp. Agric.* 42, 189–198. doi: 10.1017/S0014479705003364
- Spindel, J. E., and McCouch, S. R. (2016). When more is better: How data sharing would accelerate genomic selection of crop plants. *New Phytol.* 212, 814–826. doi: 10.1111/nph.14174
- Srikanth, A., and Schmid, M. (2011). Regulation of flowering time: All roads lead to Rome. *Cell. Mol. Life Sci.: CMLS* 68, 2013–2037. doi: 10.1007/s00018-011-0673-y
- Stuerz, S., and Asch, F. (2019). Responses of rice growth to day and night temperature and relative air humidity—Dry matter, leaf area, and partitioning. *Plants* 8, 521. doi: 10.3390/plants8110521
- USDA (2023). *USDA ERS - Rice Sector at a Glance. Rice Sector at a Glance*. Available online at: <https://www.ers.usda.gov/topics/crops/rice/rice-sector-at-a-glance/> (accessed 01 July 2024).
- Yan, W., Nilsen, K. T., and Beattie, A. (2023). Mega-environment analysis and breeding for specific adaptation. *Crop Sci.* 63, 480–494. doi: 10.1002/csc2.20895