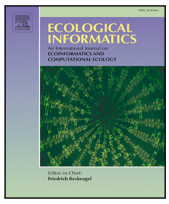
Ecological Informatics 88 (2025) 103110

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



A robust-weighted AMMI modeling approach with generalized weighting schemes

Marcelo B. Fonsêca [a,1], Vanda M. Lourenço [b,1,*], Paulo C. Rodrigues [a,c]

[a] Department of Statistics, Federal University of Bahia, Salvador, Brazil
[b] Center for Mathematics and Applications (NOVA Math) and Department of Mathematics, NOVA FCT, Quinta da Torre, 2829-516 Caparica, Portugal
[c] Statistical Learning Laboratory (SaLLy), Federal University of Bahia, Salvador, Brazil

ARTICLE INFO

Dataset link: https://github.com/marcelofonsecab/hybridRWAMMI

Keywords:
Additive main effects and multiplicative interaction model
Genotype-by-environment interaction
Robust modeling
Weighting schemes

ABSTRACT

The additive main effects and multiplicative interaction (AMMI) model and its variations are widely used to identify genotypes with specific adaptability and stability under environmental conditions in crop improvement breeding programs. However, atypical data points, arising from measurement errors, genotype characteristics, diseases, or climate phenomena, can significantly impact the model's performance, by contributing to the violation of its underlying assumptions. To address this challenge, we propose a hybrid modeling framework called robust-weighted AMMI (RW-AMMI), which combines robust and weighted algorithms to effectively model genotype-by-environment interaction (GEI) in the presence of data contamination and heteroscedasticity. We also introduce a comprehensive set of nine weighting schemes for the weighted (W-AMMI), robust (R-AMMI), and RW-AMMI models. Our extensive Monte Carlo simulations, which encompass both contaminated and uncontaminated data with and without heterogeneous error variance, demonstrate that several models within the W-AMMI, R-AMMI, and RW-AMMI classes perform competitively relative to the conventional AMMI model. Furthermore, we validate the effectiveness of the proposed approach using real crop data, where we leverage ensemble strategies to enhance genotype recommendations, providing practical evidence of its applicability. This work provides a hybrid framework for genotype selection under diverse environmental conditions, offering breeders a reliable tool for improving stability and adaptability.

1. Introduction

Genotype-by-environment interaction (GEI) refers to the relative responses of two or more genotypes to different levels of environmental conditions and is quite common in multi-environmental trials (MET). The modeling and understanding of GEI is a cornerstone in crop science to assist plant breeding under environmental changes so that the stability of genotypes, i.e., the ability of a genotype to consistently perform well across different environmental conditions, and the value of test sites can be established (Dia et al., 2016). The basis of GEI is the quantitative trait locus (QTL)-by-environment interaction (QEI), which occurs when QTL effects across environments differ, i.e., when QTLs are not stable across environments. A good understanding of these interactions is essential to selecting the best-performing genotypes under different environmental conditions (Rodrigues, 2018; Rodrigues et al., 2021).

Genotype-by-environment interaction can be quantified through several univariate and multivariate procedures based on evaluating genotypes under multiple environments. The most favored univariate methods are based on regressing the mean value of each genotype in the environmental index or marginal mean of the environments (Yates and Cochran, 1938; Finlay and Wilkinson, 1963; Eberhart and Russel, 1966). Multivariate analysis to the quantification of GEI is an alternative approach that includes complementary methods to evaluate genotype stability (Crossa, 1990), with the additive main effects and multiplicative interaction (AMMI) model of Gauch (1992) being the most widely used and preferred fixed effects model. Stability analysis, as discussed by Kang (2020), is particularly important when selecting genotypes for unpredictable and diverse growing conditions, ensuring that the chosen varieties are adaptable and perform consistently across various environments. This makes AMMI a powerful tool for breeders seeking to optimize both yield and stability in their crop selections. A notable advancement in this area is the integration of Bayesian methods with the AMMI model, which allows for a more nuanced interpretation of GEI patterns. This integration facilitates the identification of genotypes that perform consistently across diverse environments, representing a significant step forward in stability analysis.

* Corresponding author.
E-mail address: vmml@fct.unl.pt (V.M. Lourenço).
[1] Both authors contributed equally to this work and are considered co-first authors.

https://doi.org/10.1016/j.ecoinf.2025.103110
Received 23 September 2024; Received in revised form 13 March 2025; Accepted 16 March 2025
Available online 11 April 2025

Abbreviations

AMMI	Additive Main Effects and Multiplicative Interaction
ANOVA	Analysis of Variance
GEI	Genotype-by-Environment Interaction
IRWLS	Iterated Re-weighted Least Squares
MPEV	Mean Percentage of Explained Variability
MSE	Mean Squared Error
MET	Multi-Environment Trials
MtSPE	Mean Trimmed Squared Prediction Error
QEI	Quantitative Trait Locus-by-Environment Interaction
QTL	Quantitative Trait Locus
R-AMMI	Robust Additive Main Effects and Multiplicative Interaction
RW-AMMI	Robust-Weighted Additive Main Effects and Multiplicative Interaction
SVD	Singular Value Decomposition
W-AMMI	Weighted Additive Main Effects and Multiplicative Interaction

For example, studies by Bernardo Júnior et al. (2018) and da Silva et al. (2019) have demonstrated the effectiveness of Bayesian AMMI models in, respectively, handling unbalanced data and accounting for heterogeneity of variances across environments.

The AMMI model uses: (i) the analysis of variance (ANOVA) to estimate the genotypic and environmental main effects (additive effects); and (ii) the singular value decomposition (SVD) applied to the residuals of the ANOVA, to estimate the components that characterize the interaction, i.e., that summarize GEI (multiplicative effects). This method has proven to be effective because it is able to capture a large part of the sum of squares of the interaction, it allows the separation of the main and the interaction effects and, in addition, it provides a meaningful interpretation of the data (Ebdon Jr. and Gauch, 2002). In other words, it can extract a large part of GEI, and it is, therefore, more efficient in analyzing the interaction pattern (Gauch, 1992; Gauch and Zobel, 1997; Ebdon Jr. and Gauch, 2002; Paderewski, 2011; Paderewski and Rodrigues, 2014, 2018; Gauch Jr. et al., 2011).

Another interesting feature of the AMMI model is that its results can be visually assessed in graphs called biplots that display the genotype scores (scores plot) and the environment scores (loadings plot) on a single plot (Gauch and Zobel, 1997). These biplots are particularly useful for investigating the interactions between genotypes and environments, specifically aiding in the identification of genotypes that are stable and well-adapted to specific environmental conditions (Gauch, 1988). Beyond their pivotal role in stability analysis and genotype-environment interaction studies, biplots have been widely discussed in the literature, with applications ranging from ecology, biostatistics, and finance to social sciences, as outlined in the comprehensive guide on biplot analysis of Gower et al. (2011). Recently, biplots have gained attention in machine learning applications, such as digital soil mapping, where they serve as a model-agnostic tool to interpret predictions (van der Westhuizen et al., 2024), showcasing their broader applicability and value in exploring complex interactions in various fields of study.

Rodrigues et al. (2014) went one step ahead and proposed a weighted AMMI modeling approach (W-AMMI model) to account for phenotypic data with moderate to strong heterogeneous error variance across environments in the study of GEI and QEI. In the W-AMMI model: (i) a weighted least-squares fit replaces the ANOVA linear fit, and (ii) the standard SVD is replaced by a weighted SVD algorithm that uses the weights estimated in the previous step. Hadasch et al. (2018) proposed additional weighted AMMI and genotype-and-genotype–environment interaction algorithms, including a modification of Rodrigues et al. (2014) which takes correlations between genotype–environment means into account. The authors concluded that the proposed algorithms yield identical point estimates of the model effects and, therefore, can be utilized with similar efficiencies regarding the evaluation criteria considered. Further, Rodrigues et al. (2014) emphasized that their weighting scheme allows the definition of cell-specific weights, making it applicable to any situation where differential weighting of cell means is necessary. Specifically, in the context of MET, they noted that the W-AMMI approach could be an alternative to mixed model analyses, such as those described by Piepho et al. (2012), where weights are based on the standard errors of genotype means derived from the first stage of a two-stage analysis.

Data contamination is more often the rule than the exception in MET studies and other real-world research, arising from various sources, including experimental, measurement, and data entry errors, as well as inherent characteristics of genotypes, pest impacts, and climate phenomena. Such contamination poses significant challenges to statistical analysis, as it can lead to violations of fundamental model assumptions, particularly concerning the normality and homoscedasticity of the residuals. These violations can, in turn, result in unreliable model estimates and inferences. Because of these issues, the AMMI model may yield unreliable results when phenotypic data contain outliers, as these data points can disrupt ANOVA (Lourenço et al., 2011) and subsequently impact SVD (along with all other analyses relying on AMMI model results, such as stability indexes and QEI and QTL detection; Hawkins et al., 2001; Hubert et al., 2002, 2005). To address these challenges, Rodrigues et al. (2016) proposed a robust version of the AMMI model for the study of GEI (R-AMMI model). This robust model accommodates the presence of atypical data points (i.e., outliers), enabling more stable inferential results and improved biplot interpretation. The idea behind the robust AMMI model is straightforward: in (i) the ANOVA linear fit is replaced by a robust fit (M-regression; Huber, 1964; Lourenço et al., 2011) and in (ii) the standard SVD is replaced by a robust SVD algorithm (Filzmoser and Todorov, 2013).

Another potential approach to addressing data contamination in the AMMI model is to utilize robust alternating regression (Croux and Filzmoser, 1998). This method estimates both the additive and multiplicative components simultaneously, rather than employing a two-stage model where the additive part of the AMMI model is fitted first, followed by the multiplicative part. While various outlier detection methods have been proposed in the literature (Gumedze et al., 2010; Schützenmeister, 2012; Lourenço and Pires, 2014; Tanaka, 2020), integrating robust estimation within the AMMI framework may be preferable. On the one hand, classical outlier detection methods face several challenges, such as (i) true outliers are not always visible due to masking and swamping effects; (ii) gross outliers can be identified more easily than mild outliers, complicating the analysis; and (iii) outliers may still contain valuable information (Lourenço et al., 2011). On the other hand, as emphasized by Tanaka (2020), even within robust contexts, (i) there remain challenges in selecting an appropriate statistic for outlier detection and establishing thresholds for classifying outliers; (ii) current methods often consider one observation at a time, making it difficult to address swamping and masking effects, specially in complex data; and (iii) available software implementations for robust linear mixed models may lack the flexibility to accommodate various error structures, limiting their effectiveness in identifying outliers in MET data.

This work is two-fold. First, we propose a hybrid modeling framework for the AMMI model, referred to as the RW-AMMI model, which combines the weighted and robust AMMI models of Rodrigues et al. (2014, 2016), respectively, for the study of GEI and QEI, with the capability to extend to QTL detection. Second, we introduce a comprehensive set of nine weighting schemes designed for use in the

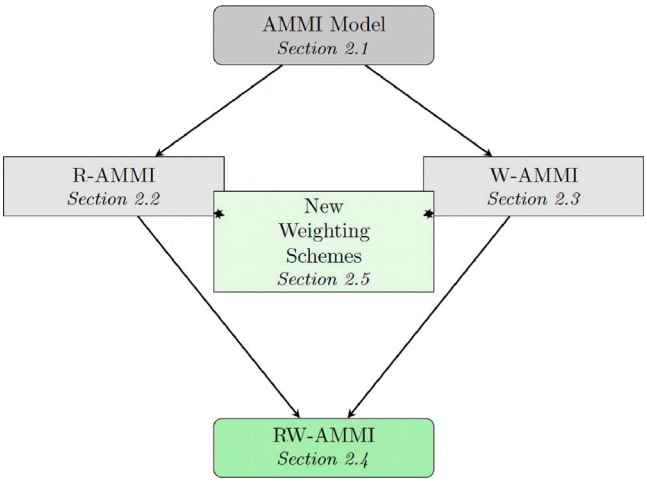


Fig. 1. Flowchart illustrating the structure and progression of the methods presented in Section 2.

weighted, robust, and robust-weighted models. The performance of the proposed hybrid approach is evaluated against the AMMI, W-AMMI, and R-AMMI counterparts, taking into account the nine weighting schemes through a Monte Carlo simulation. This evaluation specifically explores the impact of data contamination on the robustness of the aforementioned methodologies. Contamination is simulated to mimic real-world challenges, introducing deviations from the expected normal distribution. This contamination study serves as a crucial component in assessing the reliability and robustness of our proposed models under practical challenges encountered in data analysis. Additionally, the analysis of a real dataset from the doubled-haploid Steptoe x Morex barley (*Hordeum vulgare L.*) population (Hayes et al., 1993) further illustrates the usefulness of the proposed methodology. Indeed, the hybrid AMMI model can be applied in all areas where the AMMI model is considered. Overall, we aim to provide a novel framework that effectively handles data contamination and heteroscedasticity in GEI analysis, ensuring stability and offering a more reliable tool for breeders and researchers.

It is important to note that the detection of data entry errors warrants their correction or exclusion from analysis, rather than attempting to address them through robust statistical methods. Only after ensuring data integrity should robust analytical approaches be applied.

2. Materials and methods

Herein, Tables S1, S2,..., and Figures S1, S2,...,refer to Tables and Figures in the Supplementary Materials (Chapters 6–9). Fig. 1 illustrates the structural flow and dependence of the methods discussed in this section. We start with the AMMI model (Section 2.1; Gauch, 1992), progressing through its robust and weighted extensions (respectively, R-AMMI and W-AMMI; Sections 2.2–2.3; Rodrigues et al., 2014, 2016), leading to the final robust-weighted approach (RW-AMMI; Section 2.4). Finally, the new weighting schemes are presented (Section 2.5). When replacing the natural weighting schemes within the R-AMMI and W-AMMI models, these new schemes provide new formulations for both R-AMMI and W-AMMI, and thus also lead to new formulations of RW-AMMI. The performance of all these versions is evaluated in this work, with Section 3 detailing the simulation setup and comparison criteria, and Section 4 presenting the results.

Within each subsection, besides the theoretical aspects of the methods and the relevant references, a step-by-step approach is provided to help the reader follow the progression and implementation of each method more easily.

2.1. AMMI model

The additive main effects and multiplicative interaction model is a modeling technique for the study of genotype-by-environment interactions that combines the ANOVA and SVD in a two-step sequential analysis assuming that all effects, except for the error, are fixed. First, the ANOVA estimates the additive main effects of the two-way genotype-by-environment data table, and then, the SVD is applied to the residuals of the ANOVA in order to estimate the $N \leq min(I, J)$ interaction principal components (IPC). In this case, I represents the number of genotypes (rows) and J the number of environments (columns). Assuming for simplicity a completely randomized design for individual trials, the AMMI model can be written as (Gauch, 1992; Rodrigues, 2018):

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \sum_{n=1}^{N^*} \lambda_n \gamma_{n,i} \delta_{n,j} + \eta_{i,j} + \varepsilon_{i,j,k}, \tag{1}$$

where $Y_{i,j,k}$ is the phenotypic trait of the genotype i in the environment j for replicate k, μ is the grand mean (i.e., the mean value of the phenotypic trait computed across all genotypes, environments, and replicates), α_i are the genotype main effects as deviations from μ, β_j are the environment main effects as deviations from μ, λ_n is the singular value for the n-*th* IPC, $\gamma_{n,i}$ and $\delta_{n,j}$ are the ith and jth genotype and environment scores for the n-*th* IPC, i.e., the left and right singular vectors, $N^* < N$ is the number of selected IPC for the model, $\eta_{i,j}$ is the residual containing all multiplicative terms not included in the model, and $\varepsilon_{i,j,k}$ is the experimental error referring to the error for genotype i, environment j and replicate k such that $\varepsilon_{i,j,k} \underset{iid}{\sim} N(0, \sigma^2)$. The model can also be written in a matrix form as:

$$\mathbf{Y} = \mathbf{1}_I \mathbf{1}_J^T \mu + \boldsymbol{\alpha}_I \mathbf{1}_J^T + \mathbf{1}_I \boldsymbol{\beta}_J^T + \mathbf{U}_{N^*} \mathbf{D}_{N^*} \mathbf{V}_{N^*}^T + \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{Y}$ is a $(I \times J)$ matrix referring to the two-way table of phenotypic means computed across the r replicated two-way tables, $\mathbf{1}_I \mathbf{1}_J^T \mu$ is the matrix with the grand mean μ in all positions, $\alpha_I \mathbf{1}_J^T$ is a matrix of genotypic main effects as deviations from the grand mean, and $\mathbf{1}_I \beta_J^T$ is a matrix of environment main effects as deviations from the grand mean. The interaction part of the model $\mathbf{Y}^* = \mathbf{Y} - \mathbf{1_I 1_J^T} \hat{\mu} - \hat{\boldsymbol{\alpha}}_\mathbf{I} \mathbf{1_J^T} - \mathbf{1_I} \hat{\boldsymbol{\beta}}_\mathbf{J}^\mathbf{T}$ can be approximated by the product of matrices $\mathbf{U}_{N^*} \mathbf{D}_{N^*} \mathbf{V}_{N^*}^T$ with $\mathbf{U}_{N^*}$ being the $(I \times N^*)$ matrix with the left singular vectors of the interaction, $\mathbf{V}_{N^*}$ the $(J \times N^*)$ matrix with the right singular vectors of the interaction, and $\mathbf{D}_{N^*}$ the diagonal $(N^* \times N^*)$ matrix containing the N^* singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N^*} \geq 0$ of $\mathbf{Y}^*$ (with the number of non-zero singular values equal to the rank of $\mathbf{Y}^*$). The $(I \times J)$ $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ matrices contain the multiplicative terms not included in the model and the experimental error, respectively.

In practice, the fit of the AMMI model in any statistical software can be done in two sequential steps. Without loss of generality, let us assume that the data is either non-replicated or that the mean trait phenotypic values across replicates have been previously determined for each environment and genotype. Then, we can fit the model by the following steps, which refer, respectively, to the additive and multiplicative parts of the AMMI model:

*S*1: fit the linear model (i.e. ANOVA)

$$Y \sim Gen + Env + \varepsilon^*, \tag{3}$$

where Y includes the mean trait phenotypic values (computed across all replicates), *Gen* and *Env* are the factors associated with the genotypes and environments, respectively, and ε^* refers to the random errors.

*S*2: perform the singular value decomposition of the residuals $\hat{\varepsilon}$ with $N \leq min(I, J)$ and keep the decomposition matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{D}$.

Once this is done, the number of significant components, i.e., those associated with the signal, say N^*, must be selected. The selection of the proper number of multiplicative terms associated with the signal has been studied by Rodrigues et al. (2014), Malik et al. (2019), Forkman et al. (2022) and references therein. After appropriate selection, these significant multiplicative terms can then be used to obtain biplots and perform other relevant analyses.

While our main emphasis lies outside the scope of addressing missing data explicitly, it is important to acknowledge that models can still be fitted even in the presence of missing data. Imputation methods are advisable before fitting the models when applicable, and this practice, although not mandatory, is recommended for enhancing the robustness of the analysis (see Gauch and Zobel, 1990; Paderewski, 2013; Paderewski and Rodrigues, 2014; Arciniegas-Alarcón et al., 2020; Podani et al., 2021, for further details on data imputation).

2.2. Robust AMMI model (R-AMMI)

Rodrigues et al. (2016) proposed a robust version of the AMMI model (R-AMMI) for the study of GEI, where the word robust refers to the ability of the model to accommodate for outliers and other small departures from the model's assumptions. The idea behind the R-AMMI is to fit model (3) using robust M-regression and consider a robust SVD of the robust residuals in place of the traditional SVD. Specifically, assuming that either the data is non-replicated or the median trait phenotypic values across replicates have been previously determined for each environment and genotype (coincides with the mean values in the case of two-replicated data), the R-AMMI algorithm can be estimated by the following steps:

*S*1: fit the linear model (3) using Huber's loss function (Huber, 1973)

$$\rho(\epsilon_i) = \begin{cases} \frac{1}{2}\epsilon_i^2, & \text{if} \quad |\epsilon_i| \le c \\ c(|\epsilon_i| - \frac{1}{2}c), & \text{if} \quad |\epsilon_i| > c \end{cases}$$

and keep the estimated residuals, say $\widehat{\varepsilon}$ ($I \times J$ matrix; as in the AMMI model above, it is assumed that the genotypic means across replicates have been previously computed for each environment). Note that here: (i) c is some tuning constant that provides a compromise between efficiency (at the normal model) and robustness (to outlying observations at the response variable level); (ii) when c grows to infinity, the loss becomes quadratic thus allocating us to least squares estimation; (iii) a usual value of c is $c = 1.345$, which conveys 95% asymptotic relative efficiency at the normal model to the resulting M-estimator;

*S*2: perform a robust SVD of $\widehat{\varepsilon}$ using the method of Hawkins et al. (2001), keep the decomposition matrices **U**, **V**, and **D**, and select the first few components based on some criteria of interest; this robust SVD algorithm adopts a sequential estimation of the singular values and left and right singular vectors that not only ignores missing values (note that conventional SVD needs that matrix $\widehat{\varepsilon}$ is complete) but is also resistant to outliers; the method differs from the traditional alternating least squares SVD (ALS–SVD) as it considers the L1 norm (AL1–SVD) instead of the more usual least squares L2 norm, to compute a robust approximation to the SVD of a rectangular matrix.

Robust statistical models are designed to explain the bulk of the data rather than the whole data and have been successfully applied in plant breeding and quantitative genetics (Estaghvirou et al., 2014; Lourenço et al., 2017, 2020; Tanaka, 2020). In robust regression, this is achieved by down-weighting observations with larger residuals while assigning higher weights to observations with smaller residuals. The weights depend on the objective function ρ utilized. Huber's ρ, for example, uses $\psi(x) = \rho'(x)$ as the influence function, resulting in the following weights:

$$w(\epsilon_i) = \frac{\psi'(\epsilon_i)}{\epsilon_i} = \begin{cases} 1, & \text{if} \quad |\epsilon_i| \le c \\ c/|\epsilon_i|, & \text{if} \quad |\epsilon_i| > c. \end{cases} \tag{4}$$

In this case, the weights approach zero as the absolute value of the residuals increases to infinity, whereas for residuals with absolute value greater or equal to c, all the weights equal one. Since the weights in Eq. (4) are dependent on the residuals, which, in turn, depend on the estimated coefficients, the estimation process in a robust regression model is accomplished through iterated re-weighted least squares (IRWLS). In this method, initial weight values are assigned to the observations, and the regression coefficients are estimated using these weights. Then, the residuals are computed, and the weights are updated based on the values of the residuals. This process is repeated until the estimated coefficients and the weights converge to stable values. By iteratively updating the weights, the robust regression model is able to down-weight the influence of outliers and improve the accuracy of the estimated coefficients. In classical linear regression, however, the loss function used is the quadratic loss $\rho(x) = x^2$, which means that all observations are equally weighted (weights equal to one) when fitting the model.

2.3. Weighted AMMI model (W-AMMI)

The weighted additive main effects and multiplicative interaction (W-AMMI) model proposed by Rodrigues et al. (2014) is a generalization of the AMMI model that allows for individual cells, rows, or columns to be assigned specific weights. This is particularly useful when the error variance–covariance structure for the genotypic means differs between trials, replicates, and/or environments. Because blocks, incomplete blocks, and spatial trends may produce different variances and correlations, this heterogeneity needs to be accounted for in the analysis of GEI (Rodrigues et al., 2014). To fit a W-AMMI model, the weighted least squares algorithm is applied to the two-way data table to estimate the additive effects. The residuals from the weighted least squares fit are then used to estimate the multiplicative terms of the interaction with an algorithm based on weighted low-rank SVD decomposition (Srebro and Jaakkola, 2003), which is an iterative process based on the expectation–maximization (EM) algorithm, that can be written as:

$$\mathbf{X}^{(t+1)} = \text{SVD}\left[\mathbf{W} \odot \mathbf{Z} + (\mathbf{1} - \mathbf{W}) \odot \mathbf{X}^{(t)}\right] \tag{5}$$

where $\mathbf{Z}$ is the matrix that has the multiplicative terms obtained by weighted least squares, $\mathbf{W}$ is a $(I \times J)$ matrix with weights $W_{i,j}$, where $0 \le W_{i,j} \le 1$, $\mathbf{1}$ is a $(I \times J)$ matrix with 1 in all positions, $\odot$ is the Hadamard product of matrices, t is the number of iterations, and $\mathbf{X}$ is a low-rank approximation that can be initialized as $\mathbf{X}^{(0)} = \mathbf{Z}$ or $\mathbf{X}^{(0)} = \mathbf{0}$ for convergence to the global minimum. The results are the matrices $\mathbf{U}_{N^*}$, $\mathbf{D}_{N^*}$ and $\mathbf{V}_{N^*}$ such that $\mathbf{Y}^* \approx \mathbf{U}_{N^*}\mathbf{D}_{N^*}\mathbf{V}_{N^*}^T$, where N^* is the rank of the matrix or the number of components retained in the model. Summing up, the steps of the W-AMMI model, which assume that the data are replicated, are:

*S*1: fit the linear mixed effects model

$$Y \sim Rep + (1|Gen) + \varepsilon_e \tag{6}$$

for every environment (here Y refers to the phenotypic values obtained across the r replicates; Rep denotes the fixed effect of the replicates, which accounts for variability due to different experimental runs or iterations; ε_e denotes the unexplained variation, encompassing all influences not captured by the fixed and random effects included in the model) and keep the estimated residual variances, say variances $\widehat{\sigma}^2_{\varepsilon_{e_j}}$, $j = 1, \ldots, J$; then construct a $I \times J$ two-way table of weights, say W, where

$$W_e^{ij} = \frac{1/\widehat{\sigma}^2_{\varepsilon_{e_j}}}{max_{1 \le j \le J}\ (1/\widehat{\sigma}^2_{\varepsilon_{e_j}})} \times \frac{r_{ij}}{r}$$

with r the number of replicates in the study (usually two in multi-environment trials), r_{ij} the number of replications of genotype

i in the environment j (note that replications could be missing; when this is not the case, then $r_{ij}/r = 1\ \forall i, j$, as is the case of our simulated data), $i = 1, \ldots, I$; $j = 1, \ldots, J$; taking $m = max_{1\le j\le J}\ (1/\hat{\sigma}^2_{\varepsilon_{e_j}})$ and assuming that the data is complete (when this is not the case, usually imputation methods are used before fitting the models), the matrix of weights W_e takes the simpler form

$$\mathbf{W}_e = \begin{bmatrix} \frac{1/\sigma^2_{11}}{m} & \frac{1/\sigma^2_{12}}{m} & \cdots & \frac{1/\sigma^2_{1J}}{m} \\ \frac{1/\sigma^2_{21}}{m} & \frac{1/\sigma^2_{22}}{m} & \cdots & \frac{1/\sigma^2_{2J}}{m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1/\sigma^2_{I1}}{m} & \frac{1/\sigma^2_{I2}}{m} & \cdots & \frac{1/\sigma^2_{IJ}}{m} \end{bmatrix}, \tag{7}$$

where $\sigma^2_{ij} = \hat{\sigma}^2_{\varepsilon_{e_j}}\ \forall i$.

S2: fit the linear model (3) via weighted least squares using the matrix of weights W_e from the previous step and keep the estimated residuals, say $\hat{\varepsilon}$ ($I \times J$ matrix);

S3: perform a weighted low-rank SVD of $\hat{\varepsilon}$, using the algorithm defined in (5), considering the matrix of weights W_e, keep the decomposition matrices **U**, **V**, and **D**, and select the first few components based on some criteria of interest.

2.4. Robust weighted AMMI model (RW-AMMI)

The hybrid AMMI model proposed in this work (RW-AMMI), combines the strengths of both the W-AMMI and R-AMMI models described above in a couple of different ways. Specifically, one considers the following estimation steps:

S1: fit model (6) using the robust linear mixed effects model approach of Koller (2013, 2016) for every environment and keep the estimated robust residual variances, say robust variances $\hat{\sigma}^2_{\varepsilon^R_{e_j}}$, $j = 1, \ldots, J$; then construct the $I \times J$ two-way table of weights W as described in W-AMMI–S1;

S2: fit model (3) using Huber's loss function together with the matrix of weights W obtained in S1 and where now Y includes the median trait phenotypic values (computed across all replicates); keep the estimated residuals, say $\hat{\varepsilon}$;

S3: taking the estimates of $\hat{\varepsilon}$ and W from S2 above perform a weighted low-rank robust SVD of $\hat{\varepsilon}$ (W-AMMI–S2); keep the decomposition matrices **U**, **V**, and **D**, and select the first few components based on some criteria of interest.

2.5. Generalized weighting schemes for the AMMI models

In this section, we generalize the matrix of weights W used in the W-AMMI model proposed in Section 2.3. Modifications to W can lead to different parameter estimates, interpretations of effects, and ultimately, different conclusions about the data. The sensitivity of the AMMI model to changes in W implies that the robustness of the model's conclusions can vary depending on the chosen weights. This becomes particularly significant in applications where the model's predictions are used for decision-making. Evaluating different configurations of W enables researchers to investigate how robust the model's conclusions are to alternative assumptions regarding the relationships between parameters. Therefore, although the overarching methodology remains AMMI, each configuration of W can yield a unique version of the model that must be evaluated independently for its practical utility. In particular, we consider in the W-AMMI–S1 also the fit of the linear mixed effects model

$$Y \sim Rep + (1|Env) + \varepsilon_g \tag{8}$$

for every genotype (Y and Rep as in Eq. (6); ε_g denoting the unexplained variation, encompassing all influences not captured by the fixed and random effects included in the model) and keep the estimated residual variances, say variances $\hat{\sigma}^2_{\varepsilon_{g,i}}$, $i = 1, \ldots, I$. We then construct the $I \times J$ two-way table of weights W_g as

$$W_g^{ij} = \frac{1/\hat{\sigma}^2_{\varepsilon_{g,i}}}{max_{1\le i\le I}\ (1/\hat{\sigma}^2_{\varepsilon_{g,i}})} \times \frac{r_{ij}}{r},$$

with r_{ij} and r as in W-AMMI–S1. Taking $\tilde{m} = max_{1\le i\le I}\ (1/\hat{\sigma}^2_{\varepsilon_{g,i}})$ and assuming data are complete (when this is not the case, usually imputation methods are used prior to fitting the models), the matrix of weights W_g takes the simpler form of (7) where one now has $\sigma^2_{ij} = \hat{\sigma}^2_{\varepsilon_{g,i}}\ \forall j$ and $m = \tilde{m}$.

From the W_e and W_g weighting schemes, to which we will herein refer to as $\sigma^2(E)$ and $\sigma^2(G)$, respectively, now follow the additional weighting ones:

- $\sigma^2(G \odot E) = W_e \odot W_g$
- $\sigma^2(G \oplus E) = \frac{W_e \oplus W_g}{2}$
- $\sigma^2(rlm) = W_R$
- $\sigma^2(rlm \odot E) = W_R \odot W_e$
- $\sigma^2(rlm \odot G) = W_R \odot W_g$
- $\sigma^2(rlm \odot G \odot E) = W_R \odot W \odot W_g$
- $\sigma^2(rlm \odot (G \oplus E)) = W_R \odot \frac{W_e \oplus W_g}{2}$

where W_R is the matrix of weights obtained from the robust fit of the model (3) using the Huber M-estimator (*rlm* stands for robust linear model), $\oplus$ is the sum of the matrices, and $\odot$ is the Hadamard product (element-wise product). Robust versions of W_e and W_g are obtained by the robust fits of models (6) and (8), respectively. Other robust weight generalizations are possible, e.g., by considering the robust weights that are estimated directly from the single robust linear mixed effects fit of models

$$Y \sim Rep + (1|Gen) + \varepsilon \quad \text{and} \quad Y \sim Rep + (1|Env) + \varepsilon.$$

In this case, for each model fit, r ($I \times J$) matrices of weights are estimated, which may be combined into a single weight matrix, e.g., by taking the median weight value across the r matrices, resulting in yet new robust W_e and W_g matrices. The possible ways of combining the r LMM estimated weight matrices plus the combination of final W_e and W_g matrices, as previously suggested, would give rise to a huge amount of simulation scenarios. Since this is not an exhaustive simulation study and since we cover the most natural robust generalizations of the classical AMMI, this setting will not be pursued further.

As a final note, while the choice of W can significantly influence the outcomes of the AMMI model, one notes that there is no one-size-fits-all rule for selecting variations of W. Researchers should consider the specific objectives of their analysis, the characteristics of the data, and theoretical frameworks guiding their work. It is essential to conduct sensitivity analyses to understand how different configurations impact results. Ultimately, documenting and justifying the choice of weights will enhance the robustness and interpretability of the AMMI model's conclusions.

Table 1 (replicated in Table S1) presents the list of AMMI models resulting from the proposed generalized schemes and whose performance will be studied here.

All the methods were implemented in R, with the codes available to ensure result reproducibility and facilitate method adoption (Section **Data availability**). The real data application serves as a practical guide for applying the approach to other datasets, with improved documentation to support its use.

3. Simulation

3.1. Data simulation and simulation settings

We consider $m = 100$ datasets using the same structure as in Rodrigues et al. (2016). Each data consists of yield values across 100

Table 1
Notation key for AMMI model variations.

Model	Weighting scheme
W-AMMI{1}; R-AMMI{1}; RW-AMMI{1}	$\sigma^2(E)$
W-AMMI{2}; R-AMMI{2}; RW-AMMI{2}	$\sigma^2(G)$
W-AMMI{3}; R-AMMI{3}; RW-AMMI{3}	$\sigma^2(G \odot E)$
W-AMMI{4}; R-AMMI{4}; RW-AMMI{4}	$\sigma^2(G \oplus E)$
W-AMMI{5}; R-AMMI{5}; RW-AMMI{5}	$\sigma^2(rlm)$
W-AMMI{6}; R-AMMI{6}; RW-AMMI{6}	$\sigma^2(rlm \odot E)$
W-AMMI{7}; R-AMMI{7}; RW-AMMI{7}	$\sigma^2(rlm \odot G)$
W-AMMI{8}; R-AMMI{8}; RW-AMMI{8}	$\sigma^2(rlm \odot (G \odot E))$
W-AMMI{9}; R-AMMI{9}; RW-AMMI{9}	$\sigma^2(rlm \odot (G \oplus E))$

W-AMMI stands for the weighted AMMI; R-AMMI stands for the robust AMMI; RW-AMMI stands for the robust and weighted AMMI; numbers 1–9 refer to the weighting scheme being considered.

genotypes, 8 environments and 2 replicates obtained in the following way:

1. Take a (100×8) matrix $\mathbf{X}$ from a $U(-0.5, 0.5)$ distribution.
2. Perform a singular value decomposition of matrix $\mathbf{X}$ in order to obtain the matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{D}$.
3. Simulate, respectively, a global mean (μ), genotypic (α) and environmental (β) effects as: $\mu \sim N(15, 3)$, $\alpha \sim N(5, 1)$, and $\beta \sim N(8, 2)$.
4. Create the double-entry matrix $\mathbf{Y}$ of simulated yield values from the AMMI model structure with two multiplicative terms (i.e., with two components):

$$\begin{aligned}\mathbf{Y} = \mathbf{1}_I\mathbf{1}_J^T\mu + \alpha_I\mathbf{1}_J^T + \mathbf{1}_I\beta_J^T + 28 \times \mathbf{U}[,1]\mathbf{D}[1,1]\mathbf{V}[1,]^T \\ + 15 \times \mathbf{U}[,2]\mathbf{D}[2,2]\mathbf{V}[2,]^T\end{aligned} \tag{9}$$

   with the coefficients of the first and second components equal to 28 and 15, respectively. This ensures that the first component has higher importance in the simulating process.
5. Replications are obtained by adding $N(0, 1)$ random noise to the signal $\mathbf{Y}$, considering different seeds.

3.2. Data contamination

In this section, we detail the data contamination procedure within the context of Huber's contamination model—a widely recognized framework for assessing distributional robustness when facing deviations from the assumed model (Huber, 1964). The model writes as $(1 - \varepsilon) * \mathrm{F} + \varepsilon * \mathrm{G}$, where F denotes the assumed population distribution, in our case a $N(\mu, \sigma^2)$ distribution, $\varepsilon \in]0, 0.5[$ represents the contamination percentage, and G is a contaminating distribution. Consistent with this mixture model, data contamination is done by replacing a percentage of good data points (generated from $N(\mu, \sigma^2)$) with bad data points (generated from G), i.e., with several types of outliers. Specifically, one considers 2%, 5% and 10% of *shift*, *variance-inflated* and *pointmass* outliers, randomly spread across the first replicate. The outliers are generated from the following G distributions:

- *shift outliers* (one tail): $N(\mu_j + k\sigma_j, \sigma_j^2)$ with sizes $k = 4, 7, 10$
- *variance-inflated outliers*: $N(\mu_j, 5\sigma_j^2)$
- *pointmass outliers*: $N(\mu_j, \sigma_j^2/10)$

where μ_j and σ_j^2 are the phenotypic mean and sample variance corresponding to the jth environment, respectively. One tail *shift*-outliers, which are considered here, are usually more harmful than two tail shift outliers as the effect of the latter often cancels out. In addition, one-tail outliers are known to have a more deleterious effect on estimation than *variance-inflated* and *pointmass* outliers, with the latter being the least detrimental type of outliers. Also, note that the $10-shift$, 10% scenario, is a rather extreme contamination scenario, not likely to occur in field experiments, that is considered here to show how the methods compare when their breakdown points are surpassed.

It is worth noting that the simulation of outliers considered here is aimed at mimicking real-life scenarios observed in crop breeding data. While shift outliers may reflect instances where extreme values may occur due to genetic mutations or environmental factors leading to unusually high or low performance in certain genotypes (e.g., over- or under-irrigation/fertilization in a specific field location), variance-inflated outliers could simulate situations where genetic variations or environmental conditions contribute to increased trait variability (e.g., the presence of diverse soil types within a breeding program or the exposure of crops to varying climate conditions), and point-mass outliers may reflect scenarios where specific genetic traits or environmental conditions cause a reduction in trait variability (e.g., the introduction of a new genetic variant or the uniform application of a particular treatment across experimental units).

3.3. Performance assessment

To effectively compare the various methods employed in our study, we utilize a set of performance metrics that provide quantitative insights into the effectiveness of each method. These criteria are designed to capture multiple dimensions of model performance, including prediction accuracy, explained variability, and interpretability. The following criteria are used for this assessment:

3.3.1. Mean trimmed squared prediction error (*MtSPE*)

Mean trimmed prediction errors involve calculating the mean after removing a certain percentage of extreme values from both ends of the distribution of the differences between observed and predicted values. By calculating the mean after trimming a certain percentage of extreme values from the upper tail of the distribution of the squared differences, an even more robust measure of central tendency is ensured. This approach reduces the influence of outliers, offering a robust assessment of the performance of competing methods.

Let y_{ij}^s refer to the ij-th observation from the two-way data table referring to the sth simulation and $\hat{y}_{ij}^s$ its estimated value. First, the $I \times J$ squared differences $\left(y_{ij}^s - \hat{y}_{ij}^s\right)^2$ are taken and assigned to a vector, say $d_1^s, \ldots, d_{I*J}^s$. Second, the differences are sorted in ascending order such that $d_{(1)}^s < \cdots < d_{(I*J)}^s$. Finally, the 10% upper tail observations of vector $d_{(1)}^s, \ldots, d_{(I*J)}^s$ are trimmed, and the mean trimmed squared prediction error is computed as

$$MtSPE = \frac{1}{m}\sum_{s=1}^{m}\sum_{k=1}^{IJ-t}\frac{d_{(k)}^s}{IJ-t} \tag{10}$$

where t refers to the number of trimmed observations. The smaller the **MtSPE**, the better the performance of the model in predicting the bulk of the data.

3.3.2. Mean percentage of explained variability (*MPEV*)

$$MPEV = \frac{1}{m}\sum_{s=1}^{m}\frac{\hat{\lambda}_1^s + \hat{\lambda}_2^s + \cdots + \hat{\lambda}_{N^*}^s}{\lambda_1^s + \lambda_2^s + \cdots + \lambda_J^s} \tag{11}$$

where N^* is the number of estimated components from the multiplicative terms (in our case $N^* = 2$), $\lambda_j^s, j = 1, \ldots, J$, are the true eigenvalues, J is the number of columns (environments), and s is the index of the current simulation. The closer the values of **MPEV** are to 100%, the better.

3.3.3. Mean squared error j (*MSE*$_j$)

$$MSE_j = \frac{1}{m}\sum_{s=1}^{m}(\hat{\lambda}_j^s - \lambda_j)^2, \tag{12}$$

for $j = 1, \ldots, N^*$ and where λ_j refers to the jth singular value estimated from the AMMI model considering the uncontaminated data and $\hat{\lambda}_j^s$

refers to that estimated from the competing models in the $s-th$ simulation. As is, in the uncontaminated scenario, the $\mathbf{MSE}_j$ referring to the AMMI model are equal to zero. Overall, the smaller the $\mathbf{MSE}_j$, the better the performance of the model. In addition, the $\mathbf{MSE}_j$ can be summed over the number of selected components N^* for an overall comparison of the performance of the methods.

3.3.4. Mean maxsub (Krzanowski, 1979)

$$maxsub = \frac{1}{m}\sum_{s=1}^{m} arccos(\sqrt{\lambda_k^s}), \tag{13}$$

where λ_k^s is the smallest eigenvalue of the matrix $I_{N^*J}^T P_{JN^*} P_{N^*J}^T I_{JN^*}$ in the $s-th$ simulation, where I_{JN^*} are the true orthogonal *loadings*, and P_{JN^*} are the estimated *loadings*. The best value for maxsub is 0, and commonly the values are standardized by dividing with $\frac{\pi}{2}$ to make the values range from 0 to 1. Smaller values of *maxsub* represent a good approximation between the representation in the N^* dimensions obtained by a given model and the true simulated signal. Hence, values of *maxsub* close to one are not desirable.

3.3.5. Biplot interpretation

In general, each model aims to accurately represent the data in a two-dimensional referential, enabling the identification of the best genotypes for each environment. In the biplots, the loadings represent the environmental factors, and the direction from the origin to a point indicates the environment. Using an orthogonal projection, we can rank all genotypes based on their alignment with the loadings (environments). The projection is calculated using the equation $\frac{\vec{v_j}\vec{p_i}}{\|\vec{v_j}\|}$, where $\vec{v_j}$ represents the loadings vector, $\vec{p_i}$ represents the genotype score vector, and $\|\cdot\|$ represents the norm of a vector. By performing this projection, we obtain a value that reflects how well each genotype aligns with the direction of the corresponding environment. This value indicates the strength of the genotype performance in that environment. The genotype with the highest projection value is considered the best-performing genotype in that specific environment. Biplots are provided for all models for a single run. Biplots with the most similarity to the real one (obtained from the AMMI model with the uncontaminated data) in terms of scores and loadings are preferred.

While the primary focus of this work lies in the evaluation of model performance, challenges in conducting an exhaustive study need to be acknowledged, especially in the context of variance components estimation. For a more detailed exploration of this crucial aspect, readers are referred to the work of Lourenço et al. (2017), where a comprehensive investigation into variance components estimation in the Linear Mixed Model is conducted. Additionally, for a broader examination encompassing other inferential procedures such as hypothesis testing and confidence intervals, refer to the work of Hui et al. (2021).

4. Results

This section comprehends the results from the simulation study. One notes that, in the case of the simulation study, and despite a trade-off in robustness, step S3 from the RW-AMMI model (Section 2.4) was performed with the classical weighted low-rank SVD (i.e., the traditional low-rank SVD, which does not incorporate techniques to mitigate the influence of outliers or noise). This decision was prompted by occasional challenges in achieving convergence with the robust weighted low-rank SVD routine within the pre-specified maximum number of iterations, a problem that may be linked to the selection of the initial estimate $\mathbf{X}^{(0)}$ for the robust algorithm, highlighting the need for the development of a data-driven choice of $\mathbf{X}^{(0)}$ (see Section 7 for a brief discussion). Notwithstanding, robust low-rank SVD should, as proposed, be used in the RW-AMMI approach since the robust low-rank SVD method can converge with adequate initial estimates and or with the increase of the number of iterations of the algorithm. Indeed, in the numerical example ahead where biplot interpretation is assessed as well as in the real data application, step S3 of the RW-AMMI was used as defined. One would expect to see the metrics of RW-AMMI's with original step S3 at least slightly improved when compared to those reported here.

In what follows, Table 2 and Tables S2–S13 and Figures S1–S4 from the Supplementary Materials (Chapters 1–2) report the observed *MtSPEs*, *MPEVs*, and *Maxsub* values obtained across the 100 simulations referring to the uncontaminated and all contamination scenarios. Additionally, Figures S5–S8 from the Supplementary Materials (Chapter 3) present the results of the Biplot Interpretation numerical example.

The following subsections summarize the simulation findings, highlighting the strengths and weaknesses of each approach in relation to each other in terms of prediction accuracy, variability explained, and overall robustness against data contamination.

4.1. Mean trimmed squared prediction error (MtSPE)

From Tables 2 and S13, one easily acknowledges, as mentioned in Section 3.2, that the MtSPEs referring to the *shift*-outlier scenarios are greater than those observed in the *variance-inflated*- and *pointmass*-outlier scenarios. Indeed, $MtSPE_{shift} > MtSPE_{ivar} > MtSPE_{ptm}$ across all contamination scenarios and models. When the data are uncontaminated (Tables 2 and S13, 0%), as expected, all models present MtSPEs greater or equal than those reported for the AMMI model. Here, the maximum observed value of MtSPE is 0.35, which pertains to model RW-AMMI{8} and represents an error increase to that of the AMMI of almost 100%. Across the three types of data contamination, models W-AMMI{5, 9} and R-AMMI{1} performed the best across all contamination scenarios ($MtSPE_{2,5,10\%}^{shift} \leq 2$ & $MtSPE_{2,5,10\%}^{ivar} \leq 0.5$ & $MtSPE_{2,5,10\%}^{ptm} \leq 0.35$ & $MtSPE_{0\%} \leq 0.29$). If one excludes the extreme *shift* contamination scenario (i.e., $k = 10,\ 10\%$), then W-AMMI{6} is also included in this group of best-performing methods across the three types of outliers.

For the *shift* contamination type (Table 2), the best ($MtSPE_{2,5,10\%} \leq 2$ & $MtSPE_{0\%} \leq 0.29$ & $MtSPE_{0\%} \leq 0.29$) performing AMMI competing models are models W-AMMI{5,9}, R-AMMI{1,4–6,9} and RW-AMMI{5,6,9}, with all except models R-AMMI{6,9} outperforming or having a similar performance as the AMMI model across all but the 2% contamination scenarios. The worst (highest observed MtSPEs across most of the contamination scenarios) performing model here is W-AMMI{8}.

As to *pointmass* contamination (Table S13), this type of contamination does not cause a deleterious effect on the performance of the AMMI model, which outperforms all the other models. Here, models that show a similar performance to that of the AMMI are W-AMMI{1,2,4}. In addition, model RW-AMMI{8} presents the overall worst performance under this type of contamination.

The models that perform similarly or slightly better than the AMMI model in the case of *variance-inflated* type of outliers are the models W-AMMI{1,2,4}. Here, W-AMMI{8} is the worst-performing model with the highest reported values of MtSPEs in two of the three contamination scenarios.

4.2. Mean percentage of explained variability (MPEV)

Tables S2–S3 (0%) show that, with the exception of W-AMMI$\{5-9\}$ models and RW-AMMI$\{3, 5-9\}$ models, all the models with two components are able to explain more variability than the AMMI model when uncontaminated data are considered. Indeed, simulated true variability explained by the first two principal components is not 100% but rather some value close to it, which is unknown and varies across simulations.

When considering the *shift* contamination scenarios, four models stand out, which, without overestimating MPEV, produce values of $MPEV > 75\%$ across all scenarios. Specifically, models W-AMMI$\{6, 9\}$ and RW-AMMI$\{7, 8\}$ (Table S2, gray lines), with the first two models

Table 2
Mean trimmed squared prediction errors (MtSPE) for all models and all *shift* contamination scenarios.

		Shift-Outliers								
		k = 4			k = 7			k = 10		
	0%	2%	5%	10%	2%	5%	10%	2%	5%	10%
AMMI	**0.18**	0.28	0.54	1.32	0.42	1.31	3.46	0.71	2.42	6.00
W-AMMI{1}	0.18	0.29	0.54	1.27	0.42	1.16	3.46	0.64	2.16	6.89
W-AMMI{2}	0.19	**0.27**	0.53	1.35	**0.33**	1.11	3.36	0.39	1.93	6.11
W-AMMI{3}	0.19	0.44	1.05	2.23	1.01	2.00	4.20	1.79	2.77	6.05
W-AMMI{4}	0.18	0.27	0.48	1.14	0.33	0.95	3.01	0.41	1.66	5.91
W-AMMI{5}	0.24	0.31	0.48	1.02	0.33	0.58	1.48	0.34	0.63	1.80
W-AMMI{6}	0.24	0.33	0.50	1.07	0.41	0.72	1.85	0.60	1.05	2.52
W-AMMI{7}	0.24	0.33	0.55	1.22	0.37	0.83	2.09	0.39	1.02	2.69
W-AMMI{8}	0.25	0.61	1.26	2.37	1.27	2.07	3.70	2.08	2.57	4.44
W-AMMI{9}	0.24	0.31	**0.46**	**0.95**	0.34	**0.57**	1.44	0.37	0.65	1.84
R-AMMI{1}	0.24	0.32	0.52	1.12	0.33	0.61	1.79	0.32	0.64	1.96
R-AMMI{2}	0.25	0.34	0.54	1.18	0.35	0.69	2.12	0.35	0.75	2.60
R-AMMI{3}	0.26	0.34	0.55	1.16	0.36	0.68	2.02	0.35	0.75	2.43
R-AMMI{4}	0.24	0.32	0.52	1.12	0.33	0.61	1.82	0.32	0.64	1.98
R-AMMI{5}	0.24	0.32	0.52	1.13	0.33	0.62	1.86	**0.32**	0.64	1.99
R-AMMI{6}	0.29	0.38	0.57	1.14	0.38	0.61	1.61	0.36	**0.58**	1.50
R-AMMI{7}	0.30	0.39	0.59	1.19	0.39	0.64	1.74	0.37	0.62	1.68
R-AMMI{8}	0.30	0.39	0.59	1.17	0.39	0.64	1.68	0.38	0.63	1.58
R-AMMI{9}	0.29	0.38	0.57	1.14	0.38	0.61	1.62	0.36	0.58	1.50
RW-AMMI{1}	0.24	0.34	0.59	1.27	0.46	1.18	2.82	0.71	1.97	4.20
RW-AMMI{2}	0.27	0.36	0.61	1.36	0.45	1.12	2.95	0.60	1.77	4.83
RW-AMMI{3}	0.28	0.38	0.63	1.35	0.46	1.10	2.83	0.60	1.70	4.49
RW-AMMI{4}	0.24	0.34	0.58	1.24	0.44	1.13	2.75	0.67	1.90	4.04
RW-AMMI{5}	0.29	0.36	0.54	1.06	0.38	0.61	1.39	0.38	0.64	1.63
RW-AMMI{6}	0.29	0.37	0.54	1.03	0.38	0.61	1.34	0.38	0.64	1.60
RW-AMMI{7}	0.33	0.41	0.60	1.16	0.43	0.66	1.45	0.43	0.67	1.55
RW-AMMI{8}	0.35	0.43	0.62	1.20	0.44	0.68	1.49	0.45	0.71	1.70
RW-AMMI{9}	0.29	0.37	0.54	1.02	0.38	0.60	**1.29**	0.38	0.62	**1.43**

*Bold values represent the lowest observed value of MtSPE in each contamination scenario.

outperforming the last two (MPEV values always ≥ 90%). Results also show model W-AMMI{8} to be the one that greatly underestimates MPEV ($MPEV < 75\%$) across all contamination scenarios (Table S2, green line). Moreover, four models, specifically, W-AMMI{1, 4} and R-AMMI{7, 8}, overestimate MPEV ($MPEV > 100\%$) across all scenarios (Table S2, pink lines), with models W-AMMI{1, 4} overestimating more than those of R-AMMI{7, 8}. In addition, models RW-AMMI{1, 4}, besides presenting the biggest overestimation of MPEV in the extreme *shift* contamination scenario, also produce MPEV estimated values above those of the AMMI model.

Regarding the *variance-inflated* and *pointmass* contamination types (Table S3), with the exception of models W-AMMI{2 − 3, 5 − 9} and RW-AMMI{2 − 3, 5 − 9}, all the remaining models were able to keep the estimated *MPEVs* between 90–100% across all contamination scenarios while outperforming the AMMI model. Amongst these, W-AMMI{1} and R-AMMI{6 − 9} perform the best and outperform the AMMI model.

Finally, except for model W-AMMI{7}, all models showed values of $MPEV >= 75\%$, thus making W-AMMI{7} the worst performing model. When considering *pointmass* type of contamination, models R-AMMI{6 − 9} performed the best (with similar performance; $MPEV^{ivar}_{2,5,10\%} \in [95, 100]$).

4.3. Mean squared error *j* (MSE$_j$)

Tables S4–S9 display the results of the observed MSEs computed for each interaction principal component (MSE_1 and MSE_2) as well as the value of the summation of the MSEs across both interaction principal components ($MSE_1 + MSE_2$). Ideally, one would like to find a method that simultaneously provides the smallest values of MSE_1 and MSE_2. However, that is not always possible and thus one should try to find that that provides the smallest $MSE_1 + MSE_2$ instead. Although for the uncontaminated data, the highest values of MSE_1, MSE_2 and, subsequently, of $MSE_1 + MSE_2$ are those referring to models RW-AMMI{7, 8}, this does not mean, as will be seen ahead, that these models are the overall MSE worst performing AMMI competing models.

In the case of the *shift* type of contamination and considering the first interaction principal component (Table S4), the W-AMMI{9} and W-AMMI{3, 8} are, respectively, the overall best ($MSE_1^{2,5,10\%} \leq 55$ & $MSE_1^{0\%} \simeq 1.1$) and worst ($MSE_1^{2,5,10\%} \geq 100$ & $MSE_1^{0\%} \simeq 9.5$) performing models. Indeed, although the $MSE_1^{0\%}$ for the RW-AMMI{7, 8} models are greater than those of the W-AMMI{3, 8} models, the latter are the worst performing as they cannot tolerate contamination throughout ($\max MSE_1 \simeq 1853$) whereas the former can keep all errors at a substantially lower level ($\max MSE_1 \simeq 208$). In all, except for W-AMMI{9}, all methods struggled with the $10 - shift$, 10% scenario. If one disregards this scenario, then models W-AMMI{5} and RW-AMMI{5, 6, 9} come up as contenders. In fact, here, model RW-AMMI{9} is the overall best-performing model.

Results for the same type of contamination for the second interaction principal component are displayed in Table S6. In this case, the best ($MSE_2^{2,5,10\%} \leq 55$ & $MSE_2^{0\%} \leq 7.6$) and worst ($MSE_2^{2,5,10\%} \geq 45$ & $MSE_2^{0\%} \simeq 1.1$) performing AMMI competing models are, respectively, RW-AMMI{7, 8} and W-AMMI{8}. When disregarding the most extreme contamination scenario, other competing models appear again. Specifically, models W-AMMI{9} and RW-AMMI{9} are now the best overall performing models. The summation of both MSEs (Table S8) concludes in the same direction as the previous results. Namely, that models W-AMMI{9} and W-AMMI{8} are, respectively, the overall best ($MSE_1^{2,5,10\%} + MSE_2^{2,5,10\%} \leq 135$) and worst ($MSE_1^{2,5,10\%} + MSE_2^{2,5,10\%} \geq 200$) performing AMMI competing methods and that, when disregarding the most extreme contamination scenario, in particular, the RW-AMMI{9} becomes the overall best-performing model outperforming the AMMI reference model in all but the 2%, $4 - shift$ scenario.

When *variance-inflated* and *pointmass* types of contamination are considered (Tables S5, S7 and S9), the models that perform the best (tables' gray lines) across the two interaction principal components

(MSEs summation included) are the W-AMMI{4}, R-AMMI{1 − 9} and RW-AMMI{1,4}, with W-AMMI{8} performing the worst (tables' green lines). For *pointmass* contamination only, the best-performing models across the two interaction principal components and percentages of contamination, which outperform the AMMI model, are models W-AMMI{1,4} and RW-AMMI{1,4} (2% MSE_2 case excluded) with model W-AMMI{4} producing the smallest value of individual MSEs and $MSE_1 + MSE_2$ in the case of the uncontaminated data. In the case of *variance-inflated* contamination, model W-AMMI{4} is the best model outperforming the AMMI in terms of MSE_1 and $MSE_1 + MSE_2$. As to the MSE_2, no best model completely outperforms the AMMI. Here, W-AMMI{4} and R-AMMI{5} do it in two of the three contamination scenarios, respectively, 2&10% and 5&10%.

4.4. Maxsub

Tables S10–S11 show the results referring to the observed mean values of *maxsub* across all types and percentages of contamination. As expected, when uncontaminated data is considered, all the methods tend to overestimate *maxsub*. Here, models W-AMMI{1 − 9}, R-AMMI{1−5} and RW-AMMI{1,4} are able to keep these values below ≤ 0.1. As before, one sees that *shift* contamination tends to produce higher values of *maxsub* than *variance−inflated* and *pointmass* contamination.

In the case of *shift* contamination (Table S10), the best ($maxsub_{0\%} \leq 0.1$ & $maxsub_{2,5,10\%} \leq 0.15$) and worst ($maxsub_{2,5,10\%} \geq 0.2$) performing AMMI competing models are, respectively, models W-AMMI{2,7} and W-AMMI{1,3,8}. Although models R-AMMI{1 − 5}, which also present values of $maxsub_{0\%} \leq 0.1$, outperform the AMMI in the majority of the contamination scenarios (except for $k = 4$ with 2&5%), with the R-AMMI{5} appearing as the top performing method ($maxsub_{k=4;2,5\%} \leq 0.1$), these are not competitive against models W-AMMI{2,7}. When considering the scale types of contamination (Table S11), the best ($maxsub_{0,2,5,10\%} \leq 0.1$) performing AMMI competing methods across all scenarios are models W-AMMI{2,5,7} (W-AMMI{5,7} performing similarly) and R-AMMI{2,5} (similar performance), with W-AMMI{2} showing the smallest *maxsub* values. Here, the worst ($maxsub_{0,2,5,10\%} \geq 0.15$) performing AMMI competing method is model RW-AMMI{8}. Models W-AMMI{4}, R-AMMI{4} and RW-AMMI{4} are also competitive ($maxsub_{0,2,5,10\%} \leq 0.1$) in the case of the *pointmass* type of contamination. Nevertheless, none of the methods performs better than the AMMI when *pointmass* contamination is considered.

Overall, as far as *maxsub* values are concerned, models W-AMMI{2,7} are those that provide the best compromise. If thinking about model simplicity and assuming a not so extreme percentage of data contamination, then the R-AMMI5 model, which is the original R-AMMI model, is definitely a contender. Figures S2–S4 show the boxplots referring to the estimated *maxsub* values across simulations, models, and uncontaminated and contaminated scenarios. These confirm the results discussed above. In addition, the boxplots give a glimpse at the models' *maxsub* values dispersion, which is clearly and not surprisingly greater in the *shift* contamination scenarios and for higher percentages of contamination (Figure S3). Here, in particular, the R-AMMI{5} model presents higher *maxsub* dispersion than the AMMI{2,7} models, which is usually expected from robust modeling approaches. Figure S1 shows a subset of those boxplots where the level of shift contamination 4 and 10 are considered, together with the percentage of data contamination, 2%, 5%, and 10%, where similar conclusions can be taken.

4.5. Biplot interpretation

Figure S5 shows a subset of biplots obtained from the following models: AMMI (Gauch, 1992), W-AMMI{1} (Rodrigues et al., 2014), W-AMMI{5}, W-AMMI{6}, R-AMMI{1}, R-AMMI{5} (Rodrigues et al., 2016), R-AMMI{6}, RW-AMMI{1}, RW-AMMI{5} and RW-AMMI{6} (which refers to the hybrid between W-AMMI{1} and R-AMMI{5} proposed in this work). Indeed, the weighting schemes {1}, {5}, and {6} were chosen for display in Figure S5 so that biplots of the original W-, R- and RW-AMMI models can be compared. Note, however, that in the discussion above, models with weighting scheme {1} were not competitive with those of weighting schemes {5} and {6}. As expected, the scores for the genotypes and loadings for the environments are located in similar places of the biplots (note that for weighting schemes {1} and {6} the directions of the biplots are mirrored in comparison to those of the AMMI model).

In order to interpret the biplots' results for the numerical example, which considered 5% data contamination with 7-*shift* outliers, we briefly review the simulation results for this specific scenario. Specifically, within each class of models, (i) the smallest $MSE_1 + MSE_2$ observed values (Table S8), which, in the biplot interpretation, refer to the length of the loadings, are those of models W-AMMI{5,9}, R-AMMI{6,9} and RW-AMMI{6,9}, with RW-AMMI{9} showing the overall smallest value; and (ii) the smallest *maxsub* observed values (Table S10), which, in the biplot interpretation refer to the angles between the loadings, are those of models W-AMMI{2,7}, R-AMMI{2,5} and RW-AMMI{5,9}, with W-AMMI{2} showing the overall smallest value.

As to the biplots from the numerical example (Figures S6–S8), the non-robustness of the AMMI model is clear. Not only the direction of the loadings are highly impacted by contamination, but also their lengths as well as the angles between the loadings (Figure S6, first row, left- and right-hand side plots for the AMMI without and with contamination, respectively). Figure S6 also shows that, whereas the W-AMMI{3,8} models are unable to capture the correct direction of the loadings, W-AMMI{6,9} do not separate the loadings properly. Indeed here, models W-AMMI{2,5,7} seem to be those which are visually closer to the uncontaminated AMMI model results (note again that biplot directions can be mirrored). In the case of the R-AMMI biplots (Figure S7), they all capture the loadings' directions well (all but biplots referring to R-AMMI{5,6,9} are mirrored). In this case, it is very difficult to visually distinguish between the R-AMMI results as the angles between the loadings and their sizes look alike. For RW-AMMI class of models (Figure S8), those that capture the directions well (mirrored images) with visually similar loadings' lengths and angle separation between loadings, are the RW-AMMI{5,6,9} models. Overall, these results are in line with those related with the accuracy measures.

5. Real data example

The real data used to illustrate the usefulness of the methodology proposed in this paper is the Steptoe x Morex (SxM) barley mapping population, produced by the North American Barley Genome Mapping Project (Hayes et al., 1993). The original database contains 150 doubled haploid genotypes in 16 environments from 1991 to 1992 in the United States and Canada. Only trials with complete or partial replication were considered because of the need for at least one partial replicate to compute the weights for the W-AMMI models. The final data set used in this study includes 150 genotypes and eight environments. The normality and homoscedasticity of the residuals were tested with the Shapiro–Wilk and Fligner-Killeen tests, respectively. While, error normality and homoscedasticity between environments were rejected at the usual significance levels (p-values < 0.001), error homoscedasticity between genotypes was not rejected (p-value ≃ 0.937). These results are therefore consistent with data contamination.

The observed MPEV values for models with two, three, and four components were, respectively, between 48%–55%, 61%–71%, and 72%–80% with no significant differences between the different classes of models (results not shown). Despite the fact that more components explain more data variability, following Rodrigues et al. (2014) and Rodrigues et al. (2016), we opted to consider herein only two multiplicative terms in all models in order to ease result comparison.

Table 3
First best genotype, identified by each model, for each of the eight environments.

	ID91	ID92	MAN92	MIN92	MTd92	MTi91	MTi92	WA91
AMMI	DH038	DH134	DH146	DH038	DH138	DH134	DH115	DH009
W-AMMI {1}	DH012	DH112	DH091	DH091	DH076	DH081	DH076	DH091
W-AMMI {2}	DH048	DH076	DH009	DH023	DH076	DH115	DH115	DH091
W-AMMI {3}	DH126	DH081	DH091	DH023	DH076	DH054	DH076	DH091
W-AMMI {4}	DH012	DH112	DH091	DH091	DH076	DH112	DH076	DH091
W-AMMI {5}	DH038	DH134	DH146	DH038	DH115	DH115	DH115	DH057
W-AMMI {6}	DH012	DH081	DH091	DH012	DH076	DH112	DH054	DH091
W-AMMI {7}	DH137	DH137	DH113	DH137	DH115	DH113	DH115	DH057
W-AMMI {8}	DH126	DH081	DH091	DH126	DH054	DH054	DH054	DH091
W-AMMI {9}	DH012	DH081	DH091	DH012	DH076	DH112	DH054	DH091
R-AMMI {1}	DH048	DH048	DH009	DH038	DH138	DH138	DH138	DH057
R-AMMI {2}	DH048	DH138	DH009	DH048	DH115	DH138	DH115	DH009
R-AMMI {3}	DH038	DH048	DH009	DH038	DH138	DH048	DH138	DH057
R-AMMI {4}	DH048	DH048	DH009	DH038	DH138	DH138	DH138	DH009
R-AMMI {5}	DH048	DH138	DH146	DH048	DH115	DH138	DH115	DH009
R-AMMI {6}	DH048	DH048	DH009	DH048	DH115	DH138	DH138	DH009
R-AMMI {7}	DH048	DH138	DH009	DH048	DH115	DH138	DH115	DH009
R-AMMI {8}	DH038	DH048	DH009	DH038	DH115	DH048	DH138	DH057
R-AMMI {9}	DH048	DH048	DH009	DH048	DH115	DH138	DH138	DH009
RW-AMMI {1}	DH035	DH112	DH091	DH091	DH076	DH112	DH054	DH091
RW-AMMI {2}	DH054	DH054	DH102	DH081	DH125	DH113	DH125	DH081
RW-AMMI {3}	DH081	DH076	DH091	DH099	DH129	DH076	DH076	DH099
RW-AMMI {4}	DH012	DH112	DH091	DH091	DH076	DH112	DH076	DH091
RW-AMMI {5}	DH038	DH134	DH146	DH038	DH115	DH115	DH115	DH057
RW-AMMI {6}	DH012	DH112	DH091	DH091	DH076	DH112	DH112	DH091
RW-AMMI {7}	DH081	DH134	DH091	DH081	DH017	DH134	DH017	DH091
RW-AMMI {8}	DH014	DH099	DH129	DH023	DH076	DH112	DH112	DH014
RW-AMMI {9}	DH012	DH081	DH091	DH091	DH076	DH081	DH112	DH091

Besides Table 3, all Tables and Figures referring to this discussion are found in the Supplementary Materials, Chapter 4. Similarly to Figure S5, Figure S9 shows a subset of biplots obtained from the following models: AMMI (Gauch, 1992), W-AMMI{1}, W-AMMI{5}, W-AMMI{6}, R-AMMI{1}, R-AMMI{5}, R-AMMI{6}, RW-AMMI{1}, RW-AMMI{5} and RW-AMMI{6}. As reported in Rodrigues et al. (2014) and to Rodrigues et al. (2016), it is evident that the estimated loadings and scores differ between models (e.g., models W-AMMI and RW-AMMI against model R-AMMI) and within the same model class (e.g., models W-AMMI{1,6} and RW-AMMI{1,6} against, respectively, models W-AMMI{5}, and RW-AMMI{5}). In addition, most of the AMMI competing methods also show clear biplot differences in the angles and loadings when compared to the AMMI biplot results. Such differences were not evident in the numerical example where no contamination was present (Figure S5). These findings provide additional support to the hypothesis that the Steptoe x Morex data may be contaminated.

Figures S10–S12 show the biplots for the AMMI model and the three competing classes of models (W-AMMI, R-AMMI and RW-AMMI, respectively) across all weighting schemes. Again here, the differences between AMMI biplot results and those of the competing methods are apparent.

Tables 3 & S14–S16 show the first (Tables 3 & S14), second (Table S15), and third (Table S16) best-performing genotypes, respectively, for each environment, which are recommended by each model. We can see that between the R-AMMI models there is an overall agreement to select the first, second and third best-performing genotypes for each of the environments. However, this pattern is less consistent in the W-AMMI and RW-AMMI models, with only a few instances showing agreement.

The integration of the proposed generalized weighting schemes in the W-AMMI and R-AMMI models as well as in the proposed RW-AMMI model, enables the breeder to analyze the data without worrying about outlier contamination. When working with real data, it is often challenging, if not impossible, to determine the true nature of the underlying signal, let alone the percentage of contamination and the specific type of contamination. As a result, it becomes extremely difficult to recommend a particular model for fitting and analyzing such data. However, a viable approach to address this issue would be to leverage ensemble strategies, which involve combining the results of multiple models. By doing so, it is possible to generate more accurate genotype recommendations, effectively mitigating the uncertainties associated with the data's true nature. Following the above reasoning, Figures S13 and S14 show the percentage of models that recommended a given genotype as one of the three best-performing genotypes for each of the eight environments. These figures show the top 10 recommended genotypes, in the order of preference, that we would suggest to the breeder.

While our analysis provides detailed rankings, future research could explore methodologies for summarizing genotype agreements across multiple methods, potentially through visual or comparative metrics that emphasize consensus in rankings. Such developments would enhance the understanding of genotype selection dynamics throughout the breeding pipeline.

6. Discussion

This work advances the AMMI model by combining weighting and robust statistical techniques. The proposed hybrid AMMI model (RW-AMMI) and the additional generalized weighting schemes for W-AMMI, R-AMMI, and RW-AMMI models enhance the AMMI modeling approach in tackling both data heteroscedasticity and data contamination across multiple contamination scenarios. These improvements are crucial for the accuracy and reliability of analyses in plant breeding, environmental studies, and other fields where such data issues are common (Rodrigues et al., 2014, 2016; Hadasch et al., 2018; da Silva et al., 2019)

In terms of model performance, although no single model emerged as the definitive best in the simulations, several models within each class of W-AMMI, R-AMMI, and RW-AMMI demonstrated results close to the benchmark set by the AMMI model using uncontaminated data. Specifically, models W-AMMI{2,5,6,9}, R-AMMI{1-6,9}, and RW-AMMI{5,6,9} performed notably well across various comparative metrics and types of contamination. The consistent performance of weighting schemes 5, 6, and 9 across all competing AMMI models highlights their significance. In particular, these schemes, which employ robust weights, enhance the robustness of the W-AMMI{5-9} models, further strengthening their performance under different contamination scenarios.

However, the competitiveness of the RW-AMMI model relative to W-AMMI and R-AMMI counterparts is less clear in this study. This is likely due to the replacement of the robust SVD with the classical method in step S3 of the RW-AMMI procedure, due to convergence issues of the

former, which limits the utility of the robust approach in this context. Furthermore, the consideration of only two replicates in the simulation study prevents the full exploitation of the median-based approach in step S2, which could have provided more robustness in the model.

Future investigations should include scenarios with $r \geq 3$ replicated data and focus on developing a data-driven approach for determining adequate initial estimates for the robust low-rank SVD decomposition algorithm. This would help mitigate convergence challenges and improve the performance of the RW-AMMI model. This approach is especially important given the need for an increased number of iterations to achieve convergence, which raises concerns about computational efficiency. In simulation studies involving a large number of runs, the computational burden can become substantial, although for simpler data analyses, these issues may be less pronounced.

Contamination can significantly impact the AMMI model in genotype-by-environment studies, leading to inaccuracies in loadings' directions, lengths, and angles (Croux and Filzmoser, 1998; Rodrigues et al., 2016; Tanaka, 2020). Such inaccuracies can hinder genotype selection, resulting in the potential loss of time and resources. To overcome these challenges, it is essential to develop and implement competing models alongside the AMMI technique. The comparison of results from different models aids in identifying data problems and encourages critical thinking regarding data modeling and result interpretation. Nevertheless, in practical applications involving real data, determining the true nature of the underlying signal and the extent of contamination remains a challenge (Tanaka, 2020). Ensemble strategies, as suggested in the real data example analysis, that combine results from multiple models can provide more accurate genotype recommendations and reduce uncertainties associated with the true nature of the data. As the field of genotype-by-environment modeling and plant breeding continues to evolve, the hybrid approaches discussed here will likely become increasingly important in achieving reliable and robust results when addressing challenges such as data contamination and heteroscedasticity.

7. Conclusion

In conclusion, this study contributes to the AMMI modeling framework by introducing the hybrid AMMI model (RW-AMMI) and a set of enhanced weighting schemes that effectively address challenges such as data contamination and heteroscedasticity. The comparative performance of several models demonstrates that careful selection of weighting schemes can significantly improve the robustness and reliability of AMMI results. While the RW-AMMI model shows promise, future work should focus on refining its implementation, particularly with respect to computational efficiency and convergence challenges, to enhance its practical applicability. Additionally, exploring alternative weighting schemes and data-driven methods for selecting initial estimates for the robust singular value decomposition algorithm could further optimize the model's performance. By addressing these issues, future research can extend the utility of this hybrid framework for genotype selection under various environmental conditions, providing breeders with a more robust tool for improving both stability and adaptability in crop improvement programs.

CRediT authorship contribution statement

Marcelo B. Fonsêca: Writing – review & editing, Software, Methodology, Investigation. **Vanda M. Lourenço:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Paulo C. Rodrigues:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Ethics approval and consent to participate

Not applicable.

Funding

PCR acknowledges financial support from the Brazilian National Council for Scientific and Technological Development (CNPq), grant "Bolsa de produtividade PQ-2" 309359/2022-8, from the Federal University of Bahia, and CAPES-PRINT-UFBA, under the topic "Modelos Matemáticos, Estatísticos e Computacionais Aplicados às Ciências da Natureza". VML acknowledges financial support from national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications), and project 2023.14934.PEX (https://doi.org/10.54499/2023.14934.PEX; Center for Mathematics and Applications).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103110.

Data availability

Web supplementary Tables and Figures are available with this paper. The R codes, simulated datasets, and real crop data used in this study are available from https://github.com/marcelofonsecab/hybridRWAMMI. For guidance on reproducing the results and applying the methods to new datasets, we recommend consulting the README file, which provides detailed instructions and documentation. For applying the methods to new datasets, the real data example analysis is particularly helpful.

References

Arciniegas-Alarcón, S., et al., 2020. New multiple imputation methods for genotype-by-environment data that combine singular value decomposition and jackknife resampling or weighting schemes. Comput. Electron. Agric. 176, 105617. http://dx.doi.org/10.1016/j.compag.2020.105617.

Bernardo Júnior, L.A.Y., et al., 2018. Ammi bayesian models to study stability and adaptability in maize. Agron. J. 110 (5), 1765–1776. http://dx.doi.org/10.2134/agronj2017.11.0668.

Crossa, J., 1990. Statistical analyses of multilocation trials. Adv. Agron. 44, 55–85. http://dx.doi.org/10.1016/S0065-2113(08)60818-4.

Croux, C., Filzmoser, P., 1998. A robust biplot representation of two-way tables. In: Advances in Data Science and Classification: Proceedings of the 6th Conference of the International Federation of Classification Societies (IFCS-98). pp. 355–361. http://dx.doi.org/10.1007/978-3-642-72253-0_48.

Dia, M., et al., 2016. Analysis of Genotype × Environment Interaction (G×E) Using SAS Programming. Agron. J. 108, 1838–1852. http://dx.doi.org/10.2134/agronj2016.02.0085.

Ebdon Jr., J.S., Gauch, H.G., 2002. Additive main effect and multiplicative interaction analysis of national turfgrass performance trials: I, interpretation of genotype 3 environment interaction. Crop. Sci. 42, 489–496. http://dx.doi.org/10.2135/cropsci2002.0489.

Eberhart, S.A., Russel, W.A., 1966. Stability parameters for comparing varieties. Crop. Sci. 6, 36–40. http://dx.doi.org/10.2135/cropsci1966.0011183X000600010011x.

Estaghvirou, S.B.O., et al., 2014. Influence of outliers on accuracy estimation in genomic prediction in plant breeding G3: Genes, genomes. Genetics 4 (12), 2317–2328. http://dx.doi.org/10.1534/g3.114.011957.

Filzmoser, P., Todorov, V., 2013. Robust tools for the imperfect world. Inform. Sci. 245, 4–20. http://dx.doi.org/10.1016/j.ins.2012.10.017.

Finlay, K.W., Wilkinson, O.N., 1963. The analysis of adaptation in a plant breeding programme. Aust. J. Agric. Res. 14, 742–754. http://dx.doi.org/10.1071/AR9630742.

Forkman, J., et al., 2022. Testing components of two-way interaction in multi-environment trials. Comm. Statist. Theory Methods 1–20. http://dx.doi.org/10.1080/03610926.2022.2108058.

Gauch, H.G., 1988. Model selection and validation for yield trials with interaction. Biomtrics 44, 705–715. http://dx.doi.org/10.2307/2531585.

Gauch, H.G., 1992. Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs. Elsevier Science Publishers.

Gauch, H.G., Zobel, R.W., 1990. Imputing missing yield trial data. Theor. Appl. Genet. 79, 753–761. http://dx.doi.org/10.1007/BF00224240.

Gauch, H.G., Zobel, R.W., 1997. Identifying mega-environments and targeting genotypes. Crop. Sci. 37, 311–326. http://dx.doi.org/10.2135/cropsci1997.0011183X003700020002x.

Gauch Jr., H.G., et al., 2011. Two new strategies for detecting and understanding qtl×environment interactions. Crop. Sci. 51 (1), 96–113. http://dx.doi.org/10.2135/cropsci2010.04.0206.

Gower, J.C., et al., 2011. Understanding Biplots. John Wiley & Sons, http://dx.doi.org/10.1002/9780470973196.

Gumedze, F.N., et al., 2010. A variance shift model for detection of outliers in the linear mixed model. Comput. Statist. Data Anal. 54 (9), 2128–2144. http://dx.doi.org/10.1016/j.csda.2010.03.019.

Hadasch, S., et al., 2018. Weighted estimation of ammi and gge models. J. Agric. Biol. Environ. Stat. 23 (2), 255–275. http://dx.doi.org/10.1007/s13253-018-0323-z.

Hawkins, D.M., et al., 2001. Robust Singular Value Decomposition. vol. 122, National Institute of Statistical Science Technical Report.

Hayes, P.M., et al., 1993. Quantitative trait locus effects and environmental interaction in a sample of North-American barley germ plasm. Theor. Appl. Genet. 87, 392–401. http://dx.doi.org/10.1007/BF01184929.

Huber, P.J., 1964. Robust estimation of a location parameter. Ann. Math. Stat. 35, 73–101. http://dx.doi.org/10.1214/aoms/1177703732.

Huber, P.J., 1973. Robust regression: Asymptotics, conjectures and Monte Carlo. Ann. Stat. 1, 799–821. http://dx.doi.org/10.1214/aos/1176342503.

Hubert, M., et al., 2002. A fast method for robust principal components with applications to chemometrics. Chemom. Intell. Lab. Syst. 60 (1–2), 101–111. http://dx.doi.org/10.1016/S0169-7439(01)00188-5.

Hubert, M., et al., 2005. Robpca: a new approach to robust principal component analysis. Technometrics 47 (1), 64–79. http://dx.doi.org/10.1198/004017004000000563.

Hui, F.K., et al., 2021. Random effects misspecification can have severe consequences for random effects inference in linear mixed models. Int. Stat. Rev. 89 (1), 186–206. http://dx.doi.org/10.1111/insr.12378.

Kang, M.S., 2020. Genotype-environment interaction and stability analyses: an update. In: Quantitative Genetics, Genomics and Plant Breeding. CABI Wallingford UK, pp. 140–161. http://dx.doi.org/10.1079/9781789240214.0140.

Koller, M., 2013. Robust Estimation of Linear Mixed Models (Ph.D. thesis). ETH Zurich.

Koller, M., 2016. Robustlmm: an r package for robust estimation of linear mixed-effects models. J. Stat. Softw. 75, 1–24. http://dx.doi.org/10.18637/jss.v075.i06.

Krzanowski, W., 1979. Between-groups comparison of principal components. J. Amer. Statist. Assoc. 74 (367), 703–707. http://dx.doi.org/10.2307/2286995.

Lourenço, V.M., Pires, A.M., 2014. M-regression, false discovery rates and outlier detection with application to genetic association studies. Comput. Statist. Data Anal. 78, 33–42. http://dx.doi.org/10.1016/j.csda.2014.03.019.

Lourenço, V.M., et al., 2011. Robust linear regression methods in association studies. Bioinformatics 27, 815–821. http://dx.doi.org/10.1093/bioinformatics/btr006.

Lourenço, V.M., et al., 2017. A robust df-reml framework for variance components estimation in genetic studies. Bioinformatics 33 (22), 3584–3594. http://dx.doi.org/10.1093/bioinformatics/btx457.

Lourenço, V.M., et al., 2020. Robust estimation of heritability and predictive accuracy in plant breeding: evaluation using simulation and empirical data. BMC Genomics 21 (1), 1–18. http://dx.doi.org/10.1186/s12864-019-6429-z.

Malik, W.A., et al., 2019. Testing multiplicative terms in ammi and gge models for multienvironment trials with replicates. Theor. Appl. Genet. 132, 2087–2096. http://dx.doi.org/10.1007/s00122-019-03339-8.

Paderewski, J. others, 2011. Yield response of winter wheat to agro-ecological conditions using additive main effects and multiplicative interaction and cluster analysis. Crop. Sci. 51 (3), 969–980. http://dx.doi.org/10.2135/cropsci2010.05.0278.

Paderewski, J., 2013. An r function for imputation of missing cells in two-way data sets by em-ammi algorithm. Commun. Biometry Crop. Sci. 8, 60–69.

Paderewski, J., Rodrigues, P.C., 2014. The usefulness of em-ammi to study the influence of missing data pattern and application to polish post-registration winter wheat data. Aust. J. Crop. Sci. 8 (4).

Paderewski, J., Rodrigues, P.C., 2018. Constrained ammi model: Application to polish winter wheat post-registration data. Crop. Sci. 58 (4), 1458–1469. http://dx.doi.org/10.2135/cropsci2017.06.0347.

Piepho, H.-P., et al., 2012. Blup for phenotypic selection in plant breeding and variety testing. Euphytica 161, 209–228.

Podani, J., et al., 2021. Principal component analysis of incomplete data–a simple solution to an old problem. Ecol. Informatics 61, 101235. http://dx.doi.org/10.1016/j.ecoinf.2021.101235.

Rodrigues, P.C., 2018. An overview of statistical methods to detect and understand genotype-by-environment interaction and qtl-by-environment interaction. Biom. Lett. 55 (2), 123–138. http://dx.doi.org/10.2478/bile-2018-0009.

Rodrigues, P.C., et al., 2014. A weighted ammi algorithm to study genotype-by-environment interaction and qtl-by-environment interaction. Crop. Sci. 54, 1555–1570. http://dx.doi.org/10.2135/cropsci2013.07.0462.

Rodrigues, P.C., et al., 2016. A robust additive main effects and multiplicative interaction model for the analysis of genotype-by-environment data. Bioinformatics 32, 58–66. http://dx.doi.org/10.1093/bioinformatics/btv533.

Rodrigues, P.C., et al., 2021. An analysis of simulated yield data for pepper shows how genotype× environment interaction in yield can be understood in terms of yield components and their qtls. Crop. Sci. 61 (3), 1826–1842. http://dx.doi.org/10.1002/csc2.20476.

Schützenmeister, A. others, 2012. Checking normality and homoscedasticity in the general linear model using diagnostic plots. Comm. Statist. Simulation Comput. 41 (2), 141–154. http://dx.doi.org/10.1080/03610918.2011.582560.

da Silva, C.P., et al., 2019. Heterogeneity of variances in the bayesian ammi model for multienvironment trial studies. Crop. Sci. 59 (6), 2455–2472. http://dx.doi.org/10.2135/cropsci2018.10.0641.

Srebro, N., Jaakkola, T., 2003. Weighted low-rank approximations. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03). pp. 720–727.

Tanaka, E., 2020. Simple outlier detection for a multi-environmental field trial. Biometrics 76 (4), 1374–1382. http://dx.doi.org/10.1111/biom.13216.

van der Westhuizen, S., et al., 2024. Biplots for understanding machine learning predictions in digital soil mapping. Ecol. Informatics 84, 102892. http://dx.doi.org/10.1016/j.ecoinf.2024.102892.

Yates, F., Cochran, W., 1938. The analysis of groups of experiments. J. Agric. Sci. 28, 556–580. http://dx.doi.org/10.1017/S0021859600050978.