Published online September 19, 2019

Mean Performance and Stability in Multi-Environment Trials I: Combining Features of AMMI and BLUP Techniques

Tiago Olivoto,* Alessandro D.C. Lúcio, José A.G. da Silva, Volmir S. Marchioro, Velci Q. de Souza, and Evandro Jost

ABSTRACT

Additive main effect and multiplicative interaction (AMMI) and best linear unbiased prediction (BLUP) are popular methods for analyzing multi-environment trials (MET). The AMMI has nice graphical tools for modeling genotype-vs.-environment interaction (GEI) but fails in some aspects, such as accommodating a linear mixed-effect model (LMM) structure. The BLUP provides reliable estimates but new insights to deal graphically with a random GEI structure are needed. This article compares the predictive success of BLUP and AMMI, shows how to generate biplots for modeling GEI in MET analysis using LMM, and proposes a new quantitative genotypic stability measure called WAASB, which is the **W**eighted **A**verage of **A**bsolute **S**cores from the singular value decomposition of the matrix of BLUPs for the GEI effects generated by an LMM. We also introduced the theoretical basis of a superiority index that allows weighting between mean performance and stability, which was conveniently called WAASBY. The **B**LUP was found to outperform AMMI in the analysis of four real MET trials. The application of our indexes is illustrated using an oat (*Avena sativa* L.) MET dataset. It was shown that reliable measures of stability using WAASB may help breeders and agronomists to make correct decisions when selecting or recommending genotypes. In addition, the simultaneous selection index, WAASBY, will be useful when the selection should consider different weights for stability and mean performance. Some advantages over existing statistics are discussed. Finally, the implementation of the procedures of this article in future studies is facilitated by an R package containing all required functions.

Core Ideas

- The predictive accuracy of BLUP and AMMI was investigated using four real datasets.
- BLUP was found to outperform AMMI in all datasets analyzed.
- A genotypic stability index that inherits the principles of AMMI and BLUP was proposed.
- A superiority index that allows weighting between mean performance and stability was proposed.
- An R package with useful functions for MET analysis is presented.

Published in Agron. J. 111:1–12 (2019)
doi:10.2134/agronj2019.03.0220
Supplemental material available online

Breeders and geneticists continually strive to increase crop productivity to meet the growing world demand for food. In the final stage of a plant breeding program, much effort and resources need to be invested in the evaluation of the genotypes. Generally, a few hundred genotypes are investigated in a large number of environments, resulting in the well-known MET. The MET makes it possible to identify genotypes that display a small temporal variability-which is desired by and is beneficial to growers–, or cultivars that are consistent from location to location-which is desired by and is beneficial to seed companies and breeders (Yan and Kang, 2003).

The data from these trials result in a matrix $\boldsymbol{M}$ of dimension genotype × environment. Since plants respond to a host of environmental signals (both biotic and abiotic), a given genotype can perform relatively well in a given environment, but relatively poorly in others. If the genotypes' ranking changes significantly across environments, a significant GEI is observed. This interaction is known as qualitative or crossover type interaction and plays a key role in formulating strategies for crop improvement. Another form of interaction may also occur from just expansion or contraction of scale over the range of environments without a change in rank order. This form is known as quantitative or non-crossover type. At this step, one of the main challenges of MET analysis

T. Olivoto, A.D.C Lúcio, Dep. of Plant Science, Fed. Univ. of Santa Maria, Santa Maria, RS, Brazil; J.A.G. da Silva, Dep. of Agricultural Studies, Regional Univ. of Nortwest Rio Grande do Sul, Ijuí, RS, Brazil; V.S. Marchioro, Dep. of Agronomic and Environmental Science, Fed. Univ. of Santa Maria, Frederico Westphalen, RS, Brazil; V.Q. de Souza, Fed. Univ. of Pampa, São Gabriel, RS, Brazil; E. Jost, Federal Institute of Education, Science and Technology of Farroupilha, São Vicente do Sul, RS, Brazil. Received 24 Mar. 2019. Accepted 19 July 2019. *Corresponding author (tiagoolivoto@gmail.com).

Abbreviations: AMMI, additive main effects and multiplicative interaction; ASV, additive main effects and multiplicative interaction stability value; AVRC, index and the ranks of the mean yields; BLUP, best linear unbiased prediction; EV, averages of the squared eigenvector values; GEI, genotype × environment interaction; HMGV, harmonic mean of genotypic values; HMRPGV, harmonic mean of relative performance of genotypic values; IPCA, interaction principal component axis; LMM, linear mixed-effect model; MET, multi-environment trials; NF, no fungicide; RCBD, randomized complete block design; RMSPD, root mean square prediction difference; SPIC, sums of the absolute value of the IPCA scores; SVD, singular value decomposition; WAASB, weighted average of absolute scores from the singular value decomposition of the matrix of best linear unbiased predictions for the genotype × environment interaction effects generated by an linear mixed-effect model; WAASBY, weighted average of weighted average of absolute scores from the singular value decomposition of the matrix of best linear unbiased predictions for the genotype × environment interaction effects generated by an linear mixed-effect model and response variable; WF, with fungicide; Za, absolute value of the relative contribution of interaction principal component axes to the interaction.

arises: to understand the GEI, seeking new ways of exploiting it and using it to benefit the selection of highly productive genotypes targeted to specific environments or that are broadly adapted.

The breeders' desire to modeling these interactions appropriately has led to the development of procedures called stability analyses. Yates and Cochran (1938) suggested a multiplicative operator for modeling the GEI that consists of a simple regression of a genotype's performance on the environmental mean-known as joint regression analysis-and that was based from ideas presented by Mooers (1921). Years later, Finlay and Wilkinson (1963) and Eberhart and Russell (1966) popularized this approach. Methods that combine different statistical techniques have been also developed for stability analysis. For example, Gollob (1968) proposed a method which combines the features of factor analysis and analysis of variance into a single method. At that time, this method was known as FANOVA. Nowadays, this same method is known by the acronym AMMI (Gauch, 1988).

The AMMI analysis is used mainly in a fixed-effect model framework. In some cases, it may be reasonable considering genotypes or environments (or both) to be random effects (Smith et al., 2005). When one factor is fixed and others random, we have a LMM. More specifically, the BLUP offers the potential to improve the predictive accuracy of random effects (Smith et al., 2005). A study comparing the predictive success of BLUP and AMMI suggested that BLUP should be used to obtain reliable estimates in MET (Piepho, 1994). Since the 1990s, LMM has been more frequently used to analyze MET. Between 2013 and 2015, for example, the larger number of papers proposing methods to deal with GEI were related to LMM (van Eeuwijk et al., 2016).

From the practical point of view, BLUP and AMMI can be seen as two distinct approaches to achieve the same goal: to distinguish the GEI pattern from the random error. From the statistical point of view, these models are vastly different. The AMMI analysis retains most of the GEI pattern in the first interaction principal component axis (IPCA) resulting from the singular value decomposition (SVD) of the nonadditive effects matrix, while most of the random error is retained in the last IPCAs. The BLUP, on the other hand, initially estimates the effects of the ANOVA model and then attributes weights to these effects and could thus be considered a shrinkage estimator (Piepho 1994). These two models are frequently used alone in the evaluation of METs. For example, some studies were successful in estimating genotypic values in MET using BLUP (Olivoto et al., 2017; Nardino et al., 2016) while others were successful in modeling GEI patterns using AMMI (Bocianowski et al., 2019; Veenstra et al., 2019). Taking into account the importance of AMMI and BLUP, the question that becomes apparent at this point is if the benefits of these two important techniques could be incorporated into a single method. Thus, there would seem to be value in an investigation to combine the graphical tools of AMMI and the predictive accuracy of BLUP.

Our hypothesis in this study is that the shrunken GEI effects matrix generated by a BLUP-based mixed model can be subjected to an AMMI-like analysis using SVD procedure. Thus, the purposes of this study were to: (i) evaluate the predictive ability of AMMI and BLUP using real data with different GEI patterns; (ii) introduce a genotypic stability measure and a superiority index that allow weighting between performance and stability; (iii) compare these measures with worldwide-known parametric and nonparametric indexes in terms of genotype ranking; and (iv) introduce an R package that includes user-friendly functions for MET analysis, and for implementing the procedures proposed in this study.

MATERIAL AND METHODS

Basic Concepts of AMMI and BLUP

Consider a set of multi-environment trials where g genotypes are tested in each of e environments. For convenience, we will consider that in each environment the genotypes are arranged in a randomized complete block design (RCBD) with b replications. The simplest linear model with interaction effect used in the statistical analysis of this data is given in Eq. [1].

$$y_{ijk}=\mu+\alpha_i+\tau_j+(\alpha\tau)_{ij}+\gamma_{jk}+\varepsilon_{ijk} \quad [1]$$

where y_{ijk} is the response variable (e.g., grain yield) of the kth block of the ith genotype in the jth environment ($i = 1, 2,..., g$; $j = 1, 2,..., e$; $k = 1, 2,..., b$); μ is the grand mean; α_i is the main effect of the ith genotype; τ_j is the main effect of the jth environment; $(\alpha\tau)_{ij}$ is the interaction effect of the ith genotype with the jth environment; γ_{jk} is the effect of the kth block within the jth environment; and ε_{ijk} is the random error assuming $\varepsilon_{ijk} \sim \text{NID}(0,\sigma^2)$, where NID means normally, identically and independently distributed.

For cases in which a complex GEI structure is observed, a more accurate estimate of y_{ij} can be obtained using AMMI analysis (Gauch, 1988). Generally, the mean response of individual genotypes averaged over b replications within each environment is computed and used to fill a $g \times e$ matrix. Briefly, the estimate of y_{ij} is given in two steps according to Eq. [2].

$$y_{ij} = \mu + \alpha_i + \tau_j + \sum_{k=1}^{p} \lambda_k a_{ik} t_{jk} + \rho_{ij} + \varepsilon_{ij} \quad [2]$$

First, the additive effects of genotype (α_i) and environment (τ_j) are fitted by standard ANOVA procedures; then, the nonadditive-or residual-effects matrix is decomposed as $y_{ij} - \mu - \alpha_i - \tau_j = \sum_{k=1}^{p} \lambda_k a_{ik} t_{jk}$, where λ_k is the singular value for kth IPCA; a_{ik} is the ith genotype eigenvector for axis k; t_{jk} is the jth environment eigenvector for axis k. A residual, ρ_{ij}, remains if not all the p IPCA are used, that is, $p = \min(g - 1; e - 1)$. The scores for genotypes and environments are then used in biplots (Kempton, 1984), allowing a graphical interpretation of the GEI effects.

In MET trials, modeling genotypic effects as random may be preferable despite the fact that it would be classified as fixed using traditional definitions (Stroup and Mulitze, 1991). To illustrate the methodology, we will assume α_i and $(\alpha\tau)_{ij}$ to be random effects; thus the model in Eq. [1] can be conveniently rewritten in a standard linear mixed model (Yang, 2007):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \quad [3]$$

where $\mathbf{y}$ is an $n[=\sum_{j=1}^{e}(gb)]\times 1$ vector of response variable (e.g., grain yield) $\mathbf{y} = \left[y_{111}, y_{112}, \ldots, y_{geb}\right]'$; $\mathbf{b}$ is an $(eb)\times 1$ vector of unknown and unobservable fixed effects $\mathbf{b} = [\mu + \gamma_{11}, \gamma_{12}, \ldots, \gamma_{eb}]'$; $\mathbf{u}$ is an $m[= g + ge]\times 1$ vector of unobservable random effects $\mathbf{u} = \left[\alpha_1, \alpha_2, \ldots, \alpha_g, (\alpha\tau)_{11}, (\alpha\tau)_{12}, \ldots, (\alpha\tau)_{ge}\right]'$; $\mathbf{X}$ is an $n \times (eb)$ design matrix of 0s and 1s relating $\mathbf{y}$ to $\mathbf{b}$; $\mathbf{Z}$ is an $n \times m$ design

matrix of 0s and 1s relating **y** to **u**; **e** is an n × 1 vector of random errors $\mathbf{e} = \left[y_{111}, y_{112}, \ldots, y_{geb} \right]'$; and the prime (′) represents the vector transposition. Random vectors **u** and **e** are assumed to be normal and independently distributed with zero mean and variance-covariance matrices **G** and **R** respectively, such that

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

The matrices **G** and **R** are allowed to take different covariance structure (Piepho, 1998), but here we will take a simple form of **G** and **R** for a comparison with the AMMI model:

$$\mathbf{G} = \begin{bmatrix} \hat{\sigma}^2_{\alpha} \mathbf{I}_g & 0 \\ 0 & \hat{\sigma}^2_{\alpha\tau} \mathbf{I}_{ge} \end{bmatrix}$$

and $\mathbf{R} = \hat{\sigma}^2_e \mathbf{I}_n$, where $\hat{\sigma}^2_{\alpha}$, $\hat{\sigma}^2_{\alpha\tau}$, and $\hat{\sigma}^2_e$ represent variances for genotype, genotype-vs.-environment interaction and random errors, respectively; $\mathbf{I}_g$, $\mathbf{I}_{ge}$, and $\mathbf{I}_n$ are the identity matrices of order g, $g \times e$, and n, respectively.

The vectors **b** and **u** are then estimated using the well-known mixed model equation (Henderson, 1975).

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad [4]$$

where the superscript "−1" and "−" represent the inverse and generalized inverse of the matrices, respectively. Generally, variance components are usually unknown and thus their estimates ($\hat{}$ and $\hat{\mathbf{R}}$), often obtained by REstricted Maximum Likelihood (REML) using the Expectation-Maximization algorithm (Dempster et al., 1977) are substituted into Eq. [4]. The significance of the random effects may be tested by a likelihood ratio (LR) test, which for the model in Eq. [4], compares the −2(Res) log likelihoods for two models, one with all random terms (full model) and other without one of the random terms (reduced model). The probability is then obtained by a two-tailed chi-square test with one degree of freedom (χ^2_1).

Considering balanced data, the effect the ith genotype ($\hat{g}_i$) within $\mathbf{u}_\mathbf{g}$ is given in standard ANOVA notation as follows:

$$\hat{g}_i = h^2_g(\bar{y}_{i.} - \bar{y}_{..}) \quad [5]$$

where $h^2_g = \left(\hat{\sigma}^2_{\alpha\tau} + e\hat{\sigma}^2_{\alpha}\right) / \left(\hat{\sigma}^2_{\alpha\tau} + \hat{\sigma}^2_{\varepsilon} + e\hat{\sigma}^2_{\alpha}\right)$ is the shrinkage effect for the genotype effect. The BLUP of the ith genotype is then given by $BLUP_i = \mu + \hat{g}_i$. The effect of the ith genotype in the jth environment ($\hat{g}_{ij}$) within $\mathbf{u}_\mathbf{ge}$ is given as follows:

$$\hat{g}_{ij} = h^2_g(\bar{y}_{i.} - \bar{y}_{..}) + h^2_{ge}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \quad [6]$$

where h^2_g is as defined above and $h^2_{ge} = \hat{\sigma}^2_{\alpha\tau} / \left(\hat{\sigma}^2_{\alpha\tau} + \hat{\sigma}^2_{\varepsilon}\right)$ is the shrinkage effect for GEI. The BLUP of the ith genotype in the jth environment is then given by $\text{BLUP}_{ij} = \bar{y}_{.j} + \hat{g}_{ij}$. It is easy to see that the smaller the error term, the smaller the shrinkage effect, becoming one (no shrinkage effect) if error term is zero.

Source and Characterization of Data

The data used in this study comes from experiments with oat, soybean (*Glycine max*. L.), wheat (*Triticum aestivum* L.), and maize (*Zea mays* L.). Different datasets were used for covering different GEI patterns, providing more security in the selection of the most predictively accurate model. A total of 118 genotypes and 47 environments were studied. The oat dataset came from trials conducted at the Department of Agrarian Studies of the Regional University of Northwestern Rio Grande do Sul, in Augusto Pestana, RS, Brazil. Ten oat genotypes conducted in 16 environments were evaluated. The environments were defined by the combinations of eight growing seasons (2010–2017) and two fungicide managements (with fungicide [WF] and without fungicide [NF]). The characterization of the 10 oat cultivars is in Supplemental Table S1. The soybean dataset came from trials conducted at Federal Institute of Education, Science, and Technology of Farroupilha, in São Vicente do Sul, RS, Brazil. A total of 13 soybean genotypes were evaluated in five environments. Each environment was considered the combination of cultivation years (2013 and 2014) and sowing seasons (three in 2013 and two in 2014). The wheat dataset came from 17 trials performed in 2014 with 40 wheat genotypes from the wheat breeding program the Central Cooperative for Agricultural Research (COODETEC). Further details can be found in Bornhofen et al. (2018). The maize dataset came from trials with 55 maize genotypes growing in nine environments. This dataset was used by da Silva et al. (2015) and is publicly available in the online version of such article https://doi.org/10.1371/journal.pone.0131414. For all datasets, the genotypes in each environment were organized in a randomized complete block design (RCBD) with three blocks and one replication per block. The measured response variable was the grain yield (GY, Mg ha^{-1}).

Statistical Analysis

All statistical analyses were performed using R 3.5.2 software (R Core Team 2018). The functions used in this article were organized in an R package called metan-multi-environment trial analysis (Olivoto, 2019) that will be detailed later in this article.

Cross-validation Procedure

To evaluate the predictive accuracy of the AMMI and BLUP models, a cross-validation procedure was performed according to Piepho (1994). The original data was randomly split into training set-two complete and randomly selected blocks per environment–, and validation set-the remaining block per environment. Depending on the experiment, n AMMI models (AMMI0, AMMI1,..., AMMIF) were fitted to the modeling data according to Eq. [2]. The validation using BLUP considered the same steps and was based on Eq. [3]. Even though these 10 cultivars constitute a relatively small set of cultivars, they represent the majority of the area cultivated with oat in southern Brazil. Thus, it is reasonable to assume that they constitute a random sample of a population. For all models (AMMIn and BLUP) the predictive success was compared in relative terms using the root mean square prediction difference (RMSPD) between the model estimates and validation data, as follows:

$$\text{RMSPD} = \left[\left(\sum_{i=1}^{n} \left(\hat{y}_{ij} - y_{ij} \right)^2 \right) / n \right]^{0.5} \quad [7]$$

where $\hat{y}_{ij}$ is the model predicted value of the ith genotype in the jth environment and y_{ij} is the observed value of the ith genotype in the jth environment in the validation set. For all models (AMMI-model family and BLUP), this procedure was repeated 1000 times. A boxplot was used to show the distribution of the 1000 RMSPD of each model. The codes used in this section are in S3.1.

Combining the Advantages of AMMI and BLUP

As discussed earlier, only the *Oat* dataset will be used in this section. This data is publicly available (Olivoto, 2018) and will be used to reproduce all the examples. In the traditional AMMI model usage, a matrix with the residual of the additive model is decomposed into *k* IPCAs using SVD. Let $\mathbf{A_{ge}}$ be the matrix of BLUPs for the GEI effects generated by an LMM [$h^2_{ge}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$ or $\hat{\mathbf{u}}_{\mathbf{ge}}$]. Similar to the AMMI analysis, $\mathbf{A_{ge}}$ was decomposed as follows

$$\mathbf{A_{ge}} = \mathbf{U_{gp}}\mathbf{\Lambda_{pp}}\mathbf{V^T_{ep}} \quad [8]$$

where $\mathbf{\Lambda_{pp}}$ is a diagonal matrix containing *p* singular values, in decreasing order, where **p** is the rank of the matrix $\mathbf{A_{ge}}$ with $\mathbf{p} \leq \min(g - 1; e - 1)$. The matrices $\mathbf{U_{gp}}$ and $\mathbf{V_{ep}}$ are orthonormal matrices with singular vectors of $\mathbf{AA^T}$ and $\mathbf{A^TA}$, respectively, being the orthonormal basis for the genotypes and environment effects, respectively.

The coordinate points for genotypes ($\mathbf{G_{gp}}$) and environments ($\mathbf{E_{ep}}$) in the *p* possible interaction axes were then estimated as $\mathbf{G_{gp}} = \mathbf{U_{gp}}\mathbf{\Lambda^{0.5}_{pp}}$, and $\mathbf{E_{ep}} = \mathbf{V_{ep}}\mathbf{\Lambda^{0.5}_{pp}}$. The rows of the submatrices $\mathbf{G_{g2}}$ and $\mathbf{E_{e2}}$ were the points for the coordinates of genotypes and environments, respectively, and were used in an AMMI2-like biplot to model the GEI patterns. Considering that genotypes or environments have random effects-thus characterizing a mixed model-the GEI will always be random, which allows the estimation of $\mathbf{A_{ge}}$ and, consequently, the estimation of scores for genotypes and environments.

Aiming at identifying possible mega-environments as well as visualizing the "which-won-where" pattern a graphic with the nominal yield ($\hat{y}^*_{ij}$) as a function of the environment IPCA1 scores ($\mathbf{E_{e1}}$) was produced. In this graphic, each genotype is depicted by a straight line with the equation $\hat{y}^*_{ij} = \mu_i + \text{IPCA1}_i \times \text{IPCA1}_j$, where $\hat{y}^*_{ij}$ is the nominal yield for the *i*th genotype in the *j*th environment; μ_i is the grand mean of the *i*th genotype; IPC1_i is the IPCA1 score of the *i*th genotype and IPC1_j is the IPCA1 score of the *j*th environment. The winner genotype in a given environment has the highest nominal yield in that environment (Gauch and Zobel, 1997).

The Genotypic Stability Index

In the traditional AMMI model usage, when the proportion of the variance explained in IPCA1 is relatively low, there may be a biased interpretation regarding the stability of the genotypes using the AMMI1 biplot since GEI patterns are still explained in the remaining IPCA axis. To handle this problem, we propose a new stability index called WAASB, that is the **W**eighted **A**verage of **A**bsolute **S**cores from the singular value decomposition of the matrix of **B**LUPs for the GEI effects generated by an LMM, estimated as follows:

$$\text{WAASB}_i = \sum_{k=1}^{p} \left|\text{IPCA}_{ik} \times \text{EP}_k\right| \Big/ \sum_{k=1}^{p} \text{EP}_k \quad [9]$$

where WAASB_i is the weighted average of absolute scores of the *i*th genotype (or environment); IPCA_{ik} is the score of the *i*th genotype (or environment) in the *k*th IPCA, and EP_k is the amount of the variance explained by the *k*th IPCA. The genotype with the lowest WAASB value is considered the most stable, that is, the one that deviates least from the average performance across environments. Aiming at identifying highly productive and stable genotypes, we propose swapping the well-known AMMI1 biplot by a biplot with the abscissa represented by the WAASB values and the ordinate by the response variable. This biplot has the advantage of using all the estimated IPCA axes to identify the stability in a bi-dimensional plot.

To identify whether and how the ranks of genotype are altered when different numbers of IPCA are used in the WAASB estimation, the genotype's ranks were obtained considering the WAASB estimated with 1, 2,..., *p* IPCA. When using only one IPCA, WAASB = |IPCA1|. The ranking was increasing; so, the genotype with the smallest WAASB value had the first-order rank. A heatmap graph was used to show the ranks of the genotypes in the different scenarios of WAASB estimation. The codes used in the two last sections are in S3.5.

A Superiority Index that Allows Weighting between Performance and Stability

To select genotypes that combine high performance and stability we introduced the WAASBY index, which is a superiority index that allows weighting between performance (in our example, GY) and stability (WAASB index). The first step is rescaling both GY and WAASB to 0 to 100 so that they can be directly compared. Since the best values for GY is the maximum value and for WAASB is the lowest value, the transformations were performed according to the following equations:

$$rG_i = \frac{100 - 0}{G_{\max} - G_{\min}} \times (G_i - G_{\max}) + 100 \quad [10]$$

and

$$rW_i = \frac{0 - 100}{W_{\max} - W_{\min}} \times (W_i - W_{\max}) + 0 \quad [11]$$

where rG_i and rW_i are the rescaled values for GY and WAASB, respectively, for the *i*th genotype; G_i and W_i are the response variable (GY) and the WAASB values for *i*th genotype. Then the WAASBY index was calculated according to Eq. [12]:

$$\text{WAASBY}_i = \frac{(rG_i \times \theta_Y) + (rW_i \times \theta_S)}{\theta_Y + \theta_S} \quad [12]$$

where WAASBY_i is the superiority index for the *i*th genotype that weights between performance and stability, and θ_Y and θ_S are the weights for response variable and stability assumed to be 65 and 35 in this study, respectively. In addition, 21 scenarios varying θ_Y and θ_S (100/0, 95/5, 90/10,..., 0/100) were planned. For each scenario, the first-order rank was attributed to the genotype with the highest WAASBY value. The objective here is to show how the ranking of genotypes is altered depending on the weight assigned to the stability and response variable. To assist with intuitive interpretation, a heat map graph was produced. The codes used in this section are in S3.6.

Relationship between Stability Measures

In this section the indexes WAAS and WAASY (considering a fixed-effect model), and the indexes WAASB and WAASBY (considering a mixed-effect model) were compared in terms of genotypes' ranking with the following five AMMI derived stability indexes, namely: (i) absolute values of the first principal component axis, $\text{IPCA1}_i = \sum_{k=1}^{1} \left|\lambda_k^{0.5} a_{ik}\right|$; (ii) AMMI stability value (Purchase et

al., 2000), $ASV_i = \left[\left[b\lambda_1^2 / b\lambda_2^2 \times (\lambda_1^{0.5} a_{i1})\right]^2 + (\lambda_2^{0.5} a_{i2})^2\right]^{0.5}$, where b is the number of blocks; (iii) sums of the absolute value of the IPCA scores, $SIPC_i = \sum_{k=1}^{P} \left|\lambda_k^{0.5} a_{ik}\right|$; and (iv) averages of the squared eigenvector values, $EV_i = \sum_{k=1}^{P} a_{ik}^2 / P$, described by Sneller et al. (1997), where P is the number of IPCA retained via F tests; and (v) the absolute value of the relative contribution of IPCAs to the interaction (Zali et al., 2012), $Za_i = \sum_{k=1}^{P} \theta_k a_{ik}$, where θ_k is the percentage sum of squares explained by the kth IPCA. We also considered the simultaneous selection indexes (ssi) computed by summation of the ranks of the ASV, SIPC, EV, and Za indexes with the ranks of the mean yields (Farshadfar, 2008) which resulted in ssiASV, ssiSIPC, ssiEV, and ssiZa, respectively. The genotypes' rankings were calculated according to the concept of each index and for GY for the *Oat* and *Maize* datasets. The ranks obtained for the *Maize* dataset were submitted to a PCA analysis; then, a loading plot was used to explore the relationships among the studied indexes. It should be emphasized that we used the *Maize* data in this procedure due to the larger number of subjects (55 genotypes).

RESULTS

Predictive Success

The BLUP was the most predictively accurate model in all evaluated datasets (Fig. 1). For the AMMI model, the number of IPCAs retained depended on the evaluated experiment. For *Maize*, *Oat*, *Soybean*, and *Wheat* datasets, the most predictively accurate AMMI models were AMM0, AMMI5, AMMI3, and AMMI5, respectively. This result confirms the hypothesis of different GEI patterns in the analyzed trials.

Overall Performance, Variance Components, and Predicted Means

The LR test indicated highly significant effects ($p < 0.001$) for both genotype and interaction effects in the *Oat* trial (Table 1). The Supplemental Fig. S1 shows that the interaction was qualitative (i.e., crossover type), since the rank order for the genotypes changed across environments. The grand mean of GY was 2.69 Mg ha^{-1}, where the lowest mean was 1.37 Mg ha^{-1} (2014 NF) and the highest mean was 4.06 Mg ha^{-1} (2012 WF). It was noticed that, except for 2010, the use of fungicide seemed to provide higher GY. This was most evident in the years 2013 and 2015 (Supplemental Fig. S2 and S3). The individual analysis revealed that nine of 16 environments (~56.3%) showed significant differences for genotype effects. The block effect was significant for only six (37.5%) environments (Supplemental Table S2).

Proximally 54% of the phenotypic variance ($\hat{\sigma}_p^2$) was due to the residual variance ($\hat{\sigma}_\varepsilon^2$) (Table 1). The contribution of the genotypic variance ($\hat{\sigma}_\alpha^2$) was 12.6% only. Consequently, low estimates of broad-sense heritability were observed. The genotypic accuracy of selection (As), which measures the correlation between predicted and observed values was of 0.89. The genotypic CV (5.8%) was found to be less than half of the residual CV (11.9%). In addition, the high $\hat{\sigma}_{\alpha\tau}^2 / \hat{\sigma}_\alpha^2$ ratio (2.64) resulted in a low correlation between genotypic values across environments (0.38).

The G8 and G3 genotypes stood out for having the highest predicted means among the tested genotypes (Fig. 2). The genotypes G7, G2, and G4 were the other genotypes that performed

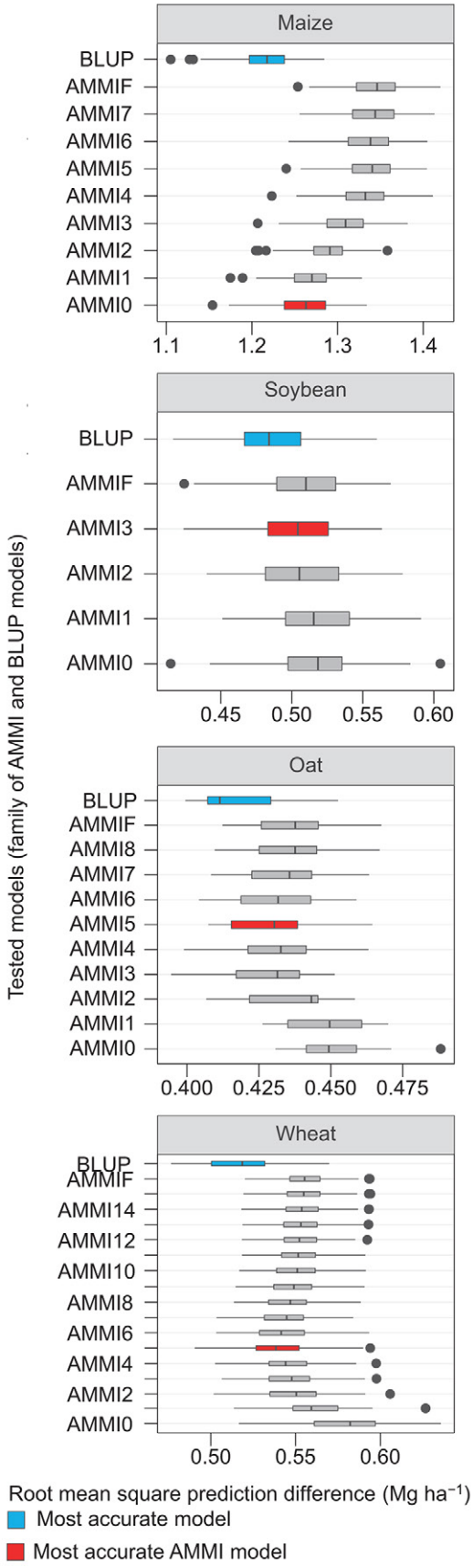

Fig. 1. Predictive accuracy of the additive main effects and multiplicative interaction (AMMI) family and best linear unbiased prediction (BLUP) for trials with four different crops. The boxplots show the distribution of the 1000 root mean square prediction difference (RMSPD) estimates.

Table 1. Deviance analysis, estimated variance components and genetic parameters for grain yield of 10 oat genotypes evaluated in 16 environments.

Statistics	Likelihood ratio test†	
	G	GEI
χ^2	19.23	61.10
p-value	1.15×10^{-5}	5.42×10^{-15}
REML‡	Variance components	
	Estimates	
$\hat{\sigma}^2_{\alpha}$	0.0241 (12.6%)§	
$\hat{\sigma}^2_{\alpha\tau}$	0.0637 (33.4%)	
σ^2_{ε}	0.1032 (54.0%)	
σ^2_{p}	0.191	
h^2_{g}	0.126	
R^2_{gei}	0.333	
h^2_{mg}	0.797	
As	0.892	
r_{ge}	0.381	
CV_g, %	5.8	
CV_e, %	11.9	
CV_g/CV_r ratio	0.483	
$\hat{\sigma}^2_{\alpha\tau}/\hat{\sigma}^2_{\alpha}$ ratio	2.643	

† G, genotype; GEI, genotype-vs-environment interaction.

‡ $\hat{\sigma}^2_{\alpha}$, genotypic variance; $\hat{\sigma}^2_{\alpha\tau}$, variance of G × E interaction; , residual variance; σ^2_{p}, phenotypic variance; h^2_{g}, broad-sense heritability; R^2_{gei}, coefficient of determination for the genotype-vs-environment interaction effects; h^2_{mg}, heritability of the genotypic mean; As, accuracy of genotype selection; r_{ge}, correlation between genotypic values across environments; CV_g (%), genotypic coefficient of variation; CV_e (%), residual coefficient of variation.

§ Parenthetical values indicate the percentage of the observed phenotypic variance (σ^2_{p}).

well (GY above the grand mean), however, with a very small difference among them.

Understanding the Genotype × Environment Interaction

Biplot Interpretation

The cumulative variance in the first two IPCA of the oat trial was 59.3% (Supplemental Table S3). From the eight cultivation years, three (2011, 2015, and 2016) had a positive correlation–since the angle among them was <90°. This suggests that the magnitude of the interaction effects tended to be the same independently on the fungicide application (Fig. 3). Negative correlations–indicated by vector angles >90°–were observed in the years 2012, 2013, 2014, and 2017.

Figure 4 allows an easy interpretation of the "which-won-where" pattern. In our example, G8 won in all studied environments. This genotype is depicted by a line with the equation $y = 3.039 + (-0.020x)$, where x is the environmental IPCA1 score. The left-most score of –0.637 implies a yield of 3.051 Mg ha^{-1}, whereas the right-most score of 0.356 implies a yield of 3.046 Mg ha^{-1}. These two coordinate pairs give two points that define the line for G8. Considering the first IPCA, G8 won in all environments because of its highest yield (3.039 Mg ha^{-1}) and the smallest IPCA1 score (which defines the slope of the

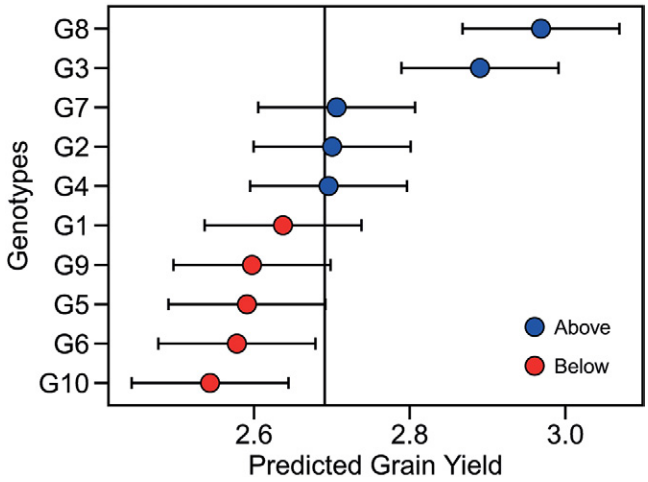

Fig. 2. Predicted grain yield (best linear unbiased prediction, BLUP) for 10 oat genotypes. Blue and red circles represent the genotypes that had BLUP above and below of BLUP means, respectively. Horizontal error bars represent the 95% confidence interval of prediction considering a two-tailed *t* test.

line) among the tested genotypes (Supplemental Table S4). At this point, we may make a relationship between Fig. 2 and 4. For example, G8 had the highest predicted mean (Fig. 2) and due to the smallest IPCA1 score, this genotype was classified as the "universal winner" (Fig. 4). On the other hand, G2 and G7 had very similar predicted means (Fig. 2) but completely distinct lines in Fig. 4. This has occurred because while G2 had an IPCA1 score of 0.063, G7 had a score of 0.669 (Supplemental Table S4). It will be discussed later in this article that these genotypes compose different groups when considering both stability and productivity for the genotypes' ranking.

The quadrants in Fig. 5 represent the four classes of genotypes/environment for a joint interpretation of performance and stability. In the first quadrant, the most unstable genotypes–the ones that contribute much to GEI–and environments with high discrimination ability are included. The magnitude of the response variable (i.e., GY), however, is below the grand mean. Oat genotypes G10 and G9 were included in this quadrant. Although they presented GY close to the grand mean, they presented the highest WAASB values. Thus, specific adaptation (Fig. 3 and 4) should be investigated for genotypes within this quadrant. We may see here how the low percentage of GEI pattern explained in IPCA1 may mask the interpretation of a biplot. Genotype G7 had a higher score (in absolute values) in IPCA1 than genotype G9 (Supplemental Table S4). However, when all IPCAs were included, G9 had a higher WAASB value than G7 (Fig. 5).

In the second quadrant, highly productive but unstable genotypes are included. The environments included in this quadrant deserve attention since, in addition to providing high magnitudes of the response variable, they present a good discrimination ability of the genotypes. In 2012 and 2016, regardless of the fungicide management, the GY was higher than the grand mean; however, the discrimination ability of the genotypes was higher in environments without fungicide application (Fig. 5).

In the third quadrant low-productive and wide-adapted genotypes are included due to the lower values of WAASB. The lower this value, the more stable is the performance of a genotype across the environments. The environments included in

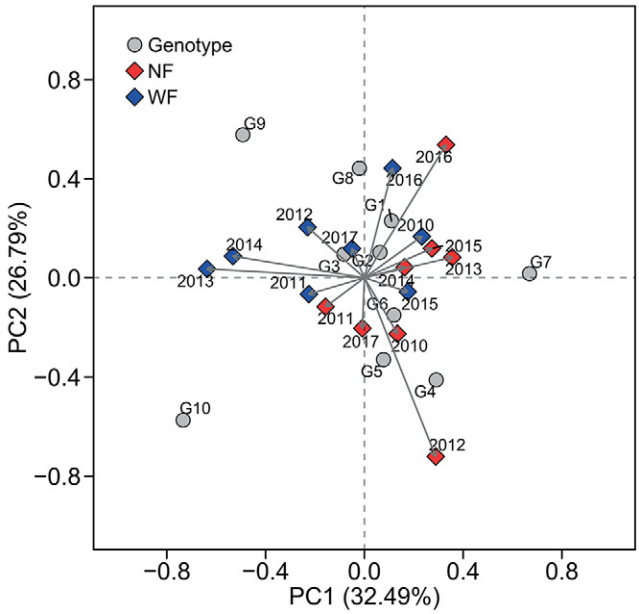

Fig. 3. Biplot of 10 oat genotypes evaluated in 16 environments (combinations of eight cultivation years with application [WF] and with no application of fungicide [NF]). The scores were obtained from fitting the singular value decomposition of the double-centered best linear unbiased prediction (BLUP) interaction effects matrix obtained in a linear mixed model with symmetric singular value partitioning (α = 1/2). The axes are equally scaled.

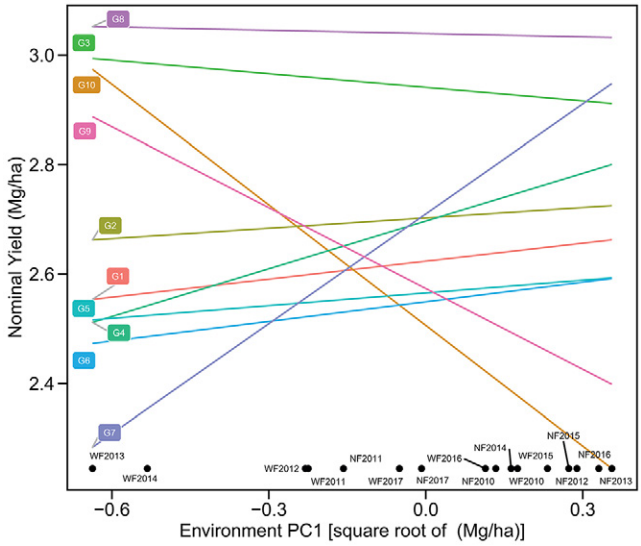

Fig. 4. Nominal grain yield for 10 oat genotypes as a function of the environment scores of the first interaction principal component axis (IPCA1).

this quadrant can be considered poorly productive and with low discrimination abilities.

The genotypes within fourth quadrant have above-mean productivity and lower values of WAASB (broadly adapted). The environments included in this quadrant, however, can be considered productive but with low discrimination abilities. We have shown previously that G8 had the smallest IPCA1, being thus the most stable when only the first IPCA is used. In our example, 67.5% of the GEI variance was not explained by IPCA1. When this information was considered (WAASB), we may see that G3 was, in fact, the most stable (smaller WAASB value).

Genotype Ranking Depending on the Number of Retained Interaction Principal Component Axis

Figure 6 shows the ranks of the genotypes in relation to stability depending on the number of IPCA used in WAASB estimation. For *Oat* data, nine axes were considered [min(10 – 1; 16 – 1)]. It is observed that the genotype ranking was altered by the extent to which IPCAs are included in the WAASB estimation. This was most evident up to three IPCA (Fig. 6). Groups of genotypes with similar stability performance may be easily identified by the dendrogram on the left side of Fig. 6. For example, the G3, G1, and G6 genotypes showed the lowest WAASB values considering four or more IPCA and were, therefore, the first, second, and third most stable, respectively (as it can be seen also in Fig. 5). The most evident change was those of the G2. When using the first and second IPCA in the WAASB estimation, that genotype was considered the second and first most stable, respectively; when more than three IPCAs were used, G2 was the sixth most stable (Fig. 6). This reinforces the benefits of using the WAASB index since it captures the variations of all IPCAs to compute the stability. If the ASV would be considered, the G2, in this case, would be ranked as most stable (Supplementary Table S5), when in fact it is not.

Genotype Ranking Depending on the Weights for Stability and Performance

Figure 7 shows the WAASBY values considering the weights for GY and WAASB equal to 65 and 35, respectively. The genotypes that had the highest values of WAASBY were G8 (92.63) and G3 (88.01). It was previously discussed that these genotypes deserved prominence because they were within the quadrant IV of Fig. 5.

Unlike Fig. 7, which shows the WAASBY values considering a fixed WAASB/GY ratio, Fig. 8 allows identifying how the ranks of the genotypes are changed depending on the weights assigned. The ranks in the left-most side are those obtained when only the stability was considered; from left to right, the weight for the response variable increase 5% each scenario. The ranks highlighted by a black rectangle are the same from those shown in Fig. 7. The ranks shown in the right-most side matches perfectly with the genotype's ranking for GY.

The clusters shown on the left side of Fig. 8 may be also used to identify groups of genotypes with similar performance regarding stability and productivity. Cluster 1 included the genotypes G3 and G8, which as previously discussed are highly productive and broadly adapted genotypes. Note that these genotypes remained the firsts-ranked regardless of the WAASB/GY ratio (Fig. 8). Cluster 2 included genotypes G2, G4, and G7, that can be considered productive, but unstable, as they were well ranked when the WAASB/GY ratio was low (greater weight for productivity). Cluster 3, conversely, included G1, G5, and G6, stable but low-productive genotypes because they were well ranked when the WAASB/GY ratio was high (greater weight for stability). Cluster 4 included G9 and G10, which we have shown in Fig. 5 are poorly productive and unstable genotypes.

Correspondence among the Stability and Simultaneous Selection Indexes

The ranks obtained for each index for both oat and maize data are shown in Supplemental Tables S5 and S6, respectively.

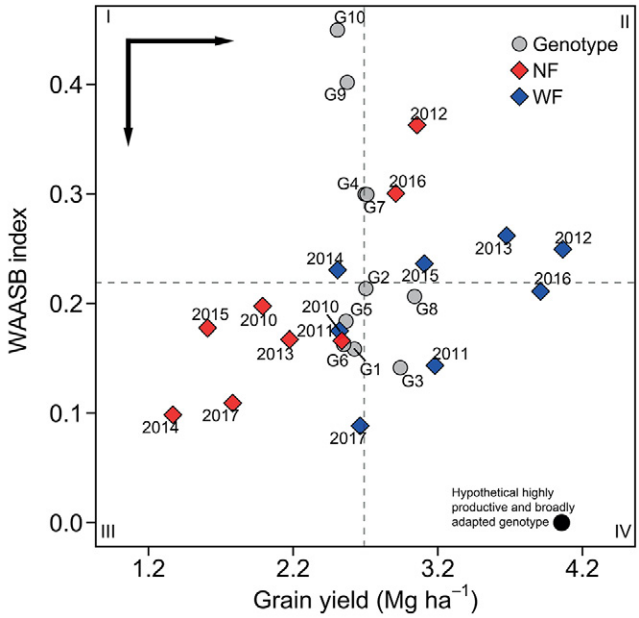

Fig. 5. Biplot of the grain yield vs. weighted average of absolute scores for the best linear unbiased prediction s of the genotype-vs.-environment interaction (WAASB) of 10 oat genotypes evaluated in 16 environments (combinations of eight cultivation years with application [WF], and with no application of fungicide [NF]). A hypothetical highly productive and broadly adapted genotype is depicted by a black circle. Horizontal and vertical black arrows indicate the direction of the increase in yielding and stability, respectively.

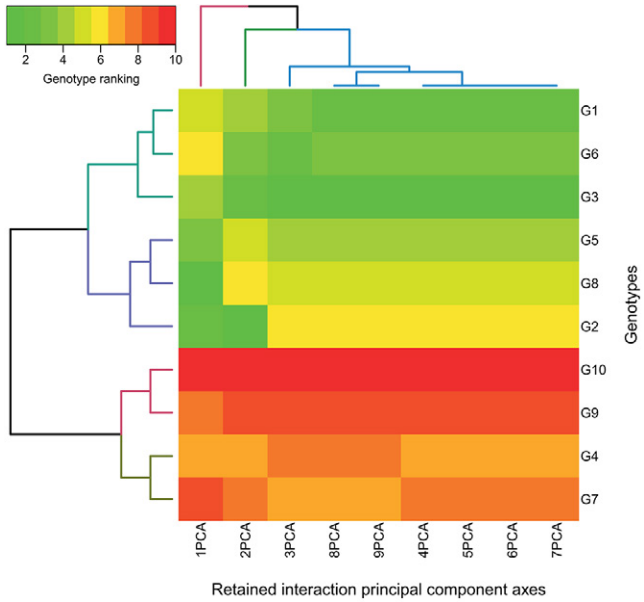

Fig. 6. Heatmap showing the ranks of 10 oat genotypes in relation to the number of interaction principal component axes (IPCA) used in the weighted average of absolute scores for the best linear unbiased predictions (BLUPs) of the genotype-vs.-environment interaction (WAASB) estimation.

Figure 9 shows the loading plot obtained in the PCA analysis considering the correlation matrix among the methods obtained from with maize data. The explained variance in the first two axes was 87.7%. Except for the IPCA1, ASV, and ssiASV, all of the other indexes were very near the edge of the circle, indicating that they were well represented by the plane of factors. All indexes had positive loadings on PCA1, and a clear separation between stability and simultaneous selection indexes was observed. Stability indexes included in the shading ellipse 1 were positively correlated with each other-the correlation between two indexes may be approximated by the cosine of the angle between its vectors-and they had positive loadings on PCA2. Our stability indexes WAAS and WAASB were highly correlated with Za, and in lower magnitude with ASV, EV, and SPIC indexes. The absence of perfect association between WAAS and WAASB suggests that beyond the relative differences, the rank order of some genotypes regarding the stability was changed when using a fixed- or a mixed-effect model (Supplemental Table S5, Fig. 9). Superiority indexes had negative loadings on PCA2 and were visually grouped into two clusters. The "ssis" indexes in shading ellipse 2 are based on AMMI principles, while the WAASY and WAASBY that inherit principles from both AMMI and BLUP formed group 3. WAASY and WAASBY were highly positively correlated and provided ranks more similar to the GY compared to the simultaneous selection indexes within shading ellipse 2. This makes sense since in this study we attributed a high weight to productivity (65%). This freedom of attributing weights for stability and mean performance in genotype ranking may facilitate the genotype selection in cases when the research wants to prioritize one of these characteristics.

DISCUSSION

Successful Statistical Analysis of Multi-Environment Trials

Multi-environment trials constitute the major efforts in plant breeding programs. Thus, the accuracy of prediction, that is, how close the predicted value is to the observed value is crucial for a successful selection, recommendation of cultivars, and delineation of mega-environments. According to Gauch and Zobel (1988), there are three main options to increase the accuracy of prediction in MET: The first, is to improve the experimental techniques, including the use of plots of ideal size and shape, the correct arrangement of plots in the experimental area, and uniform application of cultural management; The second, is to increase the number of replications, using sophisticated experimental designs; The third and last is to use statistical models with better prediction abilities. A special focus on the latter option was given in this article.

Studies comparing AMMI, BLUP, and GGE methods have been conducted (Balestre et al., 2009; Sa'diyah and Hadi 2016). In the present study, we have shown how the main advantages of AMMI and BLUP may be combined to increase the reliability of MET analysis. From the point of view of an agronomist, the biplot interpretation of the shrunken GEI effects (Fig. 3) and the "which-won-where" view (Fig. 4) may facilitate the recommendation of genotypes targeted to specific environments, thus exploiting narrow adaptations. This is important because in most cases no one genotype wins everywhere and always. On the other hand, the biplot WAASB × GY (Fig. 5) may be used for a joint interpretation of stability and productivity, thus exploiting broad adaptations. The main advantage of this biplot over the well-known AMMI1 biplot is that all IPCA axes are used, thus allowing that GEI patterns not retained in IPCA1 be considered in the genotypes' ranking.

From the point of view of a breeder, beyond the aforementioned advantages, the mixed model approach also allows the estimation

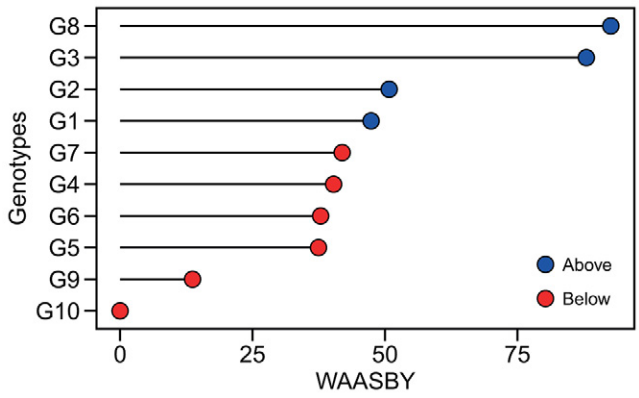

Fig. 7. Estimated values of weighted average of the stability (WAASB) and mean performance (Y) (WAASBY) for 10 oat genotypes considering the weights of 65 and 35 for yielding and stability, respectively.

of important parameters in quantitative genetics, such as the genotypic and interaction variances, broad-sense heritability, heritability on the mean basis and genetic correlations (Table 1). These pieces of information are essential in a plant breeding program and should be also exploited in MET's evaluation.

BLUP or AMMI? The Assessment will Show which Model is Better in a Given Situation

From four datasets with different GEI patterns, it is concluded that BLUP was the most predictively accurate model (Fig. 1). Our findings are according to those by Piepho (1994), who concluded that the BLUP outperforms any member of the AMMI family in predicting Faba bean (*Vicia faba* L.) yield in MET. This viewpoint, however, is not unanimous. A study evaluating rice (*Oryza sativa* L.) has shown that the estimates using the AMMI10 model was closer to the "true" value of yielding than the prediction by BLUP (Sa'diyah and Hadi 2016). Although it is not explicit, a cross-validation procedure similar to that used in this study and by Piepho (1994) was not used in such study.

Gauch (2013) pointed out that predictive accuracy merits special attention for model diagnosis in MET analysis. Due to the great data processing power of the current computers, it is reasonable to affirm that the choice of the best method to predict yield (or other response variables) should be based on the predictive ability assessment in each situation. This is due to the fact that both BLUP and AMMI have their efficacy increased depending on factors intrinsic to each trial. For example, the larger the sample size-both regarding the number of levels of each factor and the number of observations at each of these levels-the more accurate the estimates of variance components are (Smith 1978). Unbiasedness in variance component estimation results in biased repeatability coefficients. In other words, the shrinkage effect associated with random effects can be over- or underestimated. Likewise, the efficiency of the AMMI model increases with the size of trials and the increase of noisiness present in the data (Gauch and Zobel 1988).

Provided that replicated data are available, the *cv_blup()* and *cv_ammif()* functions available in the **metan** package (S3.1) may be used to evaluate the predictive ability of BLUP and AMMI methods, respectively. Thus, for each specific situation, it is easy to identify the model with the smallest bias (in relative terms) and to use it in the prediction.

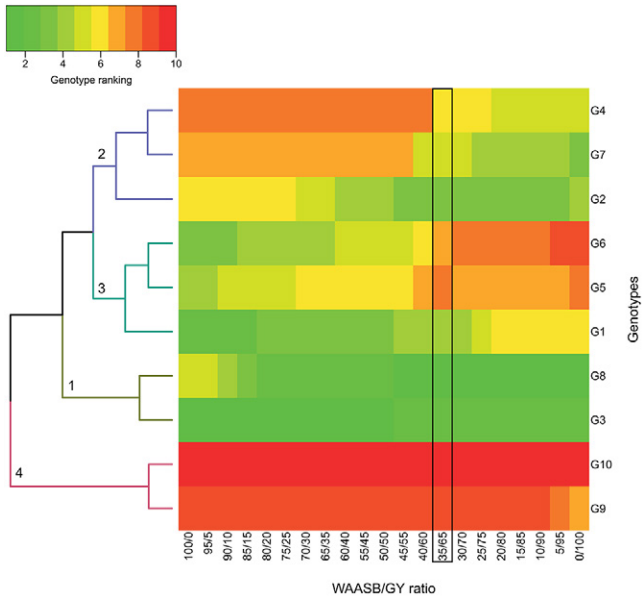

Fig. 8. Ranks of 10 oat genotypes considering different weights for stability and yielding. The most-left ranks were obtained considering the stability only. The most right-ranks were obtained considering the grain yield only. Between the extremes, the ranks were obtained different weights for stability and yielding. The four clusters represent four classes of genotypes: (1) Poorly productive and unstable genotypes; (2) productive but unstable genotypes; (3) stable but poorly productive genotypes; and (4), highly productive and stable genotypes.

Identifying Highly Productive and Broadly Adapted Genotypes with WAASB

The quantitative stability measure proposed in this study (WAASB) was found to be an important statistical tool for identifying highly productive and broadly adapted genotypes. WAASB may be seen as a mixed-effect model version of the AMMI-based stability indexes discussed here but have some advantages: (i) similar to ASV, the WAASB is a function of both cultivar and environment GEI pattern components but it is based on a mixed-effect model-shown here as to outperform fixed-effect models in terms of predictive accuracy–or even a random model; (ii) WAASB is based on absolute deviations instead squared deviations as ASV, thus, some robustness is gained due to less sensitivity to outliers; (iii) WAASB is more realistic in quantifying the stability in complex GEI structures since it is computed considering all the estimated IPCA; (iv) differently from the SIPC, which considers the sum of absolute values of the IPCA scores, the WAASB considers the weighted average of the IPCA scores; thus, more reliable results should be the obtained since high scores in the last axes will have a smaller contribution to the estimation; (v) the WAASB × GY biplot (Fig. 5) allows the joint interpretation of stability and productivity in a bi-dimensional plot considering all the IPCAs of the model.

The AMMI1 biplot is used worldwide, and some AMMI-based stability indexes, such as ASV, have recently met with some success in quantifying the stability (Adjebeng-Danquah et al., 2017; Shahriari et al., 2018). But would it be a good decision quantifying the stability using one-or even two–IPCAs? If we look back to Fig. 6 the response will probably be: No. We have shown that G2 was considered the most stable genotype when the WAASB was estimated with two

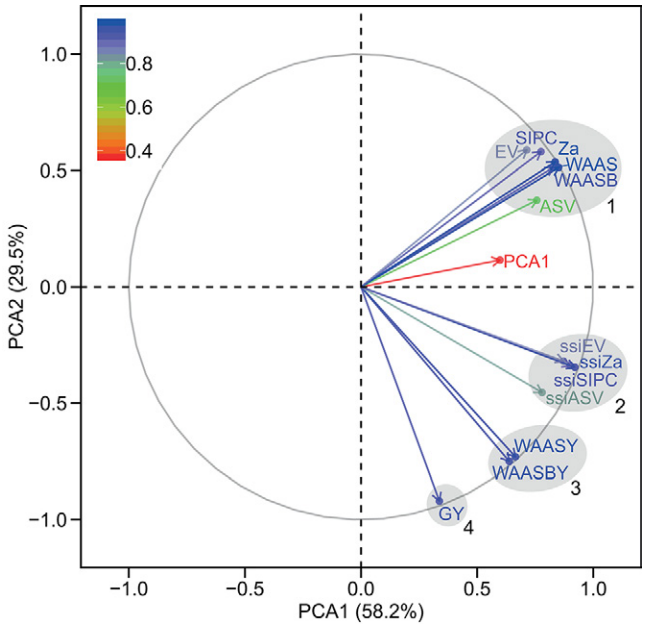

Fig. 9. Loading plot obtained in the principal component analysis with the ranks for genotypes obtained for the weighted average of absolute scores (WAAS) and weighted average of absolute scores for the best linear unbiased predictions (BLUPs) of the genotype-vs.-environment interaction (WAASB); IPCA1, absolute values of the first principal component axis; ASV, additive main effects and multiplicative interaction stability value; SIPC, sums of the absolute value of the interaction principal component axis scores; EV, averages of the squared eigenvector values; Za, absolute value of the relative contribution of interaction principal component axes to the interaction. WAASY and WAASBY are the simultaneous selection indexes using WAAS and WAASB, respectively. The "ssi" are the simultaneous selection indexes using additive main effects and multiplicative interaction-derived stability indexes. The shaded ellipses highlight four visually formed groups. Color key represents the quality of the representation for variables on the factor map.

IPCA (Fig. 6). It would be reasonable since the scores for this genotype in the first two IPCAs were very low (Supplemental Table S4). Considering the explained variance in such axes (Supplemental Table S3), the WAASB for G2 would be then $\text{WAASB}_{G2} = [(0.063 \times 32.49) + (0.103 \times 26.79)] / 59.28 = 0.081$. This genotype was also classified as the most stable by the ASV (Supplementa Table S5). Considering two IPCA, WAASB and ASV's ranks matched perfectly (data not shown). We need to be cautious since in our case only 59% of the GEI pattern was explained by the first two IPCA. The extent to which the trial networks and the complexity of GEI increase, the GEI pattern is retained in a larger number of axes, tending to be captured in the last IPCA axes. The G2 still remains a good example here. The score for this genotype on IPCA3 was 0.84, and even though IPCA3 explained only 16.9% of GEI pattern, we need to take it into account; WAASB does that. Based on this difference, for example, we may explain why the association between WAASB and ASV was lower than the WAASB and Za (Fig. 9).

Studies with a relatively low percentage of explanation in IPCA1 were observed in trials with *Brassica* spp. (38.6–54.6% [Bocianowski et al., 2019]); sugarcane (32.9-35% [Ramburan et al., 2011]); maize (26.4% [Balestre et al., 2009)] and 40.8% [Oyekunle et al., 2017]); and wheat (43.5% [Bornhofen et al., 2017]; 32.1% [Tigabu et al., 2017]; 33.1-38.2% [Veenstra et al., 2019]). In this sense, biplots with low GEI pattern recovery must be interpreted cautiously in trial networks with prevalence of crossover interaction, since only the simple part of the GEI can be represented in the first main components and the complex part of the GEI may be discarded. For example, Shahriari et al. (2018) identified genotypes of *Plantago* spp. with high mucilage content and high stability by using an AMMI1 biplot with only 36.6% of the GEI pattern explained by that axis. Thus, the use of WAASB × GY (Fig. 5) may be promising in identifying highly productive and broadly adapted genotypes in future studies. Assuming that a specific member of the AMMI family (say AMMI5) is the most predictively accurate model, the Weighted Average of Absolute Scores (WAAS) can also be obtained based on traditional AMMI usage. The estimates are also based on Eq. [9] and may be obtained with the function *waas()* (Supplemental R codes S3.3.1 and S3.3.2).

Weighting between Mean Performance and Stability with WAASBY

Yan and Kang (2003, p. 91) have said that "...stability has rarely been used by plant breeders for various reasons. One reason is that it is difficult to weigh between mean performance and stability...". There have been earlier studies that used ranking ASV and mean performance to compute a nonparametric simultaneous selection index (ssiASV) to identify genotypes that combine high performance and stability (Farshadfar, 2008; Farshadfar et al., 2011; Adjebeng-Danquah et al., 2017; Bocianowski et al., 2019). Simultaneous selection using ssiASV would be promising, provided that the rank for stability-in such studies computed by the ASV-is reliable. We have shown that the ASV's ranking may be misleading if the explanation of GEI pattern in the first two IPCA is low. Thus, we recommend careful when using this index. In a fixed-effects model framework, future studies should then consider using simultaneous selection indexes that are based on significant IPCAs, such as WAASY, ssiZa, ssiEV, and ssiSIPC.

The WAASBY was found to be a useful simultaneous selection index in future analysis of MET under a mixed-effects model framework. This index differs from the "ssis" shown in Fig. 9 in two main ways: The first and logical is that the WAASBY is based on a mixed-effects model-or even on a random-effects-model framework, and may have more reliable results than the ssiASV, for example, since the stability (WAASB) is quantified considering all estimated IPCA. The second-and perhaps the more interesting-difference is that different weights may be assigned to the performance and stability. This is important because depending on the goal of a breeding program or a cultivar recommendation trial, the researcher may want to prioritize the productivity of a genotype rather than its stability (and vice-versa). Thus, Fig. 7 and 8 should help breeders and agronomists make selection and cultivar recommendation decisions in addition to identifying groups of genotypes with similar mean performance and stability.

CONCLUSIONS

Based on a cross-validation procedure using four real MET datasets we have shown that the predictive accuracy was higher using BLUP than any member of the AMMI family. We also have shown how nice graphical tools may be obtained to model

a random GEI effect in the analysis of MET using a LMM. In our study, we considered a genotypic random effect, but for future studies, the same procedures may be used considering an LMM with random effect for environment or even a completely random-effect model. The genotypic stability index introduced in this article, WAASB, has the potential to provide reliable estimates of stability in future studies allowing a joint interpretation of performance and stability in a bidimensional plot considering two or more IPCA. This was important because the ranking of some genotypes can be mistakenly calculated when only the first IPCA is considered. In addition, our simultaneous selection index, WAASBY, may be useful when different weights should be assigned for performance and stability. Finally, the complete support provided in the **metan** R package will be useful for the reproduction and possible adaptation of all the procedures shown in this article.

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online. It contains six tables, three figures, and a brief introduction to the **metan** R package with the codes and data used to illustrate the method.

ACKNOWLEDGMENTS

We thank the National Council for Scientific and Technological Development (CNPq) and Coordination for the Improvement of Higher Education Personnel (CAPES) for fellowships and grants to the authors. The first author also thanks Professor Dr. Maicon Nardino from Fed. University of Viçosa for your numerous helpful suggestions in the planning step of the manuscript. The authors are also thankful to the anonymous reviewers and the *Agronomy Journal* editorial board for their constructive criticism and valuable suggestions during the peer-review process.

REFERENCES

Adjebeng-Danquah, J., J. Manu-Aduening, V.E. Gracen, I.K. Asante, and S.K. Offei. 2017. AMMI stability analysis and estimation of genetic parameters for growth and yield components in cassava in the forest and guinea savannah ecologies of Ghana. Int. J. Agron. 2017:1–10. doi:10.1155/2017/8075846

Balestre, M., R.G. Von Pinho, J.C. Souza, and R.L. Oliveira. 2009. Genotypic stability and adaptability in tropical maize based on AMMI and GGE biplot analysis. Genet. Mol. Res. 8:1311–1322. doi:10.4238/vol8-4gmr658

Bocianowski, J., J. Niemann, and K. Nowosad. 2019. Genotype-by-environment interaction for seed quality traits in interspecific cross-derived *Brassica* lines using additive main effects and multiplicative interaction model. Euphytica 215(7):1–13. doi:10.1007/s10681-018-2328-7

Bornhofen, E., G. Benin, L. Storck, L.G. Woyann, T. Duarte, M.G. Stoco, and S.V. Marchioro. 2017. Statistical methods to study adaptability and stability of wheat genotypes. Bragantia 76:1–10. doi:10.1590/1678-4499.557

Bornhofen, E., M.H. Todeschini, M.G. Stoco, A. Madureira, V.S. Marchioro, L. Storck, and G. Benin. 2018. Wheat yield improvements in Brazil: Roles of genetics and environment. Crop Sci. 58:1–12. doi:10.2135/cropsci2017.06.0358

da Silva, C.P., L.A. de Oliveira, J.J. Nuvunga, A.K.A. Pamplona, and M. Balestre. 2015. A Bayesian shrinkage approach for AMMI models. PLoS One 10:e0131414. doi:10.1371/journal.pone.0131414

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B 39:1–38.

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

Farshadfar, E. 2008. Incorporation of AMMI stability value and grain yield in a single non-parametric index (GSI) in bread wheat. Pak. J. Biol. Sci. 11:1791–1796. doi:10.3923/pjbs.2008.1791.1796

Farshadfar, E., N. Mahmodi, and A. Yaghotipoor. 2011. AMMI stability value and simultaneous estimation of yield and yield stability in bread wheat (*Triticum aestivum* L.). Aust. J. Crop Sci. 5:1837–1844.

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

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

Gauch, H.G. 2013. A simple protocol for AMMI analysis of yield trials. Crop Sci. 53:1860–1869. doi:10.2135/cropsci2013.04.0241

Gauch, H.G., and R.W. Zobel. 1988. Predictive and postdictive success of statistical analyses of yield trials. Theor. Appl. Genet. 76:1–10. doi:10.1007/BF00288824

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

Gollob, H.F. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. Psychometrika 33:73–115. doi:10.1007/BF02289676

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447. doi:10.2307/2529430

Kempton, R.A. 1984. The use of biplots in interpreting variety by environment interactions. J. Agric. Sci. 103:123–135. doi:10.1017/S0021859600043392

Mooers, C.A. 1921. The agronomic placement of varieties. J. Am. Soc. Agron. 13:337–352. doi:10.2134/agronj1921.00021962001300090002x

Nardino, M., D. Baretta, I.R. Carvalho, T. Olivoto, D.N. Follmann, V.J. Szareski s et al. 2016. Restricted maximum likelihood/best linear unbiased prediction (REML/BLUP) for analyzing the agronomic performance of corn. African J. Agric. Res. 11:4864–4872. doi:10.5897/AJAR2016.11691

Olivoto, T. 2018. WAASBdata, Mendeley Data, v2. doi.org/10.17632/2sjz32k3s3.2

Olivoto, T. 2019. Metan: multi environment trials analysis. R package version 1.1.0. https://github.com/TiagoOlivoto/metan (accessed 24 June 2019).

Olivoto, T., M. Nardino, I.R. Carvalho, D.N. Follmann, M. Ferrari, V.J. Szareski et al. 2017. REML/BLUP and sequential path analysis in estimating genotypic values and interrelationships among simple maize grain yield-related traits. Genet. Mol. Res. 16:gmr16019525. doi:10.4238/gmr16019525

Oyekunle, M., A. Menkir, H. Mani, G. Olaoye, I.S. Usman, S.G. Ado et al. 2017. Stability analysis of maize cultivars adapted to tropical environments using AMMI analysis. Cereal Res. Commun. 45:336–345. doi:10.1556/0806.44.2016.054

Piepho, H.-P. 1994. Best linear unbiased prediction (BLUP) for regional yield trials: A comparison to additive main effects and multiplicative interaction (AMMI) analysis. Theor. Appl. Genet. 89:647–654. doi:10.1007/BF00222462

Piepho, H.-P. 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Theor. Appl. Genet. 97:195–201. doi:10.1007/s001220050885

Purchase, J.L., H. Hatting, and C.S. van Deventer. 2000. Genotype × environment interaction of winter wheat (*Triticum aestivum* L.) in South Africa: II. Stability analysis of yield performance. S. Afr. J. Plant Soil 17:101–107. doi:10.1080/02571862.2000.10634878

Ramburan, S., M. Zhou, and M. Labuschagne. 2011. Interpretation of genotype × environment interactions of sugarcane: Identifying significant environmental factors. Field Crops Res. 124:392–399. doi:10.1016/j.fcr.2011.07.008

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (accessed 24 June 2019)

Sa'diyah, H., and A.F. Hadi. 2016. AMMI model for yield estimation in multi-environment trials: A comparison to BLUP. Agric. Agric. Sci. Procedia 9:163–169. doi:10.1016/J.AASPRO.2016.02.113

Shahriari, Z., B. Heidari, and A. Dadkhodaie. 2018. Dissection of genotype × environment interactions for mucilage and seed yield in Plantago species: Application of AMMI and GGE biplot analyses. PLoS One 13:e0196095. doi:10.1371/journal.pone.0196095

Smith, P.L. 1978. Sampling errors of variance components in small sample multifacet generalizability studies. J. Educ. Stat. 3:319–346. doi:10.3102/10769986003004319

Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. J. Agric. Sci. 143:449–462. doi:10.1017/S0021859605005587

Sneller, C.H., L. Kilgore-Norquest, and D. Dombek. 1997. Repeatability of yield sability statistics in soybean. Crop Sci. 37:383–390. doi:10.2135/cropsci1997.0011183X003700020013x

Stroup, W.W., and D.K. Mulitze. 1991. Nearest neighbor adjusted best linear unbiased prediction. Am. Stat. 45:194–200. doi:10.2307/2684289

Tigabu, D.A., Z. Tadesse, H. Zegeye, and A. Assefa. 2017. Seasonal variability and genetic response of elite bread wheat lines in drought prone environments of Ethiopia. J. Plant Breed. Genet. 5:15–21.

van Eeuwijk, F.A., D.V. Bustos-Korts, and M. Malosetti. 2016. What should students in plant breeding know about the statistical aspects of genotype × environment interactions? Crop Sci. 56:2119–2140. doi:10.2135/cropsci2015.06.0375

Veenstra, L.D., N. Santantonio, J.-L. Jannink, and M.E. Sorrells. 2019. Influence of genotype and environment on wheat grain fructan content. Crop Sci. 59:190–198. doi:10.2135/cropsci2018.06.0363

Yan, W., and M.S. Kang. 2003. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL.

Yang, R.-C. 2007. Mixed-model analysis of crossover genotype–environment interactions. Crop Sci. 47:1051–1062. doi:10.2135/cropsci2006.09.0611

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

Zali, H., E. Farshadfar, S.H. Sabaghpour, and R. Karimizadeh. 2012. Evaluation of genotype × environment interaction in chickpea using measures of stability from AMMI model. Ann. Biol. Res. 3:3126–3136.