Cornous Biology, Volume:1, Issue: 2, Pages: 1-12
https://doi.org/10.37446/corbio/rsa/1.2.2023.1-12

Research Article

Evaluation of maize varieties through multi-environment trials: application of multiplicative mixed models

Yidnekachew Marid[1*], Tarekegn Argaw[2]

Received: 09 February 2023 / Accepted: 28 June 2023 / Published: 30 September 2023

[1]EIAR, Hawassa National Maize Research sub-Center, Hawassa-Ethiopia, PO box 198 Shashamane, Ethiopia.
[2]EIAR, Department of Biometrics, PO box 2003, Addis Ababa, Ethiopia.

[*]**Correspondence**
Yidnekachew Marid
yidnems@gmail.com

Ethiopia is a key maize producer in Africa. Over the previous two decades, Ethiopia's maize sector has seen tremendous development. Farmers in Ethiopia demand a continual supply of novel and improved varieties to satisfy their ever-changing production and marketing difficulties. Breeders can no longer function without the analysis of multi-environment trials (MET) for varietal evaluation. To reliably choose better varieties that boost agricultural production, efficient statistical methods for maize variety evaluation must be used. This study used multiplicative mixed models to analyze data from multi-environment trials in order to identify outstanding maize varieties based on yield performance. In this study, 32 maize varieties, including four checks, were sown across seven major maize growing areas in Ethiopia using RCB design, with three replications during the main cropping season in 2020. The findings showed that factor analytic models were a successful approach for maize MET data analysis under the linear mixed model. The examined FA models have better data fitting, which significantly improves heritability. SXM1910008 and SXM1910007 showed good yield performance over correlated locations, including Ambo, Bako, Hawasa, and Wondogenet, and were therefore identified as potentially useful stable genotypes with a wide range of adaptability. This is because the improved analysis technique we used here showed that correlated locations were the basis for genotype selection. Data from multi-environment trials can be analyzed to provide a more reliable framework for evaluating maize varieties, giving breeders more confidence to select superior varieties for a wide range of environments. This can be done by using more efficient statistical models. In order to improve the selection of better varieties in the maize breeding program, it is vital to increase the usage of this efficient analysis technique.

Keywords: *factor analytic model, MET analysis, BLUP, mixed model, maize*

Introduction

The second most commonly grown crop across the world is maize. Between 2007 and the year 2022, maize area coverage in sub-Saharan Africa (SSA) expanded by approximately 66 percent. Maize grain yields in all nations have increased by over twofold, from around 1.6 t/ha in 1990 to 4 t/ha in the past decade, which is the highest in the region after South Africa (FAOSTAT, 2022). Maize takes up over fifty percent of the land used for grain production in the majority of SSA nations (Masuka et al., 2017). The economic well-being and food security of many millions of households in SSA depend on maize, which makes it a crucial grain (Fisher et al., 2015). Maize supplies 43 percent of protein and 45 percent of the total number of calories found in grains in eastern and southern Africa (Shiferaw et al., 2011). SSA still has the lowest maize yields worldwide, despite the crop's significance to the area (Masuka et al., 2017). Among all the grains, maize ranks second in area coverage (>2.5 million hectares), first in total output (> 10.5 million tons), and first in grain yield (4.18 t/ha) in Ethiopia (CSA, 2022). Africa's largest producer of maize is Ethiopia. Over the past 20 years, Ethiopia's maize industry has undergone a substantial development. Improved hybrid seeds and inorganic fertilizers are just two examples of modern inputs that are more readily available and used, together with enhanced extension services and rising demand (Abate et al., 2015). Due to an absence of well-adapted and improved cultivars, Ethiopia's national average maize grain yield is still quite low compared to the crop's

potential and the global average (Wolde et al., 2020). Mulatu et al. (1993) stated that maize can adapt to a broad range of environmental conditions and has been produced all over Ethiopia, from lowland areas with moisture stress to high-rainfall locations and everywhere in between. Maize is cultivated across a range of environments, and different maize varieties exhibit varying degrees of environmental adaptation. These variations are principally influenced by flowering time, abiotic stress tolerance, and disease resistance (Mercer & Perales, 2019). Ethiopia has a wide variety of climate zones, from highland regions to lowland deserts, semi-arid zones to temperate zones. Because of this, the grain production of various maize varieties may change depending on the environment. This indicates that plant output may change from one environment to the next since different genotypes will not always express their genetic potential in the same way under varied environmental conditions. In Ethiopia and other nations where the environment is extremely unpredictable and there are few effective ways to change it, performance stability is also of particular importance. In small-scale agricultural systems, poor genotype turnover and genotype-environment (GE) interaction are also major contributors to low yield (Abakemal et al., 2016; Wolde et al., 2020). Breeders must analyse multi-environment trials (MET) in order to evaluate genotype performance. Each genotype responds differently to changing climatic and soil circumstances; some show significant GE interaction while others show low GE interaction. Other researcher (Solomon Admassu et al., 2008) has looked at the estimation of G x E interaction and yield stability study of Ethiopian maize depending on the classical statistical way of analysis. Numerous G x E studies on Ethiopian maize genotypes have been conducted to increase information on the environmental and genetic factors causing the association, as well as an assessment of their value in the relevant G x E system, which could have substantial effects on plant breeding and genetic variation. Various research institutes have generated and assessed numerous maize hybrid genotypes in various places, but the majority of them were unable to adapt due to climate change effects and the changing patterns of the growing environment (Wolde et al., 2018; Solomon Admassu et al., 2008; Abate, 2020). In order to accurately choose superior varieties that increase agricultural production, the novel hybrids should be tested in multiple locations to ensure their broad adaptability. MET analysis using effective statistical methodologies must also be employed for maize variety testing. Ordinary linear models (LMs), which use ordinary least square (OLS) techniques of estimation for unknown parameters, are used to estimate the analysis of variance (ANOVA), a common technique frequently used to analyse MET data sources. The following methods are used in this strategy: Duncan testing for observed means, post hoc multiple comparisons testing with largely list significance difference (LSD), and an ANOVA table for source of variation testing with overall f-test. In order to further analyse the genotype by environment interaction (GEI) component, multivariate analysis methods like GGE (genotype and genotype by environment interaction) and AMMI (additive main effects and multiplicative interaction) are employed (Rodrigues, 2018; Yan & Tinker, 2006). This method had a major weakness in that it was unable to manage error variance heterogeneity across experiments, spatial variation within trials, imbalanced data, and missing values, as various authors (Gogel et al., 2018; Smith et al., 2005) pointed out. The linear mixed model (LMM), an extended linear model, can account for confounded factors in the experimental units by incorporating fixed and random terms into the model for systematic variability and relaxing the distributional assumptions surrounding the residual error (Kelly et al., 2007; Smith et al., 2005). According to Yang (2010), the LMM is a potent and powerful statistical model that allows for the computation of BLUPs (best linear unbiased predictions) for random effects as well as unbiased estimates of the variance component using REML (restricted maximum likelihood estimation) for random terms (Yang, 2010). LMMs can be used for both balanced and unbalanced field trial data, as well as for extended analysis with factor analytic models (Smith & Cullis, 2018). By shrinking the estimates of genotype effects closer to their true value, MET data analysis under LMM with random genotype could increase the precision for genotype ranking. Through FA models, the covariance structure of GE effects has been further improved. The Bako national maize research program regularly generates and evaluates novel maize hybrids that adapt to Ethiopia's mid-altitude subhumid and transitional highland maize growing areas. Although maize has many advantages, some of the challenges include biotic and abiotic stress, which confronts the breeder when developing improved varieties. Thus, this study was planned to evaluate the performance of promising maize varieties that might suit the local and regional market through data analysis of MET using more efficient statistical methods.

Materials and methods

Experimental sites, experimental design and crop management

Seven sites, which were chosen to reflect the main maize-growing agro-ecologies in Ethiopia, were used for the experiment. These locations vary in altitude, temperature, total annual rainfall and soil types (Table 1) and the locations represent the main maize-producing agro-ecologies of the country ranging from mid-altitude sub-humid to transitional high land sub-humid.

Table 1. Description of the study locations

Location	Altitude (m.a,s.l)	Soil type	Rainfall (mm)	Geographical position		Temperature	
				Latitude	Longitude	Max(°C)	Min(°C)
Bako	1650	Nitisol	1598	9° 06’	37°09’	29	12.78
Asosa	1547	Nitisol	1276.2	10° 02'	34° 31’	33	21
Jimma	1753	Nitosol	1561	7° 46’	36° 00’	23	18
Pawe	1120	Nitisol	1250	11°19’	36° 24’	32.6	16.5
Wondo Genet	1780	Alluvial	1128	7° 19’	38° 38’	26	11
Ambo	2175	Vertisol	1265.7	8° 57’	37° 51’	25.6	11.7
Hawasa	1650	Sandy loam	959	7° 03’	38° 30’	26.9	12.4

Source: Ethiopian institute of agricultural research (2019)

Thirty-two maize genotype with four commercial checks (BH546, BH547, BH661 and Limu) were evaluated in the multi-location trial (Table 2). These hybrids were developed or adapted by the National Maize Research Program based at the Bako Agricultural Research Centre (BARC).

Table 2. Maize hybrids tested across six locations in 2020 main growing season.

Entry	Hybrids	Pedigree	Source
1	WE6103	CKDHL0089/CML395//CKLTI0036-B-B	CIMMYT
2	WE7124	CKDHL0089/CKDHL0295//CKLTI0348-B-B	CIMMYT
3	CZH15568	CZH15568	CIMMYT
4	WE2108	CML312/CML442//CKDHL0411-B-B-B	CIMMYT
5	CZH15587	CZH15587	CIMMYT
6	WE7117	CKLTI0139/CKLMARSI0029//CKDHL120312-B-B-B	CIMMYT
7	BH 661	CML395/CML202//142-1-e	Bako
8	SXM1910008	BKL004/BKL003	Bako
9	BH 546	CML395/CML202/BKL001	Bako
10	BH 547	CML312BK/BKL002/BKL003	Bako
11	SXM1910173	SC22/124- b (113)	Bako
12	Limu	Limu	Pioneer
13	WE3105	CML444/CML442//CKDHL0295-B-B-B	CIMMTY
14	CZH15523	CZH15523	CIMMTY
15	3XM1900476	CML488/CML489/CML536	Bako
16	SXM1910007	CML444/CML536	Bako
17	WE3106	CML312/CML395//CKDHL0089-B-B-B	CIMMYT
18	WE7131	CKDHL0089/CKDHL0323//CKLTI0045-B-B	CIMMYT
19	WE7126	CML395/CML444//CKLTI0348-B-B	CIMMYT
20	WE7119	CKDHL0500/CKLTI0137//CKDHL120312-B-B-B	CIMMYT
21	WE7128	CKDHL0089/CML395//CKLTI0368-B-B-B	CIMMYT
22	WE1101	CML395/CML444//CML539-B-B-B	CIMMYT
23	WE6105	CKDHL0089/CKDHL0295//CKLTI0344-B-B	CIMMYT
24	WE6106	CKDHL0089/CKDHL0323//CKLTI0200-B-B-B	CIMMYT
25	3XM1910230	-	CIMMYT
26	CZH 131009	-	CIMMYT
27	CZH 131010	-	CIMMYT
28	CZH 131013	-	CIMMYT
29	CZH 131015	-	CIMMYT
30	CZH 132080	-	CIMMYT
31	CZH 141029	-	CIMMYT
32	WE2109	-	CIMMYT

During the main farming season of 2020, a Randomized Complete Block Design (RCBD), with three replications was employed to conduct the study. Each hybrid was planted in a two-row plot that was 5 m long, with 0.75 m between rows and 0.25 m between plants within a row. Two seeds per hill for each genotype were sown to achieve the recommended total plant population of 53,000 plants per hectare, which was then reduced to one plant at three to four

leaf stages. To ensure excellent germination and seedling growth and development, planting took place as soon as the main rainy season began, and the soil moisture level was sufficient. According to the recommendation, NPS fertilizer at a rate of 150 kg/ha was applied once during planting time at all locations, whereas urea at a rate of 200 kg/ha at Ambo and Pawe and 250 kg/ha at Hawassa, Bako, Wendo Genet, Jima, and Asosa was used in split doses, with half being applied at thinning and the other half at knee height.

Linear mixed model (LMM)

Take a look at a MET dataset that was generated using t-trials (environments might also be used) and m varieties that were planted (not all trials might have grown every variety). The k^{th} trial, where k = 1...t, is made up of n_k plots that are organized in a rectangular array with c_k columns and r_k rows ($n_k=c_kr_k$). Let y_k be the (n_k x 1) data vector for trial k, which is arranged as rows within columns, and let y be the (n x 1) data vector combining all of the trials together. The LMM for y can then be written as

$$y = X\alpha + Z_g\gamma_g + Z_p\gamma_p + \varepsilon \qquad (1)$$

where α is vector of fixed effects (including terms for the grand mean, the environment's main effects, global spatial trends at each trial, and other trial-specific fixed effects) with an associated design matrix X (assumed to be full column rank), γ_g is the mt x 1 vector of random genetic (or variety by trial) effects with associated design matrix Z_g, γ_p is a vector of non-genetic (or peripheral) random effects (including terms associated with the blocking structure at each trial, and other trial-specific random effects), with associated design matrix Z_p, and ε is the n x 1 vector of residual errors across all trials.
The random effects from the linear mixed model (equation 1) are assumed to follow a Normal distribution with mean zero vector and variance-covariance matrics, and this can be written using *E (*expectation*) and* var (variance) functions as follow

$$E\begin{pmatrix}\gamma_g\\ \gamma_p\\ \varepsilon\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}, \quad \text{var}\begin{pmatrix}\gamma_g\\ \gamma_p\\ \varepsilon\end{pmatrix}=\begin{bmatrix}G_g & 0 & 0\\ 0 & G_p & 0\\ 0 & 0 & R\end{bmatrix}$$

where α is vector of fixed effects (including terms for the grand mean, the environment's main effects, global spatial trends at each trial, and other trial-specific fixed effects) with an associated design matrix X (assumed to be full column rank), γ_g is the mt x 1 vector of random genetic (or variety by trial) effects with associated design matrix Z_g, γ_p is a vector of non-genetic (or peripheral) random effects (including terms associated with the blocking structure at each trial, and other trial-specific random effects), with associated design matrix Z_p, and ε is the n x 1 vector of residual errors across all trials.

The random effects from the linear mixed model (equation 1) are assumed to follow a Normal distribution with mean zero vector and variance-covariance matrix, that is

Model for Genetic Effects (γ_g)

Smith et al. (2001a) presented an alternative parsimonious model for γ_g using a factor analytic (FA) model approach to provide a variance structure for the genetic variance matrix G_g. This model can adequately represent the nature of heterogeneous variances and covariance found to occur in most MET data. Thus, the γ_g can be modeled with multiplicative terms. That is

$$\gamma_g = (\lambda_1 \otimes I_m) f_1 + \ldots + (\lambda_k \otimes I_m) f_k + \zeta$$
$$= (\Lambda \otimes I_m) f + \zeta \qquad (2)$$

where λ_r is the $t \times 1$ vector of loadings, f_r is the $m \times 1$ vector of factor scores ($r = 1..k$), ζ is the $mt \times 1$ vector of residuals, Λ is the $t \times k$ matrix of loadings $\{\lambda_1 \ldots \lambda_k\}$ and f is the $mk \times 1$ vector of factor scores $(f_1' f_1' .. f_k')'$. The random effects f and ζ are assumed to follow a Normal distribution with zero mean vector and variance-covariance matrix

$$\begin{bmatrix} G_f \otimes I_m & 0 \\ 0 & \Psi \otimes I_m \end{bmatrix}$$

where Ψ is a diagonal matrix of specific variances represents the residual variance not explained by the factor model, that is Ψ = diag ($\Psi_1 \ldots \Psi_t$). The factor scores are commonly assumed to be independent and scaled to have unit variance, so that $G_f = I_k$.

The genetic effects γ_g can be considered as a two dimensional (genotype by environment) array of random effects, and can be assumed to have a separable variance structure for the (mt × mt) variance matrix G_g which can be written as

$$G_g = \mathrm{G_e} \otimes G_v$$

where G_e is the $t \times t$ genetic variance matrix representing the variances at each trial and covariances between trials, and G_v is the $m \times m$ symmetric positive definite matrix represents variances of environment effects at each genotype and the covariances of environment effects between genotypes. It is typically assumed that the varieties are independent and that $G_v = I_m$. However, if the pedigree information of the varieties is available, other forms of G_v can be applicable (Oakey et al., 2006; 2007). Based on equation 2 the variance of genetic effects would be

$$\mathrm{var}(\gamma_g) = (\Lambda\Lambda' + \Psi) \otimes I_m$$
$$= \mathrm{G_e} \otimes I_m$$

Thus, the FA model approach results in the following form for G_e

$$G_e = \Lambda\Lambda' + \Psi$$

In the model, the variance parametric in these variance matrices are directly estimated using REML estimation method.

Model for Non-genetic Effects (γ_p)

The random non-genetic effects γ_p can be considered as sub- vectors $\gamma_{pj}^{(b_j \times 1)}$ for each trial, where b_j is the number of random terms for trial j. These random terms are based on terms for blocking structure (replicate blocks or other terms). In the analysis of MET data, the sub-vectors of γ_p are typically assumed to be mutually independent, with variance matrix G_{pj} for trial j, with the block diagonal form given below. Thus, there is a variance matrix for the set of none-genetic effects at each trial, That is,

$$G_p = \oplus^t{}_j G_{pj} = \begin{bmatrix} G_{p1} & 0 & \cdots & 0 \\ 0 & G_{p2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G_{pt} \end{bmatrix}$$

The most common form for the variance matrix of these extraneous effects is a simple variance component structure,

where $G_{pj} = \sigma_j^2 I_{bj}$

Estimation, testing and software

The significance of fixed effects in a linear mixed model can be evaluated through the Wald test. The traditional Wald statistic follows an asymptotic chi-squared distribution, but it is often considered as overly liberal by some researchers (Butler et al., 2009). To address this issue, Kenward & Roger (1997) introduced an F approximation and an adjusted Wald statistic, which have shown good performance in various scenarios. ASReml, a software package implemented in the R environment, was utilized to estimate the variance parameters of the linear mixed model using the Restricted Maximum Likelihood (REML) method (Butler et al., 2009). The ASReml software employs the Average Information (AI) algorithm proposed by Gilmour et al. (1997). During the estimation process of the linear mixed model, the variance-covariance parameters (represented by symbols G_g, G_p and R), as well as the fixed effects (α) and random effects (γ_g and γ_p), are all estimated. This estimation process consists of two interconnected steps. The fixed effects are estimated using Best Linear Unbiased Estimation (BLUE), while the random effects are estimated using Best Linear Unbiased Prediction (BLUP). The variance parameters of the model are estimated using the Residual Maximum Likelihood (REML) method, as described by Patterson & Thompson (1971). To determine the significance of random effects in the linear mixed model, the Residual Maximum Likelihood Ratio Test (REMLRT) is employed. The REMLRT is used when comparing the fit of two nested models with the same fixed effects.

Results and discussion

G x E analysis

For the G x E analysis, the FA models were taken into consideration while preserving the single stage-wise analysis on the individual plot yield data. The degree to which the G x E variance is explained by the factor components was used to evaluate the suitability of the two-factor FA model (Cullis et al., 2010). The factor analysis's findings are shown in Table 3. Excluding for the two trials Jimma and Asosa, nearly all of the trials were largely explained by the FA models, and the two factor components gave admirable explanation to the genetic variance. This comprises the total proportion of (G x E) variation explained by the factor components of the model for both the individual trials as well as the entire sample of trials. Due to the insufficient dataset fit, the factor analytic model was not taken into account for more than two factors. The two multiplicative terms of the FA model explained nearly seventy percent of the variance in the G x E effects, with the first multiplicative term accounting for fifty-three percent of that variance.

Table 3. Results from fitting FA model.

	Factor1	Factor2	All
Ambo	99.5	0.5	100
Asosa	22.69	1.47	24.15
Bako	73.48	1.05	74.53
Hawasa	80.24	19.76	100
Jimma	18.88	4.98	23.86
Pawe	0.02	70.33	70.35
Wondogenet	61.55	2.17	63.72
%var FA-1= 53.51,		% var FA-2=70.19	

%varFA-1= percentage of GxE variance explained from fitting FA model with a single factor; %var FA-1=percentage of GxE variance explained from fitting FA model with two factors.

The FA models do not adequately explain Assosa and Jimma, which can occur because these trials lack correlation with the other trials or are unique in comparison to the others.

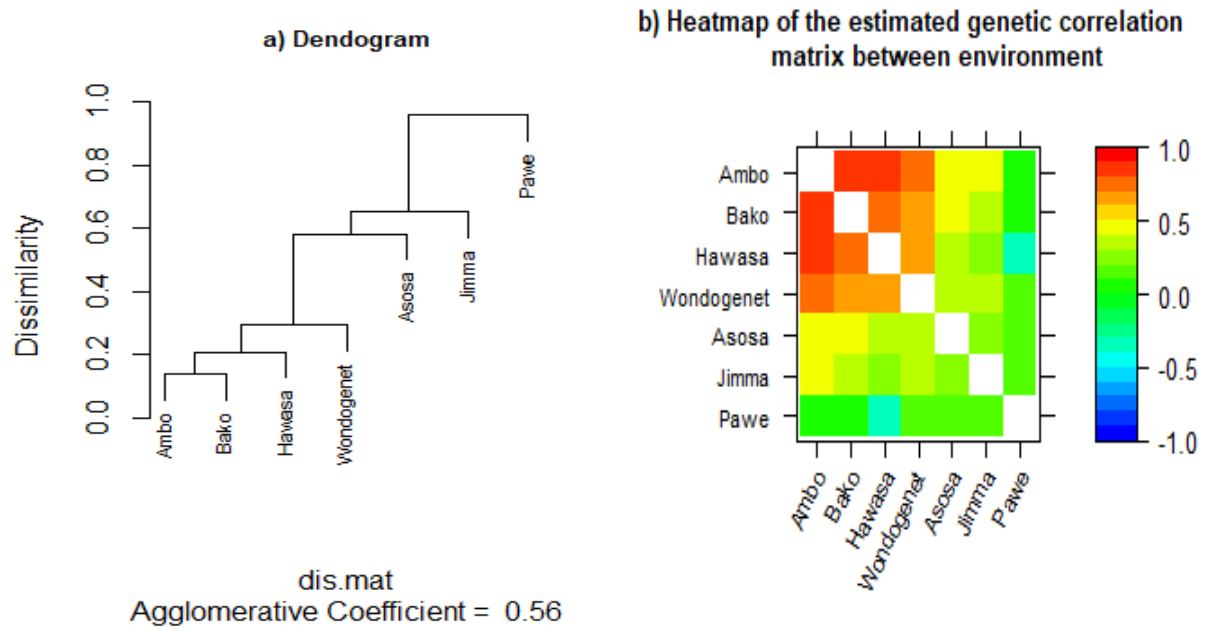


Figure 1. Dendrogram of the dissimilarity matrix (a) and heatmap representation of the genetic correlation matrix (b).

Factor analysis also produces another important summary of statistics when cluster analysis is performed using a dendrogram. The cluster analysis grouped the trials according to how environmentally related they were using the dendrogram in Figure 1(a). According to Cullis et al. (2010) estimate of the variation cut-off (around about 0.6) at which clusters form, the dendrogram shows that there may be three clusters of trials, with the first cluster having a maximum of four trials. This demonstrates that, whereas genotype rankings differ for trials located in different clusters, they are substantially the same for trials located inside these established clusters. Given that the produced clusters are logically reasonable for doing genotype selection independently for every one of the clusters, genotype selection was performed for each cluster individually utilizing average BLUPs as a selection index. A heatmap depicting the genetic links among all trials is another popular component of factor analysis reports, in addition to the dendrogram. The correlation patterns between trials are shown in Figure 1(b), which illustrates the similarity of the trials. Only a few of the trials had a poor correlation, as evidenced by the heatmap, which reveals that the majority of the trials are highly connected. This suggests that almost all of the trials in the first cluster with the red hue can be used to average genotype means for genotype selection. Additionally, there are trials with a negative genetic association, such as the one between Pawe and Hawasa (Table 4), which suggests that genotype rankings may have reversed in these trials.

Table 4. Genetic correlation between environments

	Ambo	Asosa	Bako	Hawasa	Jimma	Pawe	Wondogenet
Ambo	1						
Asosa	0.48	1					
Bako	0.86	0.42	1				
Hawasa	0.86	0.37	0.72	1			
Jimma	0.45	0.23	0.40	0.3	1		
Pawe	0.08	0.11	0.10	-0.4	0.19	1	
Wondogenet	0.79	0.39	0.69	0.6	0.37	0.14	1

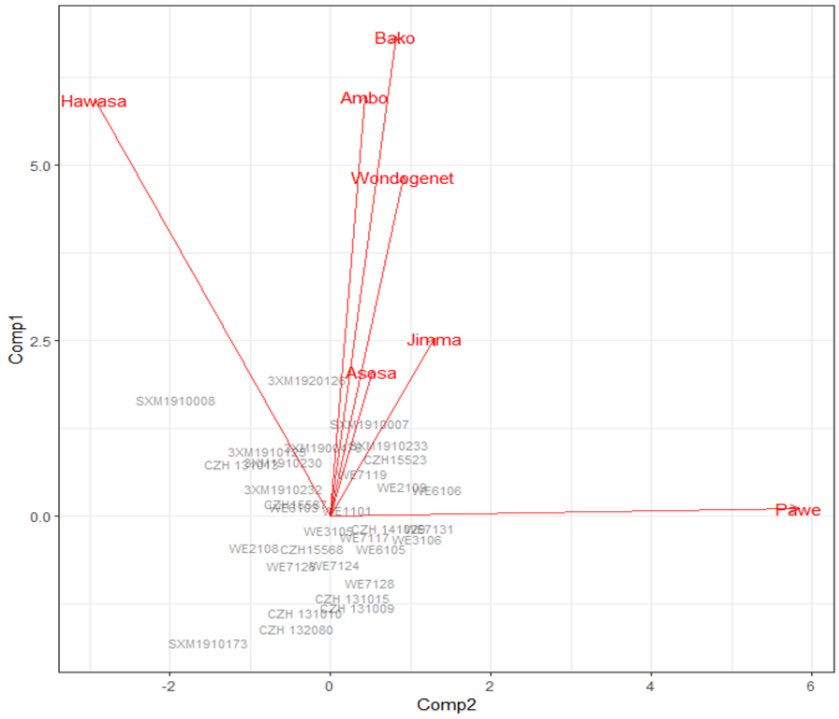


Figure 2. Bi-plot analysis

The biplot in Figure 2 shows the concept of genotype performance, consistency across environments, and the discriminating abilities of each trial. The trials that have a lengthy arm from the bi-plot's centre have comparatively strong genotype discrimination power compared to the others and show a large genetic variance (Tesfaye et al., 2023). In comparison with the other location, Pawe, Hawasa, and Bako thus showed higher genetic variance. As a result, we evaluated three Clusters of trials (C1, C2, and C3), with Ambo, Bako, Hawasa, Wendogenet, and Assosa residing on C1, Jimma residing on C2, and Pawe residing on C3. These trials were grouped together based on the dendrogram and heatmap in Figure 1, the bi-plot in Figure 2, and the genetic association as well from Table 4. In this study, we ranked average BLUPs within clusters as a selection index to choose superior and stable varieties, focusing on the first cluster (C1) being utilized for selection because it has a comparatively higher correlation of trials.

Variance components

The REML estimation produces unbiased and efficient estimates for variance component parameters at each trial (Smith et al., 2005). Table 5 shows the genetic variance, error variance, and heritability from the final fitted FA model for each trial. Variance component parameter estimates range from 0.52 to 1.78 for genetic variance, 0.56 to 3.33 for error variance, and 62.05 to 90.92 for heritability. Bako had more genetic variation. This indicates that the genotype discrimination power at these testing sites was relatively high.

Table 5. Variance component results MET analysis using FA models

	Genetic variance	Error variance	Heritability
Ambo	1.00	0.56	90.92
Asosa	0.52	0.68	73.32
Bako	1.78	3.33	77.26
Hawasa	1.22	1.40	84.75
Jimma	0.94	2.28	62.05
Pawe	1.35	1.55	75.29
Wondogenet	1.05	0.59	86.99

This could be attributed to Bako's significantly higher rainfall amounts and distribution during that growing season. This highlights the value of using meteorological information from a given cropping season to suggest the best

genotype for that cropping season as well as its wider application to the various agro-ecologies across the country. Plant breeders commonly measure both narrow and broad sense heritability on a genotype-mean basis to quantify and finally assess the accuracy of METs. The latter is the portion of phenotypic variability that can be attributed to the total genotype variability, which includes additive, dominance, and epistatic variability. The accuracy of a single field experiment or a series of field trials is frequently evaluated by plant breeders using the heritability approach (Piepho & Möhring, 2007). The preferred models for plant breeding field trial data analysis are linear models. When the underlying assumptions of classic regression models are violated, however, these models often perform poorly and produce misleading parameter estimates (Cullis et al., 2010; Smith et al., 2005). This typically happens when the data is incomplete, imbalanced, and filled by outliers. Due to these issues, estimates of the heritability and prediction power of genetic and non-genetic effects are inaccurate. Robust statistical techniques offer a theoretically sound and intuitively appealing framework for getting around some of the limitations of traditional analysis, most notably its limitation in the analysis of incomplete and correlated MET data (Having a precise and accurate understanding of heritability is essential for the plant breeding program to be successful. Learning more about the genetic components that contribute to significant character variations is of primary interest to plant breeders. Due to this, it is essential from the perspective of plant breeding programs to quantify various genetic variances and make decisions regarding their inheritance based on estimates of various genetic characteristics acquired by using reputable statistical techniques like FA mixed mode statistics. So, using both randomized complete block (RCB) and FA analysis, Figure 3 illustrates the heritability of yield at each trial. It demonstrates how applying FA analysis strengthens heredity. By properly utilizing the data recorded in the MET dataset, processing this dataset with factor analytic model often increases genotype generation precision and accuracy (Smith & Cullis, 2018; Cullis et al., 2010).

Best linear unbiased prediction (BLUPs) for genotypes across trials

A commonly used technique for calculating random effects in a mixed model is the BLUP approach. BLUPs have the property of a minimum mean square error of prediction, which allows them to create a more accurate estimation of the underlying effects Genotype effects are generally fitted as random factors in the context of plant breeding, where precise genotype ranking is essential for the selection of superior genotypes (Piepho & Möhring, 2007). This is especially important for early generation trials with a large number of entries. Genotype performance can be graded based on the averaged values of BLUPs across the correlated environments of the first cluster (C1), excluding Jimma and Pawe because they are in distinct clusters. More than 31% (10) of the 32 genotypes had average grain yields that were more than 6.5 t/ha, as shown in Table 6.

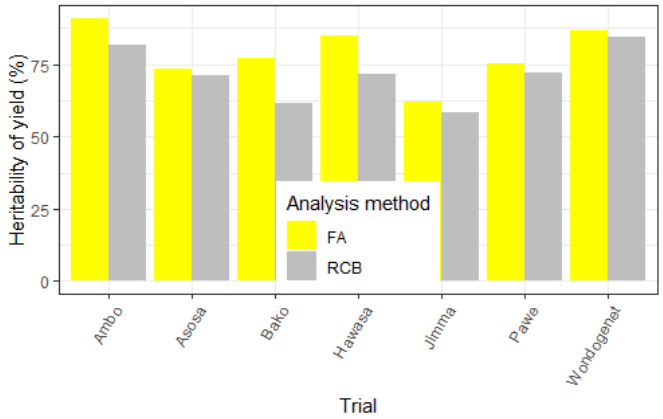


Figure 3. Improvements in heritability through the applications of FA models

Table 6. BLUPs for genotype means across clusters of correlated environments (C1)

Genotype	Ambo	Bako	Hawasa	Wondogenet	Asosa	Jimma	Pawe	Average at C1
SXM1910008	8.91	10.91	6.29	5.70	6.97	7.40	4.53	7.95
BH661	9.33	11.43	5.78	5.44	6.91	7.68	6.34	7.78
SXM1910007	8.76	11.32	4.78	5.38	6.67	8.03	7.35	7.56
Limu	8.46	10.44	4.35	5.49	6.83	6.37	8.22	7.19
3XM1900476	8.38	10.51	4.72	4.53	6.49	7.27	6.93	7.04
3XM1910230	8.12	9.97	4.75	5.00	6.77	7.66	6.38	6.96
BH546	8.27	10.50	5.01	3.52	6.51	7.08	6.14	6.82

CZH 131013	8.06	9.52	4.99	4.42	6.54	7.11	5.83	6.75
CZH15523	8.27	10.33	4.12	4.21	7.53	7.94	7.58	6.73
WE7119	8.04	9.99	4.11	4.48	6.30	7.41	7.81	6.66
BH547	7.75	10.17	4.37	3.66	7.38	6.98	5.74	6.49
WE6106	7.87	9.35	3.43	5.10	7.31	7.24	8.55	6.44
WE2109	7.89	9.93	3.69	4.20	5.72	7.22	8.10	6.43
WE6103	7.49	9.25	4.05	4.86	6.15	7.98	5.98	6.42
CZH15587	7.55	9.67	4.10	3.42	6.04	6.62	6.29	6.19
WE1101	7.50	9.12	3.68	4.27	6.75	6.31	7.56	6.14
WE3105	7.19	9.26	3.51	4.45	5.80	7.63	6.93	6.10
WE7117	7.13	8.94	3.20	4.55	5.63	6.75	6.99	5.96
WE7131	7.32	9.59	2.93	3.93	7.15	8.23	8.24	5.94
WE3106	7.15	9.65	2.85	3.91	7.18	7.83	7.89	5.89
CZH 141029	7.28	8.83	3.17	4.13	6.38	7.12	8.00	5.85
WE2108	6.88	8.99	3.73	3.24	5.85	7.43	5.52	5.71
CZH15568	6.92	9.09	3.36	3.20	5.60	6.49	7.08	5.64
WE6105	6.99	8.45	2.95	3.80	6.37	7.88	7.80	5.55
WE7126	6.66	8.72	3.25	3.21	6.50	6.01	6.14	5.46
WE7124	6.71	8.61	2.99	3.44	5.90	6.04	7.27	5.44
WE7128	6.49	8.66	2.53	3.22	5.72	6.27	7.53	5.23
CZH 131015	6.26	7.45	2.42	3.46	6.00	6.61	7.17	4.90
CZH 131009	6.12	8.00	2.25	2.62	5.71	6.56	7.27	4.75
CZH 131010	6.00	7.48	2.50	2.87	5.40	6.38	6.38	4.71
CZH 132080	5.77	7.60	2.33	1.94	5.79	6.27	6.23	4.41
SXM1910173	5.50	6.20	2.68	2.21	6.30	5.41	5.00	4.15

The estimated mean grain yield, however, revealed two genotypes that had a greater mean yield throughout trials in the first cluster (C1): one is SXM1910008, and the other is the check, BH661 (Table 6). BLUP analysis also revealed that these two genotypes did poorly at Asosa, Jimma, and Pawe, implying that these sites were not found to be ideal for selecting maize genotypes for this study. According to the enhanced method of analysis we used here, cluster one (Cl) would be the basis for genotype selection, and thus the genotypes SXM1910008 and SXM1910007 had good yield performance over correlated trials, Ambo, Bako, Hawasa, and wondogenet, and can potentially be used as stable genotypes with broad adaptability.

Conclusion

Farmers in Ethiopia require a steady supply of new and improved varieties to help them meet their constantly changing production and marketing challenges. Breeders no longer can function without the analysis of multi-environment trials (MET) for varietal evaluation. Each cultivar responds differently to shifting climatic and soil conditions; some show high GE interaction while others show low GE interaction. ANOVA-based techniques might not be useful for assessing MET data because it isn't always balanced and/or comprehensive. The linear mixed model relaxes the ANOVA distributional assumptions about the residual error and offers a robust framework for handling unbalanced and/or incomplete data. The linear mixed model with the FA models showed to be an effective data analysis technique for this investigation. The evidence of heredity measure reveals that the multiplicative mixed model analysis considerably enhances the findings of the MET data analysis. The analysis has improved because the GE effects are now modelled using FA models. The investigated FA models exhibit improved data fitting, resulting in a significant improvement in heritability. SXM1910008 and SXM1910007 were found to be potentially useful as stable genotypes with a wide range of adaptability because they demonstrated good yield performance over correlated locations, including Ambo, Bako, Hawasa, and Wondogenet. This is due to the fact that the enhanced method of analysis we employed here revealed that correlated locations served as the base for genotype selection.

Acknowledgement

The Ethiopian Institutes of Agricultural Research, the Hawassa Maize research sub-center, and the employees of the Hawassa Maize research sub-center are all gratefully acknowledged by the authors for their financial and technical assistance.

Author contributions

Yednekachew Merid developed the research idea, the work proposal, and the experiment. Tarekeng Argaw and Yednekachew Merid analyzed this research. Additionally, they wrote the manuscript's final draft and published it.

Conflict of interest

The authors declare no conflict of interest.

Ethics approval

Not applicable.

References

Abakemal, D., Shimelis, H., & Derera, J. (2016). Genotype-by-environment interaction and yield stability of quality protein maize hybrids developed from tropical-highland adapted inbred lines. *Euphytica*, *209*, 757-769.

Abate, M. (2020). Genotype by environment interaction and yield stability analysis of open pollinated maize varieties using AMMI model in Afar Regional State, Ethiopia. *Journal of Plant breeding and crop science, 12*(1), 8-15.

Abate, T., Shiferaw, B., Menkir, A., Wegary, D., Kebede, Y., Tesfaye, K., ... & Keno, T. (2015). Factors that transformed maize productivity in Ethiopia. *Food security, 7*, 965-981.

Butler, D. G., Cullis, B. R., Gilmour, A. R., & Gogel, B. J. (2009). ASReml-R reference manual. *The State of Queensland, Department of primary industries and fisheries, Brisbane*.

CSA (Central Statistical Agency). 2022. The Federal Democratic Republic of Ethiopia Central Statistical Agency Agricultural Sample Survey Report on Crop and Livestock Product Utilization (Private Peasant Holdings, Meher Season) Volume Vii.Mengesha, Addis Ababa, Ethiopia.

Cullis, B. R., Smith, A. B., Beeck, C. P., & Cowling, W. A. (2010). Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis. *Genome, 53*(11), 1002-1016.

FAOSTAT (Food and Agriculture Organization Corporate Statistical Database). 2022. Statistical databases and data sets of the Food and Agriculture Organization of the United Nations.

Fisher, M., Abate, T., Lunduka, R. W., Asnake, W., Alemayehu, Y., & Madulu, R. B. (2015). Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. Climatic Change, 133, 283-299.

Gilmour, A. R., Cullis, B. R., & Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics,* 269-293.

Gogel, B., Smith, A., & Cullis, B. (2018). Comparison of a one-and two-stage mixed model analysis of Australia's National Variety Trial Southern Region wheat data. *Euphytica, 214*, 1-21.

Kelly, A. M., Smith, A. B., Eccleston, J. A., & Cullis, B. R. (2007). The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Science, 47*(3), 1063-1070.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 983-997.

Masuka, B., Magorokosho, C., Olsen, M., Atlin, G. N., Bänziger, M., Pixley, K. V and Cairns, J. E. 2017. Gains in maize genetic improvement in Eastern and Southern Africa: II. CIMMYT open-pollinated variety breeding pipeline, *Crop Science, 57*(1):168-179.

Mercer, K. L., & Perales, H. (2019). Structure of local adaptation across the landscape: flowering time and fitness in Mexican maize (*Zea mays* L. subsp. mays) landraces. *Genetic resources and crop evolution, 66*, 27-45.

Mulatu, K., Bogale, G., Tolesa, B., Worku, M., Desalegne, Y., & Afeta, A. (1993). Maize production trends and research in Ethiopia. In 1. National Maize Workshop of Ethiopia, Addis Abeba (Ethiopia), 5-7 May 1992. IAR.

Piepho, H. P., & Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics, 177*(3), 1881-1888.

Rodrigues, P. C. (2018). An overview of statistical methods to detect and understand genotype-by-environment interaction and QTL-by-environment interaction. *Biometrical Letters, 55*(2), 123-138.

Shiferaw, B., Prasanna, B. M., Hellin, J., & Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food security, 3*, 307-327.

Smith, A. B., & Cullis, B. R. (2018). Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica, 214*, 1-19.

Smith, A., Cullis, B., & Gilmour, A. (2001a). Applications: the analysis of crop variety evaluation data in Australia. *Australian & New Zealand Journal of Statistics*, *43*(2), 129-145.

Smith, A. B., Cullis, B. R., & Thompson, R. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *Journal of Agricultural Science, 143*(6), 449–462. https://doi.org/10.1017/S0021859605005587

Solomon Admassu, S. A., Mandefro Nigussie, M. N., & Habtamu Zelleke, H. Z. (2008). Genotype-environment interaction and stability analysis for grain yield of maize (*Zea mays* L.) in Ethiopia.

Tesfaye, K., Alemu, T., Argaw, T., de Villiers, S., & Assefa, E. (2023). Evaluation of finger millet (*Eleusine coracana* (L.) Gaertn.) in multi-environment trials using enhanced statistical models. *Plos one, 18*(2), e0277499.

Wolde, L., Keno, T., Abebe, A., Abakemal, D., Terefe, W., & Negera, D. (2020). Genotype x environment interaction and stability analysis of grain yield in QPM hybrid varieties. *Maydica*, *64*(3), 10.

Wolde, L., Keno, T., Tadesse, B., Bogale, G., & Abebe, B. (2018). Mega-environment targeting of maize varieties using Ammi and GGE bi-plot analysis in Ethiopia. *Ethiopian Journal of Agricultural Sciences*, *28*(2), 65-84.

Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian journal of plant science, 86*(3), 623-645.

Yang, J., Benyamin, B., McEvoy, B.P., et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet., 42*, 565–569.