SPATIAL ANALYSIS IN MULTI-ENVIRONMENT TRIALS OF MALT BARLEY IN ETHIOPIA

D. TADESE, B. LAKEW 1 and G TAYE 2

Ethiopian Institute of Agricultural Research, P. O. Box 2003, Addis Ababa, Ethiopia
1 Holeta Agricultural Research Center, Ethiopia
2 School of Public Health, College of Health Science, Addis Ababa University, Ethiopia

Corresponding author: diribatadese@gmail.com

(Received 24 February 2019; accepted 19 August 2019)

ABSTRACT

Selection of superior genotypes and measuring heritability are some of the basic objectives of plant breeding. For this purpose, plant breeders grow crops across environments. Understanding the pattern of response across environments is an integral component of selection of superior and stable genotypes. The objective of this study was to improve selection strategies in barley breeding of Ethiopia through modeling spatial field trend. A set of multi-environment trials (MET) data from the national variety trial series conducted over four years, was taken from the Ethiopian Barley Breeding Programme, spanning stages from early generation to national variety trial testing for yield, was used in this study. The trials were analysed in a linear mixed model framework. Then, fitting a one-stage model for MET data, including a correlated spatial process for field trend within each trial, and combining a factor analytic (FA) model for genotype by environment interaction was conducted. The genetic correlations from this MET analysis were then used to cluster the environments based on their similarity. Performance of genotypes across these environmental clusters indicate broad (Bekoji-2005 and Bekoji-2004) and specific adaptation (Sgonder-2007 and Sgonder-2006) of genotype to certain types of environments. In addition, analysis of this historical MET data shed light on how breeding programme design can be improved to capture responses across the target population of environments, as it can inform the adequacy of the current number of barley grown areas in Ethiopia and the improvement in measuring heritability.

Key Words: Barley MET, heritability, linear mixed model

RÉSUMÉ

La sélection de génotypes supérieurs et la mesure de l’héritabilité font partie des objectifs fondamentaux de la sélection végétale. Dans ce but, les sélectionneurs de plantes font pousser des cultures dans tous les environnements. Comprendre le modèle de réponse dans les environnements fait partie intégrante de la sélection de génotypes supérieurs et stables. L’objectif de cette étude était d’améliorer les stratégies de sélection dans l’élevage d’orge en Ethiopie en modélisant la tendance des champs
D. TADESE et al.

Un ensemble de données d’essais multi-environnementaux (MET) de la série d’essais de variétés nationaux menés sur quatre ans a été tiré du Programme de sélection éthiopien de l’orge, qui couvre les stades de la première génération aux essais de variétés nationaux pour le rendement, a été utilisé dans cette étude. Les essais ont été analysés dans un cadre de modèle mixte linéaire. Ensuite, on a ajusté un modèle en une étape pour les données MET, y compris un processus spatial corrélé pour la tendance de terrain dans chaque essai, et combiné un modèle d’analyse factorielle (FA) pour une interaction génotype par environnement. Les corrélations génétiques de cette analyse MET ont ensuite été utilisées pour regrouper les environnements en fonction de leur similitude. La performance des génotypes de ces groupes environnementaux indique une adaptation large (Bekoji-2005 et Bekoji-2004) et spécifique (Sgonder-2007 et Sgonder-2006) à certains types d’environnements. En outre, l’analyse de ces données MET historiques a permis de mieux comprendre comment améliorer la conception du programme de sélection pour capturer les réponses dans la population visée d’environnements, car elle peut contribuer à l’adéquation du nombre actuel de zones de culture d’orge en Éthiopie et à l’amélioration de la mesure d’héritabilité.

**Mots Clés:** Orge MET, héritabilité, modèle mixte linéaire

---

**INTRODUCTION**

The aim of plant breeding is most often to select either high performing genotypes for target environments or stable genotypes across a given set of environments. Breeders are also interested in stable genotypes, given that such genotypes are among the winner genotypes. But the challenge is to get high performing stable genotypes, mainly due to interactions between genotypes and environments. Therefore, in order to achieve this objective, breeders usually design trials across several locations within a given agro-ecology and for a number of years. Several statistical techniques have been proposed to facilitate estimation and interpretation of genotype by environment interactions (Welham et al., 2010).

The classical statistical techniques mostly used for analysis of multi-environment trial (MET) is the two cross classifications (genotype by environment interaction) of Analysis of Variance (ANOVA). However, while this technique can adequately explain only the main effects, and identify genotype environment interactions as a source of variation, it fails to decompose and analyse the inherent interaction effects. This is due to the additive nature of the ordinary ANOVA model, which does not allow for analysis of non-additive interaction components. Therefore, other statistical approaches are required to identify the pattern and complexity of interaction. Oakey et al. (2007) compared the performance of the ANOVA method with the regression method, and found that ANOVA fails to detect significant interaction components; and the regression approach accounts for only a small portion of the interaction sum of squares, only when the pattern fits a specific regression model.

MET is usually analysed using a two-stage approach, in which variety means are first estimated separately for each trial and then combined to form the data for an overall analysis. The latter methods include mixed effect models (Talbot, 1984; Smith et al., 2005) and the fixed effects AMMI (additive main effects and multiplicative interactions) model (Welham et al., 2010). The two-stage approach is an approximation of the combined analysis of the raw plot data from all trials. If there is error variance heterogeneity between trials and spatial variation or unequal replication within trials, the approximation may be poor in estimation by classical ANOVA. Smith and Cullis (2001) presented a weighted mixed model for the second-stage analysis that aimed at accommodating these sources of error variation, thereby reducing efficiency losses.
The superior approach, however, is the spatial MET analysis of Kelly et al. (2007), in which individual plot data are analysed and a separate spatial covariance structure and error variance allowed for each trial.

Spatial variation within trials and error variance heterogeneity between trials occurs routinely, and if unaccounted for, will result in biased and inefficient estimates of variety effects. Historically, many authors assume common error of variance within trials; which is inappropriate for all trials. In MET data analysis, usually there is interest in the analysis of both overall performance of each variety (across all environments) and whether variety performance is affected by these interaction effect (genotype by environment interaction).

The approach of multiplicative models for the interaction effect allows for a separate genetic variance for each trial, and provides a parsimonious, and interpretable model for the genetic covariances between pairs of trials. Then, the genetic model can be regarded as a random effects analogue of the AMMI model. Hence, this approach combines the strength of AMMI with the advantages afforded by the mixed model framework.

Therefore, in this study, a linear mixed effect model was used to analyse MET data, with the specific objective of clustering barley testing environment based on the correlation (the way in which the locations respond to each genotype) of the trials, to show advantages of using advanced statistical methods for precise estimation and improvement in heritability in barley breeding programme in Ethiopia. So, to achieve this objectives, the field trend at each trial was included in LMM to capture spatial field trend to increase precision and accuracy of the estimation. Then, at the second stage of analysis under LMM, combining the trial through Factor Analytic (FA) model to capture correlated trial, as well as clustering of each trial, based on their correlations.

MATERIALS AND METHODS

The data used for this study were taken from the malt barley breeding programme of the Ethiopian Institutes of Agricultural Research (EIAR). The trial was grown across different environments in Ethiopia; while the environment was considered as the combination of years and locations. The trials included Preliminary Variety Trial (PVT in 20014), National Variety Trials (NVT1 in 2005, NVT2 in 2006 and NVT3 in 2007) across five locations (Adet, Asasa, Bekoji, Holeta and Sgonder) for four consecutive years (2004-2007) of environments. The preliminary variety trial in 2004 of Adet was excluded from the combined analysis, due to zero genetic variance and this occurred since high damage of the trial in this specific year. Hence, 19 trials were included in this study with their respective row, column and genotypes in each trial. No missing plot was observed in all of the trials.

At PVT level the number of genotypes included in the study were different from that of NVT since the NVT genotypes were selected from their previous PVT trials, and this is common in plant breeding programmes, since a high yielder genotype is promoted for the next trial for further evaluation. Therefore, 49 genotypes were tested at PVT; while 15 were selected for NVT1, NVT2 and NVT3. All trials were designed as Randomised Complete Block design (RCBD), with four replication.

Modeling of field trend in MET. Most often, field experiments are laid out in a rectangular form like jth trials j= 1, . . . p, which consist Nj plots with rj rows and cj columns (Nj = rj x cj) (Smith and Cullis, 2001). Therefore, the vector \( y_j^{(Nj \times 1)} \) is sorted accordingly rows with in columns. Hence, the combined vector of data across all environments can be modeled as:
\( y^{(n \times 1)} = \{y_j\},\ n = \sum_{j=1}^{p} N_j \) is given by

\[ y = X \tau + Zu + e \] ................................ (1)

Where:

\( \tau^{(tx1)} \) and \( u^{(bx1)} \) are vectors of fixed and random factors, respectively; while \( X^{(n \times r)} \) and \( U^{(b \times 1)} \) be associated design matrices for fixed and random effects with the former assumed to be of full column ranks and \( e \) is vector of the residual. The joint distribution of \( (\tau', u')' \) is assumed to be:

\[
\begin{bmatrix}
G(\tau) & 0 \\
0 & R(\theta)
\end{bmatrix}
\]

Where:

\( \gamma \) and \( \theta \) are vectors of variance parameters.

Then, the distribution form of the vector data \( y \), is a Gaussian distribution with mean \( X \) and variance matrix \( H = ZGZ' + R \). In addition, the error term contains a vector of sub-error \( \{\epsilon_j\} \), where is vector of plot errors for a given \( j^{th} \) trial and decomposed into a spatially dependent process; while the independent white noise process is \( \eta_j \). The matrix of the error term for trial \( j \) can be written as:

\[ R_j = \sigma^2 \sum_j (\alpha_j) + \sigma^2_{\eta} I_{N_j} \]

Where:

\( \Sigma_j \) is the matrix of spatial correlation which is a function of \( \alpha_j \) with associated variance \( \sigma^2 \); while \( \sigma^2_{\eta} \) is variance parameters of the white noise process.

The spatial process \( \eta \) is assumed to be the second order stationary, given that the correlation between plot depends on the lag distance. Furthermore, the column and row dimensions are separable. Hence, one can write:

\[ \Sigma_j = \Sigma_{cj} \otimes \Sigma_{rj} \]

Where:

\( \Sigma_{cj} \) and \( \Sigma_{rj} \) are the two dimensional correlations respectively.

However, a number of research findings show that the first ordered autoregressive which is denoted by \( AR1 \times AR1 \), most often provides an adequate variance structure for local spatial trend (Smith et al., 2001).

The random effect \( u \) consists of sub-vectors \( \{\mu_i\} \), where \( \mu^{(bx1)} \) is the vector of effect for the \( i^{th} \) random term, \( i=1, \ldots, q \). The matrix \( Z \) is partitioned conformably as \( [Z_1 \ldots Z_q] \). It was assumed that the sub-vectors of \( u \) were mutually independent. Variance matrix \( G_i \) for the \( i^{th} \) random term has many possible forms including the standard variance component structure:

\[ G_i = \sigma^2 \cdot I_{b_i} \cdot I_{e_i} \cdot I_{v_i} \]

Let \( u_g \) be the \( mp \times 1 \) vector of genetic effect for \( m \) varieties for each \( p \) environments, ordered as varieties within environments. It represents a two dimensional (varieties by environment) arrays of effect, namely, where \( U_g = vec(U_g) \). It was assumed that the associated variance structure had separable forms with:

\[ var(u_g) = G_e \otimes G_v \]

Where:

\( G_e \) and \( G_v \) are the symmetric \( p \times p \) and \( m \times m \) component matrices for environment and varieties, respectively. When \( G_e = I_e \), just for simplicity, therefore, \( var(u_g) = G_e \otimes I_m \) and the matrix \( G_e = \{\sigma_{ij}\} \) is the so called genetic variance matrix. The diagonal elements are genetic variance for individual environments, and the off–diagonal elements are genetic covariance between pairs of environments.
Spatial analysis in multi-environment trials using linear mixed model

The spatial mixed model for MET can then be written as:

\[ y = X_\tau + Z_\nu + e \]
\[ = X_\tau + Z_0u_0 + Z_\nu u_\nu + e \]  \hspace{2cm} (2)

The fixed effect \( \tau \) includes environmental main effects and trial specific effects for extraneous field variation (Gilmour et al., 1997). \( \nu \) includes variety effects at each environment with associated design matrix \( Z_\nu^{(n x mp)} \) and \( \mu \), comprising of an additional random effect, with design matrix, and variance matrix.

In breeding programmes, there are many possible forms of genetic variance matrix structures. Under the linear mixed model, the standard structure is given by:

\[ G_\nu = \sigma_v^2 I_p + \sigma_{ve}^2 I_p \]

Where:

- \( \sigma_v^2 \) and \( \sigma_{ve}^2 \) are the variance components for variety main effects and interaction effects, respectively;
- \( j_p \) is a \( p \times p \) matrix of ones.

This implies that all environments had constant genetic variance and all pairs of environment had the same genetic covariance. In this case, since inefficient estimation variance covariance, Smith and Cullis (2001) consider an alternative variance structure model; which is known as Factor Analytic model; which is analogous of AMMI model.

In multi-environment trial data, this model captures the nature of heterogeneous variance covariance. The Factor Analytic (FA) model is a regression-type model \( y=ax+b \), which can be fitted for an increasing number of dimensions, \( k \). The model for factor analytic is given as:

\[ u_\nu = (\Lambda_v \otimes I_m)f_0 + \delta \]
\[ = (\sigma_v 1_p \otimes I_m)f_0 + (\Lambda \otimes I_m)f + \delta. \]

Therefore, in this study, the first spatial field trend and global variability (variability from one corner to the other in the experiment) were considered for each specific trials, and considered as the first stage of analysis. Then, combined analysis followed through FA model, considering all information at individual trials through Linear Mixed Model (LMM).

RESULTS AND DISCUSSION

A visual display of residual variation before spatial adjustment and residual variation after adjustment are presented in Figure 1a and b. In large field trials, field variations are a substantial source of error, since neighbouring plots show similar characteristics compared to those far apart. Therefore, unless accounted for, this spatial variability can bias the estimated potential of a given genotype, when compared with other candidate genotypes. Hence, neighbour lines of plots was used to model the systematic field trends to minimise bias, and reduce error to increase precision in estimation for some trials in this study. Figure 1a indicates areas of high and low yields in the field. It also shows that neighbouring plots tended to be more similar than those far apart. Thus, there is need to include spatial correlation in variance-covariance analysis, to handle field trend that extend from one point to the other at the plot level.

This study identified the relative genetic merits of different lines or hybrids where trials are correlated. When trials are correlated (similar response of genotypes at some environment) selecting best materials in a given environment is the same as selecting
Figure 1. Residual variation in field trial before spatial adjustments and after spatial adjustments.
Spatial analysis in multi-environment trials using linear mixed model

best material in another environment. Then, information from multiple environment can be combined to improve accuracy of genetic gains of particular trials. In this case, MET analysis can also help to understand the broad and specific adaptation of genotypes over a range of target populations.

In this study, the factor analytic model was used for MET data analysis. Then, the correlations between environments ranged from -1 and 1. Correlations of -1 indicates that the performance of the environments fall in opposite direction (the angle between the two environment is more than 90 degree), implying that the highest performing genotypes in one environment were the lowest performing genotypes in other environment.

Correlation of +1 is an indication of perfect similarity between two environments, hence selection of superior genotypes based on one environment is the same as selection for another environment. In Figure 2, for instance, Holeta-2004 had nearly 0 correlation with Adet-2005, Asasa-2006, Bekoji-2007, Sgonder-2004 and Asasa-2005; while Holeta-2006 with Holeta-2007 had nearly perfect correlation. But sometimes, there may be indication of none or even reverse correlations between trials within the same locations, but for different years due to seasonal effect or other factors; for example Holeta-2004 with Holeta-2005 (Fig. 2).

Generally, correlations ranged from negative to positive (Figs. 2 - 4) both sides.

Figure 2. Heatmap to show patterns of genetic correlation in barley experiment of fifteen trials in Ethiopia.
For example, Bekoji-2004 had negative correlation with Adet-2005; hence selection of best genotypes based on one trial may be lower in the other trials. Holeta-2004 had no correlation with Adet-2005, Holeta-2006 with Holeta-2007. Therefore, it is important to classify the trial for selection strategies into different clusters based on the correlation structure in a trials.

Graphical description of MET data is commonly used to explain genotype by environment. In Figure 3, the concepts of genotypes won and where is illustrated. Plots show that the environment with longest line from the centre measures the discriminativeness of that environment when compared with others. For example, Holleta-2007, Bekoji-2006, Adet-2006 and Asasa-2006 were among the most discriminative environments; this means that environments had considerable contributions in discriminating genetic variations. On the other hand, environments with less distances from the centre were those stable environments, hence they explained less genetic variations. In addition, when a specific genotype is close to a given environment, it indicates that the genotype is the winner for that specific environment. That means, that this genotype is the best performer for that trial. For example, EH1847 was the winner genotype at Bekoji-2006, Asasa-2007, and also in other environments.

Figure 4 is about correlations between the trials. Environments with less angle (less than 90 degree) between their two lines are more
correlated. Then, selection of the best genotypes based on one of the two environments does not change the ranks of the genotypes in another environment, for example Holeta-2006 and Bekoji-2004. When an angle between two lines is greater than 90 degrees, selection based on the two environments can change the ranks of genotypes; for example, Asasa-2004 and Sgonder-2006. It is also important to cluster environments based on their correlation. According to Figure 5, two large clusters are evident; but such type of clustering depends on the required homogeneity in the clusters.

Heritability analysis is another objective of breeding programmes, which requires specialised statistical analysis methods for best selections. Heritability can be estimated more precisely through appropriate statistical methods, without additional costs. Figure 6 indicates improvement in heritability generated from different statistical models; including heritability in classical RCBD, in spatial modeling and in spatial plus correlated trials (MET). In Figure 6, the horizontal axis is the list of barley trials; while the vertical line is the percent of heritability for each trial in different statistical models. In most of the trials, there was an improvement in heritability when spatial plus correlated trials were modeled simultaneously. This means that if there is a spatial field trend in the trial, heritability improves through modeling field trends. Also, in MET there is improvement in heritability since it borrows information from different trials.

It is well known that the reliability of varietal selection from a single trial is greatly improved with the use of sound design and analysis techniques. Consideration of spatial trend in field experiments increased gains in accuracy and precision, as this study indicated,
Figure 5. Dendrogram to show clustering of 19 malt barley trials in Ethiopia.
Spatial analysis in multi-environment trials using linear mixed model

Figure 6. Improvement in heritability only through use of advanced statistical models.

for which also many authors recognised the need to model spatial trend within a field trial (Cullis et al., 2006).

If not considered, spatial variability can bias the selection of genotypes; implying that genotypes are selected only due to the plot where they are planted. For example Figure 1a indicates potential existence of field trends, which is usually expected in large trials. Hence, for selection purposes, adjustment for spatial variability is important. The adjustment is simultaneously fitted under linear mixed model, with other source of variability at a time.

The Factor Analytic (FA3) model was used to indicate the correlation between trials. In Figure 2, there are environments with positive correlations such as Adet-2005, Sgondor-2007, Adet-2006, Holeta-2007, with Holeta-2005. Environments with negative correlations include Asasa-2004 with Adet-2005 and Holeta-2005; and Bekoji-2004 with Adet-2005 and Holeta-2005.

This study showed which genotype was the winner and in which environments, as well as the correlations between environments based on angles of the two environmental axes (Figs. 3 and 4). Discriminative environments are also identified in this study (Fig. 3); this means environments with high interaction with genotypes. Environments with smaller distances from centre are indications of stable environment. Performance of each genotype is determined by the closest environment, and one genotype can be closer to one or more environments; that means this genotype is superior in these environments; an observation that agrees with the ideas of Yan and Tinker (2006).
TABLE 1. Summary of malt barley trials including rows, column and number of genotypes

|            | Rows | Columns | Genotypes | Mean yield |
|------------|------|---------|-----------|------------|
| Adet-2005  | 15   | 4       | 15        | 2.11       |
| Adet-2006  | 15   | 4       | 15        | 4.84       |
| Adet-2007  | 15   | 4       | 15        | 4.31       |
| Asasa-2004 | 7    | 14      | 49        | 2.74       |
| Asasa-2005 | 15   | 4       | 15        | 3.95       |
| Asasa-2006 | 15   | 4       | 15        | 1.64       |
| Asasa-2007 | 15   | 4       | 15        | 1.58       |
| Bekoji-2004| 7    | 14      | 49        | 4.16       |
| Bekoji-2005| 15   | 4       | 15        | 4.48       |
| Bekoji-2006| 15   | 4       | 15        | 3.43       |
| Bekoji-2007| 15   | 4       | 15        | 1.65       |
| Holeta-2004| 7    | 14      | 49        | 3.39       |
| Holeta-2005| 15   | 4       | 15        | 3.3        |
| Holeta-2006| 15   | 4       | 15        | 2.85       |
| Holeta-2007| 15   | 4       | 15        | 2.85       |
| Sgonder-2004| 7   | 14      | 49        | 2.95       |
| Sgonder-2005| 15  | 4       | 15        | 1.16       |
| Sgonder-2006| 15  | 4       | 15        | 4.22       |
| Sgonder-2007| 15  | 4       | 15        | 3.89       |

Based on the closeness in terms of discriminating the genotypes, the 19 malt barley environments were clustered into two mega-environments using a dendrogram (Fig. 5). There is high improvement in heritability through use of spatial plus MET analysis (Fig. 6). The improvement in heritability under the different statistical methodologies are illustrated in Figure 6; for example heritability gained under classical RCBD, heritability under spatial analysis and heritability under spatial plus MET through use of correlated environments are compared. In general, there is high improvements of heritability through modeling spatial field trend with MET under linear mixed effect model.

CONCLUSION

This study underscores the importance of spatial correlation plus the use of the Factor Analytic model to account for structural correlation among testing environments in Ethiopian malt barely breeding programme. Maximum likelihood estimations of linear mixed model is more appropriate and applicable for the analysis of MET, since it can handle and fit simultaneously all source of variations (extraneous variations, random variations and trend variations) at a time. In addition to select best performing genotype, it is better to consider the correlation between environments and cluster them accordingly. Smith and Cullis (2001) also strongly suggested the need of considering field trend in experiment for precise estimation.

ACKNOWLEDGEMENT

Thanks to Ethiopian Institutes of Agricultural Research (EIAR) especially Barley Breeding Programme who provide the data used in this paper.
REFERENCES

Cullis, B.R., Smith, A.B. and Combes, N.E. 2006. On the design of early generation variety trials with correlated data. American Statistical Association and the International Biometric Society. Journal of Agricultural, Biological, and Environmental Statistics 11(4):381–393.

Gauch, H.G. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Amsterdam: Elsevier.

Kelly, A., Smith, A., Eccleston, J., Cullis, B. 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Science 47:1063–1070.

Oakey, H., Verbyla, A., Cullis, B., Pitchford, W. 2007. Joint modelling of additive and non-additive (genetic line) effects in multi-environment trials. Theoretical and Applied Genetics 114:1319–1332.

Smith, A.B., Cullis, B.R. and Thompson, R. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138–1147.

Smith, A.B. and Cullis, B.R. 2001. The analysis of crop variety evaluation data in Australia. Australian and New Zealand Journal of Statistics 43:129-145.

Smith, A.B., Cullis, B.R. and Thompson, R. 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Sciences 143:1–14.

Talbot, M. 1984. Yield variability of crop varieties in the U.K. Journal of Agricultural Science, Cambridge 102:315-321

Welham, S., Gogel, B., Smith, A., Thompson, R., Cullis, B. 2010. A comparison of analysis methods for late-stage variety evaluation trials. Australian and New Zealand Journal of Statistics 52:125–149.

Yan and Tinker. 2006. Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, 960 Carling Ave., Ottawa, Ontario, Canada K1A 0C6.