Spatial Field Trend Analysis and Factor Analytic Model Under Linear Mixed Model in Multi-Environment Trial Data: The Case of Durum Wheat in Ethiopia

Diriba Tadese[1] and Mekuria Dejene[2]*

[1]*Ethiopian Institutes of Agricultural Research, Ethiopia*

[2]*Debre zeit Agricultural Research Center, Ethiopia*

ABSTRACT

In multi-environment trial (MET) data analysis, breeders are usually interested in performance of genotype across target environment (broad adaptation) and performance of genotype at individual environment (specific adaptation). Hence, the main objective of this study was to predict genetic potential of durum wheat across environment and to observe improvement in heritability across different statistical methods. MET data of durum wheat was taken from Debre zeit Agricultural Research Center (DzARC) of Ethiopian Institutes of Agricultural Research (EIAR) which was planted in 2011 and 2012 year of main season. Then, for the data analysis the standard linear mixed model (LMM) is applied using ASREML packages of R software considering spatial field trend at individual trial and factor analytic model among trial. The result of this study (Table 1) show that on average, DSP2009-Off.F4.1H.378-meh.4H.187 genotype perform better than all other genotypes while, the performance of UC1113GPCLR198001/59 genotype is less compared to other genotypes. In addition, when looking at improvement in heritability (Figure 1) across three different statistical methods spatial field trend plus MET together indicate high improvement in heritability than the classical methods.

KEYWORDS: Durum wheat; Heritability; MET; LMM

INTRODUCTION

Durum wheat is one of the widely grown crop and economic advantages in Ethiopian. Hence, lots of research has been done to improve seeds and quality of durum wheat Hailu 1991. Accordingly, 40 improved durum wheat has been released for commercial use since 1996. Potentially, durum wheat is grown at the highlands (1800-2800 meters) of Ethiopia undervert soil Tesemma [1]. Currently, the production and productivities of the durum wheat could be 80% of the bread wheat. Historically, the yield level of durum wheat in Ethiopia is about 80% of that of bread wheat, which has been attributed partly to less favourable crop growing environments and management practices Srivastaval et al. [2]. However, new high yielding semi-dwarf durum have been developed that have yield potential equal, or even superior to the highest yielding bread wheat in some areas Abinasa et al. [3]. As the price of durum is often higher than that of bread wheat, it is a promising and viable alternative cash crop for farmers. Statistical analysis of multi-environment trial (MET) data currently increasing from time to time in plant breeding to precision and accuracy in estimating genetic potential. In MET data analysis, there are several information about variation in each specific trial like spatial field

trend, extraneous variation and global variability Damesa et al. [4]. In combined analysis, factor analytic (FA) model is the most powerful in MET data analysis in capturing correlation among trials and how this correlation would affect genetic selection. Routinely, field spatial variability within a trials and heterogeneity between trials usually occurs in field experiment and need to be accounted unless it will result in biased and inefficient estimates of genetic effect Tadese et al. [5]. In a classical statistical analysis, historically authors usually consider a common error in experimental field but, which is sometimes not correct.

Table 1: Summary statistics of durum wheat across ten trials in Ethiopia.

Trials	Number of Rows	Number of Columns	Number of Genotype	Mean Yield
AK_NVTOH2012	18	4	18	1.12
AK_PVTOH2011	23	3	23	0.66
CD_NVTOH2012	18	4	18	3.22
CD_PVTOH2011	23	3	23	2.58
DN_NVTOH2012	18	4	18	2.23
DN_PVTOH2011	23	3	23	2.94
DZ_NVTOH2012	18	4	18	3.48
DZ_PVTOH2011	23	3	23	2.38
MJ_PVTOH2011	23	3	23	2.2
RB_PVTOH2011	23	3	23	4.39
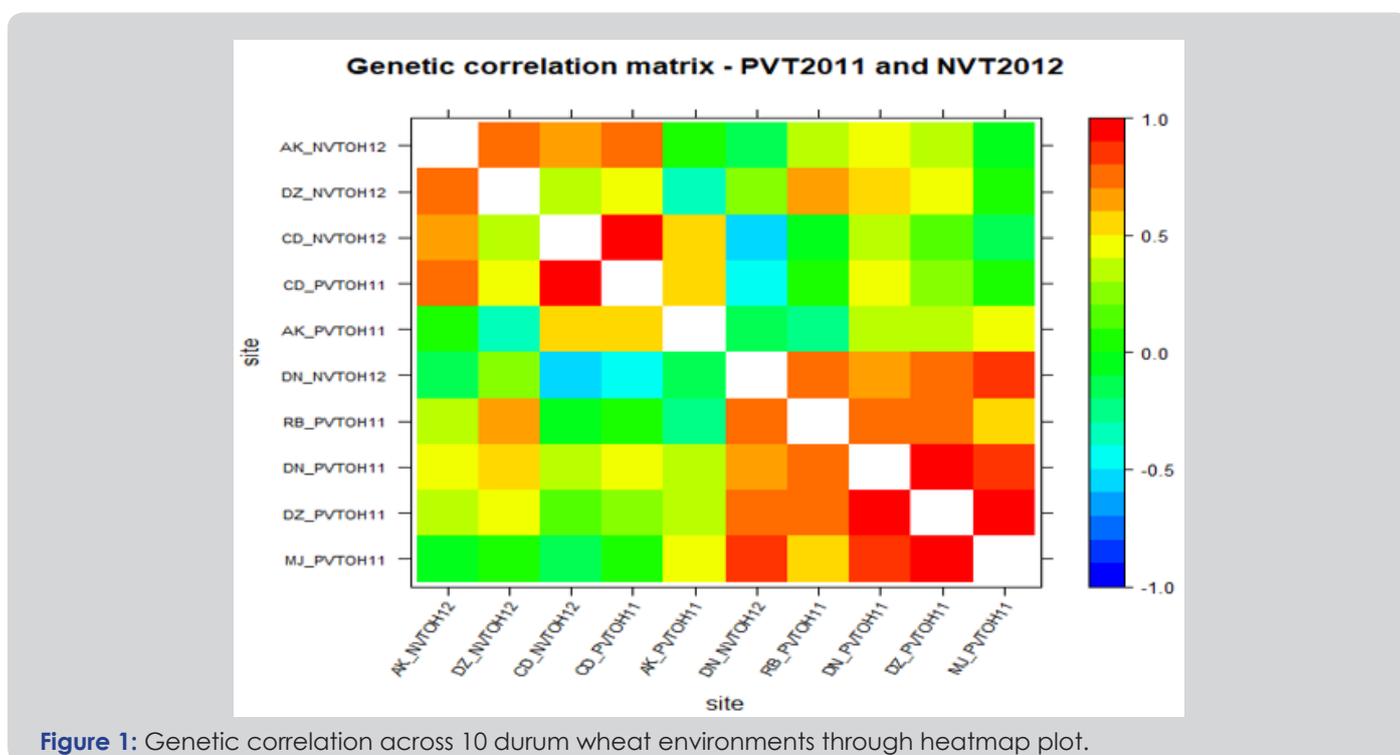
Figure 1: Genetic correlation across 10 durum wheat environments through heatmap plot.

Breeders are mostly interested in overall performance of genotypes across target environment as well as specific adaptations of each genotype when conducting MET data analysis.

The techniques of genotype by environment interaction effect models permit for separate genetic variance for individual trials and provides parsimonious, and interpretable model for the genetic covariances between pair of trials.

Therefore, in this study two phase of statistical analysis is employed by using linear mixed model (LMM). In the first stage, individual trials are analysed for the spatial field trend, extraneous variation and global variability. In the second stage, using information from each trial, the trials are combined through factor analytic model (FA) and included in the standard LMM.

MATERIALS AND METHODS

Data used for this study was taken from Debre zeit Agricultural Research Center (DARC) of the Ethiopian Institutes of Agricultural Research (EIAR). The experiment was conducted in 2011 and 2012 in randomized complete block design of 23 genotypes in 2011 and 18 genotypes in 2012.

The layout of an experiment is a rectangular arrangement like j^{th} trials j=1. . . p, with Nj, plots consist of r_j, rows and c_j columns s that $N_j = r_j \times c_j$; Smith [6]. Then, for the analysis of field trend, the vector of the response variable $y_j^{(NjX1)}$ would be sorted with rows within columns for all environments before analysis.

Then, the standard linear mixed model approaches employed would be computed as,

$y^{(nx1)} = \{y_j\}$, n $= \sum_{j=1}^{p} N_j$ is given by

y = X+ Zu+ e ----------------------------(1)

where $\tau^{(tx1)}$ and $u^{(bx1)}$ are vectors of fixed effect and random effect factors respectively where as $X^{(nxt)}$ and $U^{(bx1)}$ be associated design matrices for fixed and random effects while the former one assumed to be of full column ranks and e is vector residual. Therefore, the joint distribution of (u', e') assumed to be,

$$\begin{bmatrix} G(\gamma) & 0 \\ 0 & R(\emptyset) \end{bmatrix}$$

Where γ and ∅ are variance parameters. Then, the distribution of y would be, Gaussian distributions with mean Xτ and variance matrix H= $ZGZ' + R$.

In addition, the error term contains vectors of sub error {ej}, where $e_j^{(Njx1)}$ is vector of plot errors for a given jth trial and decomposed into a spatially dependent process ξ_j while independent white noise process η_j. The matrix of the error term for trial j can be extended as $R_j= \sigma_j^2 \sum_j(\alpha j) + \sigma_{\eta j}^2 I_{Nj}$,

where Σ_j is the matrix of spatial correlation associated with α_j with variance σ_j^2 while $\sigma_{\eta j}^2$ is variance parameters of the white noise process. The spatial process ξ_j is assumed to be the second order stationary given that the correlation between plot depends on the lag distance given that the columns and row dimensions are separable. Hence, one can write $\Sigma_j = \Sigma_{cj} \otimes \Sigma_{rj}$, where Σ_{cj} and Σ_{rj} are the two-dimensional correlations respectively. Number of research findings Oakey [7]; Tadese et al. [5] show that the first order autoregressive which can be denoted by ar1xar1 explain an appropriate variance structure in field trend.

u is a random factor with the sub vectors of $\{u_i\}$, where $u_i^{(bix1)}$ is the vector of effect for the ith random term, i=1...q. The matrix Z therefore can be partitioned in to [Z1 ...Zq] and the sub vectors of u are assumed to be mutually independent. But the variance matrix G$_i$ for the ith random term has many possible forms including the standard variance component structure G$_i$ = $\sigma_i^2 I_{bi}$.

In a plant breeding program, breeders want to measure how well total genetic effects can be predicted through stability analysis and this requires special statistical methods. This describes the proportion of the genetic variance to the total variance. If heritability is low, then little advance can be made from phenotypic selections while high in heritability best lines and parents can be selected for future crossing so that further enhancing genetic gain in the breeding program.

Let u$_g$ be the mpx1 vector of genetic effect for m varieties for each p environments ordered as varieties within environments. It represents a two-dimensional (varieties by environment) array of effect, namely $U_g^{(mxp)}$, where $u_g = vec(U_g)$. Assuming, the associated variance structure has separable form with $var(u_g) = G_e \otimes G_v$, where G_e and G_v are the symmetric p x p and m x m component matrices for environment and varieties, respectively. When $G_v = I_m$, just for simplicity, therefore $var(u_g) = G_e \otimes I_m$, and the matrix $G_e = \{\sigma_{jj'}\}$ is the so-called genetic variance matrix. Therefore, the standard linear mixed model can be extended as;

$$y = X_\tau + Z_u + e$$

$$= X_\tau + Z_0 u_0 + Z_g u_g + e$$, ----------------(2)

The fixed effect τ includes environmental main effects and trial specific effects for extraneous field variation Oakey et al. [7].

u_g is variety effects at each environment with associated design matrix $Z_g^{(nxmp)}$ and u_0 comprise an additional random effect with design matrix Z_0.

Therefore,

$$u_g \sim N(0, \sigma_g^2 I_m)$$

$$u_i \sim N(0, \sigma_i^2 I_{bi})$$

In multi-environment trial (MET) data analysis, there are many possible form of genetic variance matrix structures while using linear mixed model and the standard structure is given by $G_e = \sigma_v^2 J_p + \sigma_{ve}^2 I_p$, where σ_v^2 and σ_{ve}^2 are the variance components for variety main effects and interaction effects respectively, J_p is a p x p identity matrix. This implies that all environments have constant genetic variance and all pair of environment has the same genetic covariance. In this case, since inefficient estimation, Piepho [8]. consider an alternative variance structure model which is known as Factor Analytic model which is analogous of AMMI model. In addition, this model captures the nature of heterogeneous variance covariance structures.

The model for factor analytic is given as

$$u_g = (\Lambda_v \otimes I_m)f_v + \delta$$

$$u_g = (\Lambda_v \otimes I_m)f_v + \delta$$

Where $\Lambda_v^{p(k+1)} = [\sigma_v 1_p \ \Lambda]$,

$$f_0 = u_v/\sigma_v$$

$$f_v' = (f_0', \ f')$$.

While fitting linear mixed model in this study, spatial field trend fitted first for each environment and tested for the potential existence of field trend between the neighbor plots. Furthermore, global variability and extraneous variation has checked and included into the standard linear mixed model. Finally, trial across environment is combined keeping their specific trial information like spatial field trend and included in a LMM through factor analytic model.

RESULTS

The summary statistics (Table 2) indicate that the average performance of all genotypes at RB_PVTOH2011 is higher (4.39 t/ha) than other trials while the potential of AK_PVTOH2011 trial is less than all others. Furthermore, the number of rows and columns used for spatial field trend analysis are included in Table 1 with the number of genotypes in 2011 and 2012.

In multi-environmental breeding program, looking at performance of each genotype and the rank change across testing environments is so important for selection. Therefore, the response of these genotypes in respective environment goes to determine selection of genotype for the next trial or for release. Hence, when trials are correlated, ranking of genotypes similar so that the one best performing genotype/s at specific environment has similar performance with the one highly correlated environment.

In Figure 1 MJ_PVTOH2011 has strong positive correlation with DZ_PVTOH2011, DN_PVTOH2011 and DN_PVTOH2012 as an example. So, if a researcher selects the best material based on MJ_PVTOH2011, it will be similar selection for DZ_PVTOH2011, DN_PVTOH2011 and DN_PVTOH2012. But if environments are negatively correlated, selection of best genotype is little

bit challenging since high rank change among genotype from environment to environment. That means, the one with good performing in one environment will be poor performing genotype in another environment with the one negatively correlated. In addition, when the environments not correlated either negatively or positively, they independent environment so selecting best material in one environment has an average performance for another environment with the independent environment. The correlation among trials can also be plot using biplot graphical analysis (Figure 2) in which the angle between two lines indicate the strength of correlation between two environments. Less angle indicates high positive correlation approximately less than 90' while near 90' between the angle indicate the environments are independent. If the angle between the line is greater than 90', then the environments are more negatively correlated. In field experiment, heritability analysis is also one objective in plant breeding program. High in heritability implies high in selection since it measures how accurate we predict in a total genetic effect.

In case of low heritability, few advancements can be made from phenotypic selection a plant breeding. Therefore, heritability calculation across three statistical methods is employed in this study. High improvement in heritability observed through the use field trend spatial analysis plus MET data analysis as indicated in Figure 3. In MET improvement in heritability comes from correlated trial that means while calculating heritability it uses information from the one correlated with it. Predicted value in Table 2, indicate that DSP2009-Off.F4.1H.378-meh.4H.187 genotype perform better (2.92 t/ha) than the rest of candidate materials on average while relatively UC1113GPCLR198001/59 genotype perform lower (2.075 t/ha) than the others on average. Generally, most genotypes perform better than the local check which is 2.52 t/ha as indicated in Table 2. When looking at potential environment, RB_PVTOH11 environment perform better than other environment with 4.39 t/ha yield and in reverse environment AK_PVTOH11 is low yielder (0.65 t/ha) when compared with others.

Table 2: Predicted values of each genotype across each environment of durum wheat in Ethiopia.

List Genotypes	AK_NVTOH12	AK_PVTOH11	CD_NVTOH12	CD_PVTOH11	DN_NVTOH12	DN_PVTOH11	DZ_NVTOH12	DZ_PVTOH11	MJ_PVTOH11	RB_PVTOH11	Grand Total
DSP2009-0FF.f3.2H-22-Meh.1H.26	1.100043	0.584609	3.214071	2.593866	2.194883	3.054085	3.467246	2.429289	2.308702	4.285916	2.523271
DSP2009-0FF.f4 .1H.378-Meh.4H.187	1.291736	1.104221	3.517416	2.968217	2.02147	3.222147	3.710237	2.480105	2.30262	4.387972	2.700614
DSP2009-0FF.F4 .1H-783-Meh-4H.259	1.101666	0.838815	3.264549	2.655242	2.123601	3.000478	3.47789	2.384809	2.188604	4.167063	2.520272
DSP2009-0FF.F4 .1H-785-Meh-2H.262	1.117186	0.595736	3.055899	2.402219	2.449392	3.28195	3.455764	2.607531	2.628847	4.796502	2.639103
DSP2009-0FF.F4 .1H-977-Meh-1H.293	0.734814	0.351861	3.045319	2.377774	1.928995	2.263868	3.077623	1.948886	1.584716	3.644062	2.095792
DSP2009-0FF.f4 .2H.712-Meh.1H.248	1.065168	0.45828	3.148221	2.512806	2.241969	3.035755	3.421127	2.430038	2.28229	4.896405	2.549206
DSP2009-0FF.f4 .2H.735-Meh.2H.251	1.113268	0.695108	3.497452	2.938553	1.800475	2.764227	3.531051	2.186775	1.841628	4.288965	2.46575
DSP2009-0FF.f4 .3H.639-Meh.4H.240	1.119387	0.564015	3.380606	2.796779	1.979317	2.914922	3.516129	2.307003	2.023759	4.444761	2.504668
DSP2009-0FF.f4 .3H.976-Meh.2H.292	1.078331	0.46447	3.270945	2.662306	2.081581	2.93026	3.456075	2.338013	2.118706	4.369491	2.477018
DSP2009-0FF.F6/Off/1508/2009	1.128595	0.794773	3.373853	2.788854	2.002062	2.94749	3.523979	2.329435	2.096032	3.943274	2.492835
DSP2009-F6/Off/1508/2009	1.039845	0.651341	3.257672	2.645013	2.046993	2.841811	3.390568	2.283163	2.051633	4.243622	2.445166
DSP2009-F9/0FF/1615/2009	0.749259	0.516317	2.861427	2.154798	2.217202	2.514006	3.058877	2.14652	1.916748	4.25629	2.239144
DSP2009-Off.F3.2H.22-meh.1H.26	1.316874	0.812659	3.492082	2.938201	2.093611	3.318974	3.768758	2.547805	2.391087	4.579241	2.725929
DSP2009-Off.F4.1H.378-meh.4H.187	1.488003	0.894019	3.56861	3.036364	2.221897	3.692152	3.899293	2.775282	2.716788	4.892244	2.918465
DSP2009-Off.F4.1H.783-meh.4H.259	1.095965	0.616708	3.053909	2.399159	2.422555	3.22706	3.425528	2.572293	2.547313	4.700511	2.6061
DSP2009-Off.F4.1H.785-meh.2H.262	0.997192	0.600938	3.126157	2.483939	2.178866	2.877948	3.5067	2.331867	2.16151	4.339651	2.460477
DSP2009-Off.F4.2H.712-meh.1H.248	0.912621	0.560788	3.088551	2.435688	2.115156	2.69328	3.392788	2.219249	2.000187	4.184636	2.360294
DSP2009-Off.F4.2H.735-meh.2H.251	1.034134	0.633745	3.199577	2.574258	2.123647	2.893153	3.283165	2.327636	2.135138	4.319147	2.45236
DSP2009-Off.F4.3H.639-meh.1H.240	1.02412	0.619572	3.159965	2.52583	2.167335	2.91167	3.301714	2.347258	2.175776	4.355212	2.458845

|										
DSP2009-Off.F4.3H.976-meh.2H.292	0.970052	0.606386	3.182516	2.551588	2.058711	2.740093	3.377586	2.231585	1.993473	4.183991	2.389598
Hitosa	1.220156	0.664692	3.049144	2.397134	2.603526	3.567143	3.598319	2.793902	2.899445	4.700818	2.749428
IDON2009-Meh/1220/2009	0.953572	0.568772	3.204251	2.577495	2.003946	2.670726	3.321362	2.182396	1.89926	4.093334	2.347511
IDON-MD-2009-off/12/2009	1.119061	0.565926	3.204062	2.582283	2.236117	3.116821	3.499578	2.471923	2.361645	4.40994	2.556736
IDON-MD-2009-off/25/2009	1.161213	0.793831	3.219274	2.602042	2.273018	3.212922	3.471468	2.531366	2.430885	4.858647	2.655467
IDON-MD-2009-off/34/2009	1.290411	0.808659	3.39446	2.818796	2.198782	3.359803	3.73245	2.593045	2.488383	4.914177	2.759897
IDON-MD-2009-off/53/2009	1.320668	0.693057	3.38245	2.805121	2.258684	3.455122	3.899747	2.657208	2.58821	5.070363	2.813063
IDON-MD-2009-off/70/2009	0.984112	0.567242	3.189656	2.560689	2.068008	2.769774	3.348797	2.249475	2.013759	4.333489	2.4085
Kilinto	0.968591	0.859609	3.308694	2.70484	1.872798	2.591228	3.35486	2.110752	1.781352	3.580123	2.313285
Local	1.123399	0.78902	3.294259	2.691996	2.110763	3.02491	3.233975	2.394959	2.240933	4.259868	2.516408
Mangudo	1.174739	0.706349	3.283936	2.68101	2.197749	3.175095	3.618947	2.494403	2.366245	4.54307	2.624154
UC1113GPCLR198001/59	0.621034	0.358816	2.784409	2.05734	2.149751	2.256987	2.918902	1.994519	1.74855	3.861374	2.075168
Ude	0.746881	0.386483	2.591478	1.826759	2.607235	2.817659	2.917136	2.395407	2.430071	4.692382	2.341149
Yerer	1.110657	0.680344	3.241783	2.627856	2.169376	3.050854	3.421023	2.421871	2.253278	4.3292	2.530624
Grand Total	1.068871	0.648702	3.209293	2.587116	2.158166	2.975587	3.435717	2.379266	2.211139	4.391689	2.506554
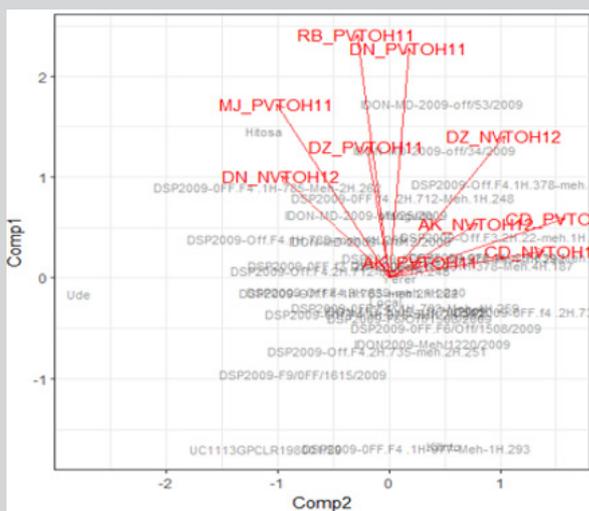
Figure 2: Biplot of genotype by environment in 10 durum wheat environments in Ethiopia.
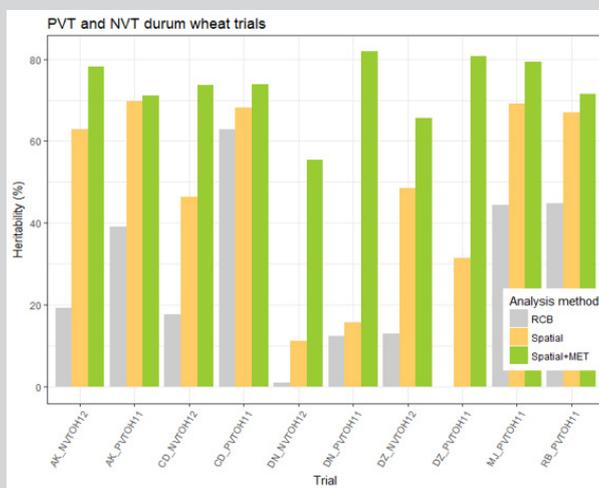

Figure 3: Improvement in heritability across three statistical methods in MET data of durum wheat in Ethiopia.

DISCUSSION

In this study, the advantage of modelling spatial correlation through neighborhood plot and variance covariance structure between trial through factor analytic model in genetic prediction is boldly reviled. Furthermore, modelling field spatial correlation plus MET through FA together under linear mixed model also indicate high improvement in heritability analysis which is also strongly recommended by recent study. For MET breeding program, restricted maximum likelihood estimation of LMM is more robust model in prediction of genetic potential across environment. Among the candidate genotype tested across ten environments, on average DSP2009-Off.F4.1H.378-meh.4H.187 genotype is the best performing material while UC1113GPCLR198001/59 genotype is performed less than the other on average.

ACKNOWLEDGMENT

KEYMESSAGE

The study demonstrates that the use of spatial field trend under linear mixed model (LMM) of combined data improve genotype selection in durum wheat data. Accordingly, on average DSP2009-Off.F4.1H.378-meh.4H.187 genotype indicate better performance than other genotypes.

REFERENCES

1. Tesemma T, Belay (1991) Morphological diversity in tetraploid wheat landrace populations from the central highlands of Ethiopia. Hereditas 114: 171-176.

2. Srivastava JP, Damania AB, Pecetti L (1988) Landraces, primitive forms and wild progenitors of macaroni wheat, Triticum durum: Their use in dry land agriculture. In: Miller TE, Koebner RMD(eds) Proceedings of 7th International Wheat Genetics Symposium, Oxford University Press: Oxford, England, pp. 153-158.

3. Abinasa M, Amsalu Ayana, Geremew Bultosa (2011) Genetic variability, heritability and trait associations in durum wheat (Triticum turgidum L. var. durum) genotypes. African Journal of Agricultural Research 6(17): 3972-3979.

4. Damesa TM, Möhring J, Forkman J, Piepho HP (2017) One step at a time: Stage-wise analysis of a series of experiments. Agron J 109: 845-857.

5. Tadese D, Lakew B, Taye G (2019) Spatial analysis in multi environment trials of malt barley in Ethiopia. African Crop Science Journal 27(3): 515-527.

6. Smith AB, Cullis BR (2001) The analysis of crop variety evaluation data in Australia. Australian and New Zealand Journal of Statistics 43:129-145.

7. 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.

8. Piepho HP (1997) Analyzing genotype-environment data by mixed models with multiplicative terms. Biometrics 53: 761-766.