Indian J. Genet., 81(1): 87-92 (2021)
DOI: 10.31742/IJGPB.81.1.9

R-AMMI-LM: Linear-fit Robust-AMMI model to analyze genotype-by environment interactions

B. C. Ajay*, K. T. Ramya[1], R. Abdul Fiyaz[2], G. Govindaraj[3], S. K. Bera[4], Narendra Kumar[4], K. Gangadhar[4], Praveen Kona[4], G. P. Singh[5] and T. Radhakrishnan[4]

Regional Research Station, ICAR-Directorate of Groundnut Research, Anantapur; [1]ICAR-Indian Institute of Oilseeds Research, Hyderabad; [2]ICAR-Indian Institute of Rice Research, Hyderabad; [3]ICAR-National Institute of Veternary Epidomology and Disease Informatics, Bangalore; [4]ICAR-Directorate of Groundnut Research, Junagadh, Gujarat; [5]ICAR-Indian Institute of Wheat and Barley Research, Karnal

(Received: June 2020; Revised: December 2020; Accepted: January 2021)

Abstract

Outliers are a common phenomenon when genotypes are evaluated over locations and years under field conditions and such outliers makes studying genotype-environment Interactions difficult. Robust-AMMI models which use a combination of robust fit and robust SVD approaches, denoted as 'R-AMMI-RLM' have been proposed to study GEI in presence of such outliers. Instead of 'R-AMMI-RLM' we propose a model which uses a combination of linear fit and robust SVD to study GEI in presence of outliers and we denote this model as 'R-AMMI-LM'. Here we prove that 'R-AMMI-LM' was superior over 'R-AMMI-RLM' as it recorded very low residual sum of squares and low RMSE values. Thus proposed, 'R-AMMI-LM' model could explain the GEI more precisely even in presence of outliers.

Keywords: Genotype x environment interactions (GEI), R-AMMI-RLM, R-AMMI-LM, Outliers

Introduction

Genotype-by-environment interactions (GEI) is an important component of any plant breeding programs for identification of a cultivar for a target location. Understanding the GEI is important at all stages of breeding programs and can be used to establish breeding objectives, identify ideal test conditions and formulate recommendations for cultivar adaptation (Sabaghnia et al. 2008). In order to understand GEI among genotypes, multi-environment trials are conducted wherein different genotypes show differential response across environments for a trait of interest enabling the identification of genotypes with broad and narrow adaptations. Additive main effects and multiplicative interaction (AMMI) analysis is one of the most popular multivariate methods to predict adaptation and stability of cultivars over multiple environments (Chuni Lal et al. 2019; Ajay et al. 2019). AMMI model is fit in two stages; first, the main effects of the model are estimated using the additive two-way analysis of variance (ANOVA) by least squares. Then, the singular value decomposition (SVD) is applied to the residuals from the ANOVA, i.e. to the interaction, to obtain the estimates for multiplicative terms of AMMI model.

AMMI model assumes equal weights for all genotypes being tested and that data is free of outliers. Outliers are common in field data such as multi-environment trials which lead to biased results, mis-interpretation of data and bad decisions. Since data contamination is more a rule than an exception in real-life data, there has been growing interest in the use of these statistical methodologies which allow for valid results even if model assumptions are violated (Copt and Heritier 2007; Lourenco et al. 2011). To overcome the problem of analysing two-way contaminated data and to understand GEI rationally, Rodrigues et al. (2016) introduced a robust AMMI model where the linear fit, underlying the ANOVA method, is replaced by a robust fit (M-regression; Huber, 1964) and the use of the standard SVD by a robust SVD approach. We designate robust AMMI model of Rodrigues et al.

---

*Corresponding author's e-mail: ajaygpb@yahoo.co.in

Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

(2016) as R-AMMI-RLM. In the present study, for GEI analysis of data with outliers we suggest using linear fit for ANOVA and Robust SVD for analysing interaction effects and we designate this model as 'R-AMMI-LM'. The new model being proposed was tested using datasets from peanut and wheat to establish the usability of the model across the crops.

Materials and methods

Dataset 1

Raw data : Raw data set consisted of 52 peanut genotypes involving 47 advanced breeding lines; two cultivated varieties of different seed size (BAU-13, Somnath); two P-efficient (ICGV-86590, SP250A) and a P- inefficient (NRCG-7320) lines. These genotypes were evaluated during 2013 and 2014 rainy seasons under two levels of P i.e. without P application (LP) and with application of 50 kg/ha P_2O_5 (as single super phosphate) (MP). Nitrogen (as urea) and potash (as murate of potash) were applied at 50 kg/ha N and 60 kg/ha K_2O equally for both the treatments. The recommended crop management practices were adopted for raising a healthy crop. Field screening, was conducted at ICAR-Directorate of Groundnut Research, Junagadh (lat 21°31'N, long 70°36'E), India, in a medium black calcareous (17% CaCO3) clayey, Vertic Ustochrept soil having moderate available phosphorus (15kg/ha P), 7.5 pH, 0.7% organic C, 268kg/ha N, 300 - 400kg/ha K, 5kg/ha available S and 1.6, 15, and 0.78kg/ha DTPA extractable Fe, Mn, and Zn, respectively. Crop was harvested at maturity and pod yield was recorded.

Simulated data

Using this raw data four more data-sets were simulated and were designated as 2SD5, 2SD10, 4SD5 and 4SD10. In-order to simulate these data sets standard deviation (SD) was calculated for each environment separately. For 2SD5 data set, 5 per cent outliers were introduced into raw data randomly at the rate of 2 times SD. For 2SD10 data set, 10 per cent outliers were introduced into raw data randomly at the rate of 2 times SD. On the other hand, for 4SD5 and 4SD10 data sets 5 and 10 per cent outliers were introduced into raw data sets as contamination at the rate of 4 times SD (Fig. 1).

Dataset 2

The dataset 2 comprised of 36 bread wheat genotypes, of which ten genotypes were from international core set for abiotic stress developed by International Maize and Wheat Improvement Centre (CIMMYT), Mexico

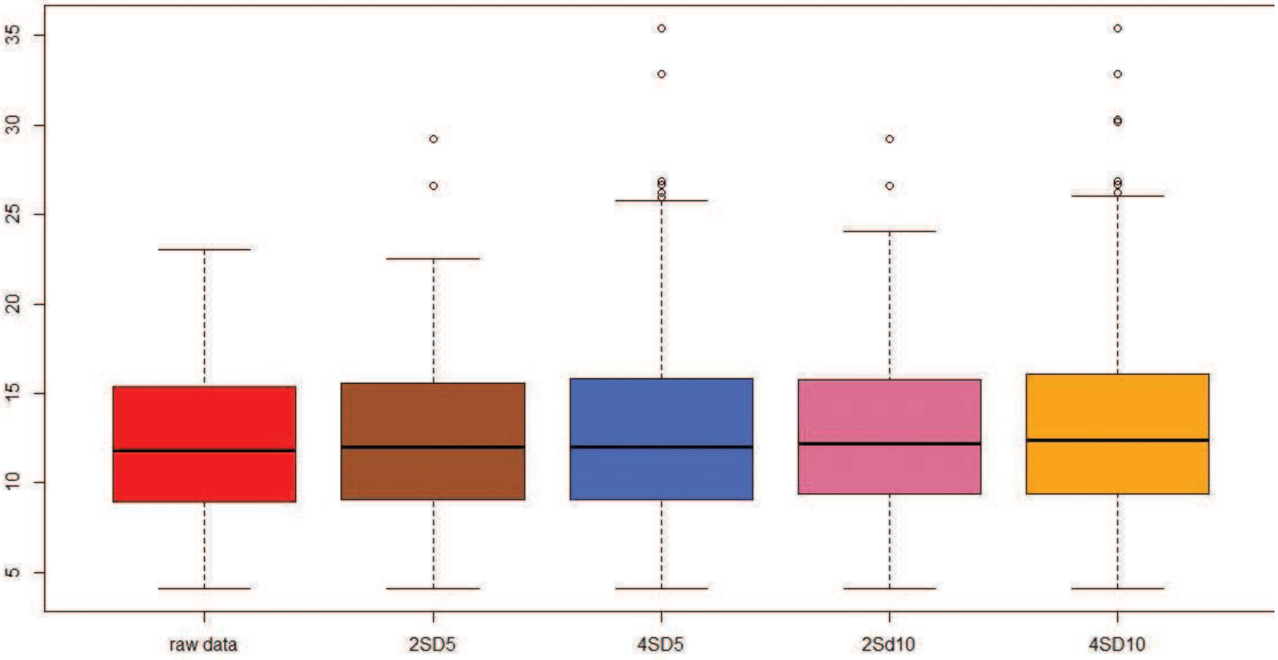

Fig. 1. **Comparison of raw data set and simulated data sets (2SD5 = 5% data outliers with 2 standard deviation; 4SD5 = 5% data outliers with 4 standard deviation; 2SD10 = 10% data outliers with 2 standard deviation; 4SD10 = 10% data outliers with 4 standard deviation)**

and 26 elite Indian genotypes released for different agro-ecological conditions. The experiments were conducted at experimental farm, ICAR-Indian Agricultural Research Institute, New Delhi, India. The latitude of the research farm is 280 38’23"N, longitude 770 09’27"E and altitude is 228.61 m above mean sea level. The experiment was laid out in a 6×6 simple lattice design with two replications and two dates of sowing with the help of Weintersteiger seed drill. The crop was sown on 15th November (Timely sowing-TS), 15th December (late sowing-LS) and 6th January (Very late sowing-VLS) to create temperature effect during growth stages during 2012 and 2013. Standard cultivation practices prescribed for wheat under irrigated conditions were precisely followed. Crop was harvested at maturity and seed yield per plot was recorded.

Robust-AMMI model analysis

AMMI model combines the features of ANOVA and SVD as follows: first the ANOVA estimates the additive main effects of the two-way data table using linear models (LM); then the SVD is applied to the residuals from the additive ANOVA model, estimating interaction principal components (IPCs). AMMI model may be written as follows,

$$Y_{i,j,k} = \mu+\alpha_i+\beta_j+ \sum_{n=1}^{N} (\lambda_n\gamma_{i,n}\delta_{n,j})+\rho_{i,j}+\varepsilon_{i,j,k}$$

where $y_{i,j,k}$ is the yield of the ith genotype in the jth environment for replicate k; μ is the grand mean; α_i is the genotype deviations from μ; β_j is the environment deviations from μ; λ_n is the singular value of the IPC analysis axis n; $\gamma_{i,n}$ and $\delta_{n,j}$ are the ith and jth genotype and environment IPC scores for axis n, respectively; $\rho_{i,j}$ is the residual containing all multiplicative terms not included in the model; $\varepsilon_{i,j,k}$ is the experimental error; and N is the number of principal components retained in the model.

Unlike AMMI model, Robust-AMMI model uses robust linear model (RLM) based on M-Huber estimator (Huber, 1981) instead of LM model to calculate ANOVA followed by the use of robust SVD procedure instead of standard SVD (Rodrigues et al. 2016) and we designate this Robust-AMMI using RLM model as ‘R-AMMI-RLM’. Here we propose a new Robust-AMMI model which calculate ANOVA using linear models (LM), instead of robust linear model (RLM), followed by robust SVD procedure. This modification reduces unaccounted variability explained by residuals and improves the accuracy of AMMI model. All the statistical analysis involving AMMI model were performed in R (R core team 2018) using packages MASS and pcaMethods. Root mean square error (RMSE) was calculated to compare the outcomes of LM and RLM based Robust-AMMI models.

Results and discussion

Robust AMMI models are used for studying GEI when there is contamination in the data and when it is difficult to assign weights to outlying observations. One of the important features of AMMI model is ANOVA showing proportion of sum of squares due to main effects using linear models at the first stage and computation of successive IPCAs (Rodrigues et al. 2016) to the residuals from the ANOVA to obtain the estimates for multiplicative terms of AMMI model by using SVD. In any plant breeding trials from field data outliers are a common problem. In order to overcome th problem of outliers Rodrigues et al. (2016) introduced a robust AMMI model where the linear fit, underlying the ANOVA method, is replaced by a robust fit (M-regression; Huber, 1964) and the use of the standard SVD by a robust SVD approach. We designate robust AMMI model of Rodrigues et al. (2016) as R-AMMI-RLM. In the present study, for GEI analysis of data with outliers we propose to estimate ANOVA by linear fit models at first stage and Robust SVD at later stages over residuals from ANOVA for analysing interaction effects and we designate this model as ‘R-AMMI-LM’. Results of both ‘R-AMMI-LM’ and ‘R-AMMI-RLM’ models are presented and discussed.

Table 1 presents ANOVA of ‘R-AMMI-LM’ and ‘R-AMMI-RLM’ models for raw data and four simulated data sets. Significant differences were observed for genotype (G), environment (E) and GEI for pod yield in raw data (0), 2SD5, 2SD10 and 4SD5 for both ‘R-AMMI-LM’ and ‘R-AMMI-RLM’ models. 4SD10 data had significant influence of E and G under both the models whereas GEI was significant only with model ‘R-AMMI-LM’.

Variability percentage explained by different components of ‘R-AMMI-LM’ and ‘R-AMMI-RLM’ models under different levels of outliers is presented in table 2. Major portion of variability was explained by E followed by GEI under both the models at all levels of outliers. Also, as the extent of outliers increases the proportion of variability explained by interaction component (GEI) and unpredictable component (residuals) increases, whereas, variability explained by main components such as G and E

Table 1. Mean sum of squares of Robust-AMMI with linear fit(R-AMMI-LM) and Robust-AMMI with robust fit (R-AMMI-RLM) at different levels of outliers among 52 peanut genotypes

Source of Variation	Df	0	2SD5	2SD10	4SD5	4SD10
R-AMMI-RLM						
Environment (E)	3	1265.8**	1185.85**	1152.93**	1166.05**	1082.20**
Genotype (G)	51	18.02**	18.62**	18.85**	18.66**	19.24**
Replication	4	2.61	2.76	4.45	2.71	4.24
E x G	153	7.60**	7.70**	7.66**	8.04*	8.28
IPCA1	53	9.33**	8.16**	12.88**	13.00**	9.09
IPCA2	51	5.92**	6.92**	6.19**	8.67	8.92
IPCA3	49	6.52**	13.22**	6.58**	8.34	23.46**
Residuals	204	1.55	2.52	3.16	6.05	9.59
R-AMMI-LM						
Environment (E)	3	1290.4**	1323.2**	1365.23**	1372.67**	1439.20**
Genotype (G)	51	19.58**	21.49**	20.88**	24.75**	26.39**
Replication	4	3.53	2.05	4.39	2.48	12.25
E x G	153	8.23**	9.44**	10.04**	12.90**	15.39**
IPCA1	53	9.86**	13.09**	14.31**	16.88**	9.01
IPCA2	51	5.52**	6.46**	7.44**	9.71*	8.92
IPCA3	49	6.93**	9.85**	8.98**	9.38*	23.46**
Residuals	204	1.55	2.52	3.16	6.04	9.52

Table 2. Percent of variability explained by different sources of variation under Robust-AMMI with linear fit (R-AMMI-LM) and Robust-AMMI with robust fit (R-AMMI-RLM) at different levels of outliers

Source of Variation	0	2SD5	2SD10	4SD5	4SD10
R-AMMI-RLM					
Environment (E)	61.2	57.3	55.3	50.5	**43.47**
Genotype (G)	14.8	15.3	15.4	13.7	13.14
Replication	0.2	0.2	0.3	0.2	0.23
E x G	18.7	19.0	18.7	17.8	16.96
Residuals	5.1	8.3	10.3	17.8	26.21
R-AMMI-LM					
Environment (E)	59.9	56.4	55.7	47.9	43.1
Genotype (G)	15.5	15.6	14.5	14.7	13.4
Rep	0.2	0.1	0.2	0.1	0.5
E x G	19.5	20.5	20.9	23.0	23.5
Residuals	4.9	7.3	8.8	14.3	19.4

decreases. Among four data sets with outliers, unpredictable component was low under ‘R-AMMI-LM’ compared to ‘R-AMMI-RLM’ model indicating that ‘R-

Table 3. Percentage of variability explained by Interaction principal components (IPCs) and Root Mean Square error (RMSE) under Robust-AMMI with linear fit (R-AMMI-LM) and Robust-AMMI with robust fit (R-AMMI-RLM) at different levels of outliers

Source of variation	0	2SD5	2SD10	4SD5	4SD10
R-AMMI-RLM					
IPCA1	42.54**	28.77**	46.17**	38.26**	20.40
IPCA2	25.97**	23.49**	21.34**	24.56	19.27
IPCA3	27.47**	43.10**	21.83**	22.70	48.68**
Total	95.98	95.36	89.34	85.52	88.35
RMSE	1.75	1.88	1.97	2.23	2.46
R-AMMI-LM					
IPCA1	43.76**	45.97**	48.06**	46.52**	20.4
IPCA2	23.55**	21.85**	24.05**	25.75*	19.27
IPCA3	28.41**	31.99**	27.89**	23.9*	48.68**
Total	95.72	99.81	100	96.17	88.35
RMSE	1.74	1.86	1.92	2.18	2.38

Table 4. Mean sum of squares of Robust-AMMI with linear fit (R-AMMI-LM) and Robust-AMMI with robust fit (R-AMMI-RLM) among wheat genotypes evaluated under six environments

Source of variation	Df	R-AMMI-RLM		R-AMMI-LM	
		MSS	% SS	MSS	% SS
Environment (E)	5	4714.10**	80.22	5030.30**	79.56
Genotype (G)	35	53.30**	6.35	61.40**	6.79
Replication	6	6.20	0.13	6.10	0.12
E x G	175	18.80**	11.20	20.90**	11.58
IPCA1	39	70.78**	73.30	69.92**	73.97
IPCA2	37	11.64**	11.43	13.55**	13.60
IPCA3	35	5.69**	5.29	3.96ns	3.76
IPCA4	33	6.45**	5.65	6.53**	5.84
Residuals	210	2.90	2.10	2.90	1.95
RMSE		3.00	2.91		

AMMI-LM' model is better in analyzing the data with outliers when compared to than 'R-AMMI-RLM'.

Per cent variability explained by first three interactions principal components (IPCAs) and root mean square error (RMSE) are presented in Table 3. For raw data (0), 2SD5 and 2SD10 IPCA1, IPCA2 and IPCA3 were highly significant in both 'R-AMMI-LM' and 'R-AMMI-RLM' models; for 4SD5 only IPCA1 was significant whereas for 4SD10 IPCA3 was significant under 'R-AMMI-RLM' models. Whereas in 'R-AMMI-LM' model for 4SD5 all three IPCAs showed significant variation but in 4SD10 only IPCA3 was significant. Unlike regular SVD based models such as AMMI, robust-SVD based models such as 'R-AMMI-LM' and 'R-AMMI-RLM' being discussed here differs from regular SVD based AMMI models in two aspects. In the conventional SVD models all the eigenvectors are orthogonal whereas in robust-SVD models eigenvectors are not orthogonal. In conventional SVD successive interaction principal components (IPCAs) are found in descending order whereas it's not the case in robust-SVD models. Larger IPCAs may follow the smaller ones (Hawkins et al. 2001) as observed in Table 3 in the present study.

In 'R-AMMI-RLM' model total percentage GEI variability explained by first three IPCAs for different levels of outliers ranged from 85.52 to 95.98 with raw data having high total percent variability (95.98%) and 4SD5 explaining the least (85.52%). Whereas in 'R-AMMI-LM' model percentage GEI variability explained ranged from 88.35 to 100% with 2SD10 having high total percent variability (100%) and 4SD10 explaining the least (88.35%). Root mean square error (RMSE) was also lower for 'R-AMMI-LM' compared to 'R-AMMI-RLM' at all levels of outliers.

In order to prove the superiority of 'R-AMMI-LM' over 'R-AMMI-RLM', wheat data set with 36 genotypes evaluated under three different dates of sowing for two consecutive years 2012 and 2013 was used. ANOVA obtained for both the models, as presented in Table 4, indicate that environment (E), genotype (G) and GE interactions (GEI) were highly significant with environment explaining maximum variation followed by GEI and G. In both the models GEI was further partitioned into four interaction principal components (IPCAs). IPCA1 and IPCA2 together explained 83% variability of GEI under 'R-AMMI-LM' whereas under 'R-AMMI-RLM' they explained 82% variability of GEI. RMSE was calculated for both the models and 'R-AMMI-LM' had lower RMSE values when compared to 'R-AMMI-RLM'. These results are in accordance with peanut data set wherein 'R-AMMI-LM' model had low RMSE values and slightly higher variability as explained by their IPCAs. This example further reinforces the superiority of model 'R-AMMI-LM' over 'R-AMMI-RLM'.

'R-AMMI-LM' model proposed in this study is more effective in studying genotype-environment-interactions even in presence of outliers with very low error variation compared to other AMMI model.

Authors' contribution

Conceptualization of research (BCA, SKB, TR); Designing of the experiments (KTR, GG, BCA, RAF); Contribution of experimental materials (BCA, KTR, NK, GPS); Execution of field/lab experiments and data collection (BCA, KTR, GPS, PK, KG); Analysis of data and interpretation (BCA, RF); Preparation of manuscript (BCA, TR, GPS).

Declaration

The authors declare no conflict of interest.

Acknowledgements

Authors would like to thank ICAR and Director, ICAR-Directorate of Groundnut Research for supporting this project.

References

Ajay B. C., Aravind J., Abdul Fiyaz R., Narendra Kumar, Chuni Lal, Gangadhar K., Praveen Kona, Dagla M. C. and Bera S. K. 2019. Rectification of modified AMMI stability value (MASV). Indian J. Genet., **79**(4):

Chuni Lal, Ajay B. C., Chikani B. M. and Gor H. K. 2019. AMMI and GGE biplot analysis to evaluate the phenotypic stability of recombinant inbred lines (RILs) of peanut under mid-season water stress conditions. Indian J. Genet., **79**: 420-426.

Copt S. and Heritier S. 2007. Robust alternatives to the F-Test in mixed linear models based on MM-estimates. Biometrics, **63**: 1045-1052.

Hawkins, Douglas M., Li Liu and S. Stanley Young (2001) Robust Singular Value Decomposition, National Institute of Statistical Sciences, Technical Report Number 122.

Huber P. J. 1964. Robust estimation of a location parameter. Ann. Math. Stat., **35**: 73-101.

Lourenco V. M. et al. (2011) Robust linear regression methods in association studies. Bioinformatics, **27**: 815-821.

R Development Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rodrigues P. C., Monteiro A. and Lourenco V. M. 2016. A robust AMMI model for the analysis of genotype-by-environment data. Bioinformatics, **32**: 58-66.

Sabaghnia N., Sabaghpour S. H. and Dehghani H. 2008. The use of an AMMI model and its parameters to analyse yield stability in multi-environment trials. J. Agr. Sci., **146**: 571-581.