

BAYESIAN APPROACH FOR ANALYZING GENOTYPE BY ENVIRONMENT INTERACTION



by

UMAR WAQAS KHAN
MSC (STATISTICS)
2013-AG-6536

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF PHILOSOPHY
IN
STATISTICS**

**DEPARTMENT OF MATHEMATICS & STATISTICS
FACULTY OF SCIENCES,**

**UNIVERSITY OF AGRICULTURE, FAISALABAD
PAKISTAN
2015**

Declaration

I hereby declare that contents of the thesis “Bayesian approach for analyzing genotype by environment interaction” are product of my own research and no part has been copied from any published source (except the references, standard mathematical or genetic models/ equations/ formulae / protocols etc.). I further declare that this work has not been submitted for award of any other diploma/ degree. The university may take action if the information provided is found incorrect at any stage (in case of any default, the scholar will be proceeded against as per HEC plagiarism policy).

Umar Waqas Khan
(2013-ag-6536)

To,

The Controller of Examinations,
University of Agriculture,
Faisalabad.

We, the supervisory committee, certify that the contents and form of thesis submitted by Umar Waqas Khan, Regd. No. 2013-ag-6536 have been found satisfactory and recommend that it be processed for evaluation by the External Examiner(s) for the award of degree.

SUPERVISORY COMMITTEE

Supervisor

Dr. Muhammad Yaseen

Member

Mr. Muhammad Arif

Member

Miss Madiha Ghamkhar

Dedication

The Sublime Love

Of

My Beloved and Kind Parents

Who taught me,

The first step to take,

The first word to speak,

The first Alphabet to write,

Inspired me to higher ideas of life,

Whose hands always rose in prayer for me,

Who are with me to feel the bud of their wishes and prayers blooming into

a flower And

Under whose feet my heaven lies

Acknowledgements

Bounteous praise for “**ALMIGHTY ALLAH**”, the magnificent, the merciful, the propitious, the supreme, the omnipotent, the omnipresent, the omniscient and sovereign whose blessing and glories flourish my thoughts and ambition and all the praises for the “**HOLY PROPHET MUHAMMAD (P.B.U.H.)**” for enlightening our conscience of faith in ALLAH, converging all His kindness and mercy upon him.

I feel much honor to express my deepest sense of gratitude and indebtedness to my honorable supervisor, **Dr. Muhammad Yaseen** (Assistant Professor), Department of Mathematics & Statistics, University of Agriculture, Faisalabad from the core of my heart for his dynamic supervision, marvelous guidance, keen interest and encouraging behavior. With humble, profound and deepest sense of devotion I wish to record my sincere appreciation to **Mr. Muhammad Arif** (Lecturer), Department of Mathematics & Statistics, University of Agriculture, Faisalabad and **Miss Madiha Ghamkhar** (Lecturer), Department of Mathematics and Statistics, University of Agriculture, Faisalabad for their sincere help, dynamic supervision and inspiring guidance throughout the course of this research work.

I want to express my great appreciation and sincerest gratitude to my friends and class fellows for their dexterous, dynamic, untiring help, friendly behavior and moral support during my whole study.

Last but not least, no acknowledgement could ever adequately express my obligation to my affectionate Parents and my brother **Sufian Khan** whose endless effort and best wishes sustained me at all stages of my life and encouraged me for achieving high ideas of life. May ALLAH bless all these people with long, happy and peaceful lives (Aameen)!

Umar Waqas Khan

Table of Contents

<i>Abstract</i>	ix
<i>Chapter 1</i> <i>Introduction</i>	1
1.1 Objectives	3
1.2 Significance of study	4
1.3 Structure of study	4
<i>Chapter 2</i> <i>Review of Literature</i>	5
2.1 Classical approach	5
2.2 Bayesian approach	14
<i>Chapter 3</i> <i>Materials and Methods</i>	18
3.1 Material	18
3.2 Classical Approach	20
3.2.1 AMMI model	20
3.2.1.1 Determination of optimal number of interaction components	23
3.2.1.2 AMMI Biplots	24
3.2.1.3 AMMI stability value (ASV)	25
3.2.1.4 Genotype Selection Index (GSI)	26
3.2.2 GGE-biplot	26
3.2.2.1 Biplot inner product property	27
3.2.2.2 Data centering prior to singular value decomposition	28
3.2.2.3 Data scaling and singular values decomposition and Singular values partitioning	29
3.2.2.4 Column-Metric preserving and associated interpretation	30

3.2.2.5	<i>Row-Metric preserving and associated interpretation</i>	30
3.2.2.6	<i>Symmetrical partitioning and associated interpretation</i>	31
3.3	<i>Bayesian Approach</i>	31
3.3.1	<i>Von Mises-Fisher distribution</i>	33
3.3.2	<i>Likelihood Function</i>	34
3.3.3	<i>Prior Distribution</i>	34
3.3.4	<i>Posterior Distribution And GIBBS Sampler</i>	38
3.3.4.1	<i>Gibbs Sampler</i>	40
Chapter 4	<i>Results and Discussion</i>	41
4.1	<i>Classical approaches analysis</i>	41
4.1.1	<i>Combined Analysis of Variance</i>	41
4.1.2	<i>AMMI analysis and bi-plot representation</i>	42
4.1.2.1	<i>AMMI bi-plot</i>	43
4.1.2.2	<i>AMMI Stability Value (ASV)</i>	46
4.1.2.3	<i>Genotype selection index (GSI)</i>	47
4.1.3	<i>GGE-biplot</i>	47
4.1.4	<i>Cluster Analysis</i>	49
4.2	<i>Bayesian analysis</i>	50
4.2.1	<i>Uncertainty and confidence regions of first two bilinear terms in Bayesian AMMI</i>	51
4.2.2	<i>Credible intervals for genotypes</i>	54
4.2.3	<i>Credible intervals for environments</i>	54
4.2.4	<i>Credible Regions of the First Two Bilinear Terms of the Linear Bilinear Model</i>	54
Chapter 5	<i>Summary</i>	62
References		64

List of Tables

<i>3.1 List of locations</i>	18
<i>3.2 List of genotypes</i>	19
<i>3.3 The General Analysis of Variance for the AMMI Model</i>	23
<i>3.4 The General Analysis of Variance for the AMMI Model</i>	24
<i>4.1 Combined Aanlysis of Variance</i>	41
<i>4.2 Combined analysis of variance (ANOVA) according to the AMMI</i>	42
<i>4.3 AMMI stability values</i>	46
<i>4.4 HPD table for genotypes</i>	55
<i>4.5 HPD table for environments</i>	56
<i>4.6 HPD table for significant bilinear components</i>	56

List of Figures

4.1	Biplot of IPCA axis-1 versus Axis-2	44
4.2	Biplot of IPCA axis-1 versus mean yield	45
4.3	GGE-biplot based on genotype-focused scaling for genotypes	47
4.4	GGE-biplot based on environment-focused scaling for environments	48
4.5	Polygon views of GGE-biplot based on symmetrical scaling	48
4.6	Average environment coordination (AEC) views of the GGE-biplot	49
4.7	Comparison of the genotypes with ideal genotype	49
4.8	comparison the environments with ideal environment	49
4.9	Cluster analysis of genotypes	50
4.10	Cluster analysis of environments	50
4.11	Trace for MCMC samples	52
4.12	Posterior densities and regions for singular values and cumulative proportion	53
4.13	Bayesian biplot for environments	57
4.14	Bayesian biplot for genotypes	58
4.15	Bayesian dendrogram for environments	59
4.16	Bayesian dendrogram for genotypes	60
4.17	Comparison between OLS and Bayes method	61

Abstract

Genotypes of different genetic structure behave differently in different environmental conditions. Genotype-by-environment interaction (GEI) is referred as differential responses of different genotypes across different environments; GEI has great importance because of higher performance of genotypes to be assessed by GEI. But presence of GEI makes analysis more complicated. To up-root these assessment complications several methods have been proposed such as Principal Component Analysis (PCA), Cluster Analysis, Additive Main effects and Multiplicative Interaction (AMMI) models and Genotype plus Genotype by Environment interaction (GGE). These methods neither overcome the problem of over parameterization nor use the prior information. The aim of this study is to use such technique which can address these problems. For this purpose a wheat crop data comprised of 30 genotypes test across 13 different locations of punjab, Pakistan for two consecutive years was used. The layout of the experiment was Randomized complete Block Design (RCBD). In this study a comparison was made between Classical methods AMMI, GGE biplot and Bayesian approach using Von-Mises Fisher distribution as prior. classical methods showed that genotype V-11098 was the most desirable genotype based on stability and high yield performance. Bayesian approach was used for GEI because it makes statistical interpretation rather easy by relaxing some constraints and it uses the prior information, also provides solution for these by using MCMC algorithm. Bayesian strategy for analysis of GEI was used to assess the general, specific performance of genotypes and risk related to genotype. Analysis revealed that bilinear terms $u_{25,1}$ for genotype NS-10 genotype and $v_{13,1}$ for environment S13 (Piplan-14) were found significant indicated that these have effect on interaction. It was observed that Bayesian approach can nicely explore GE interaction.

Chapter 1: Introduction

Multi-environment experiments are commonly conducted to identify genotypes that have high yield and are less sensitive to adverse changes in environments. Genotypes do not perform similar in different environments so the presence of genotype by environment interaction (GEI) is of a great importance; it is necessary to analyze genotype environment interaction (GEI) to assess that how well genotypes perform in different environments.

Superior genotypes or environments can be identified by exploring GEI ([Allard *et al.*, 1964](#)). GEI revealed a basic principal is that, same genotypes do not exhibit similar behavior or pattern when environment condition change. Therefore, stability of genotypes is normally assessed by testing these genotypes in different conditions. these conditions may be locations, years or seasons etc. Combination of physical, chemical and biological factors are refereed as environment which may influence the growth of object. GE appear when genotypes are tested in multi-environmental conditions ([Becker *et al.*, 1988](#)). [Magari *et al.* \(1993\)](#) found that environmental factors contribute in the stability of yield an significantly effect the heterogeneity.

Yield and grain quality are two very vital features to study for the evaluation of wheat crop. Characteristics such as quality of wheat and yield vary differently in different environments ([Uhlen1998](#)), which takes into account environmental consequence on these characteristics. Biological processes and number of genes are highly interaction with traits like yield and quality . The mixture of genetic and ecological factors such as soil characteristics, precipitation, fertilization, soil and air temperature, as well as the genotype \times environment interaction can be used to define these traits ([Peterson *et al.*, 1992](#), [Johansson *et al.*, 2003](#)).

In Pakistan Environmental variation are very large so $G \times E$ is expected will be high. So different genotype may have variation in yield in different environments. For selection of superior genotypes across location, it is important to determine magnitude

instead of calculating average yield (Gauch *et al.*, 1997). According to Simmonds (1962) adaptation is a mechanism in which a genotype shows the better stability in different environments. So always researcher aim is to produce such genotypes which have good adaptation ability.

In multi environment trials (METs) the GE effects are more evident, that have three main aims : a) Estimation and forecasting of experimental data yield level can be done more precisely ; b) yield stability and genotypes adaptations to different environments can be determined; and c) selection of ideal genotype can be made which can provide guidance in future for suitable sites (J. Crossa *et al.*, 1990).

There are many methods to analyze GEI such as principal component analysis (PCA), cluster analysis and genotype and genotype by environment interaction (GGE) bi-plot analysis (Yan and Tinker, 2005). But these methods have some drawbacks such as PCA fails in separation and identification of significant genotype and environment main effect, Cluster analysis just provide the graphical grouping of genotypes or environments ; so AMMI can be used to overcome these problems because it accomplishes the GE much efficaciously by having maximum variation explained by the interaction sum of squares (Zobel *et al.*, 1988). Stability analysis can effectively be analyzed by AMMI model because it explores the GEI; that is why AMMI models are frequently used for GEI (J. Crossa *et al.*, 1990). Further more AMMI model can be visualized by using bi-plots. The AMMI model which can be defined as follow:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^t \lambda_k u_{ik} v_{jk} + \epsilon_{ij} \quad (1.1)$$

Where Y_{ij} is response, λ_k = The singular value for k-th principal component axis and u_{ik} = Element of left singular vector and v_{jk} = Element of right singular vector

Although AMMI models are very useful for GEI but in AMMI model have some problems such as;

- It is difficult to explain the $G \times E$ structure by AMMI model when data set is incomplete or have missing values (PEREZ-ELIZALDE *et al.*, 2011).
- The problem of over parameterization arises which make model much more com-

plex, and also the model does not use prior information (JOSSE *et al.*, 2014).

- Outliers or extremes can lead to misleading interpretation using Classical AMMI as it use ordinary least square method for estimation (OLS) , which can be significantly effected by those outliers (Rodrigues *et al.*, 2015)

It was suggested that these problems can be tackled by Bayesian methodology . Bayesian approach is very useful because it make statistical interpretation easy by using standard MCMC algorithm (Gelman, Carlin, *et al.*, 2004). Viele *et al.* (2000) applied uniform prior on the first column of u_i as well as on the v_j and obtain useful results. Gibbs sampling with embedded Metropolis-Hasting random walks is another tool of Bayesian approach to obtain estimates of interaction. Bayesian paradigm for principal component analysis was presented by Hoff (2007) and Simdl *et al.* (2007). Although it provides similar computational environment but it differ in linear part which contain grand mean and column effect. Bayesian approach is widely used to analyze problems like GEI because it provides relaxation to model constraints and uses prior knowledge of the data.

Silva *et al.* (2015) suggested a Bayesian shrinkage AMMI instead of classical Bayesian, although it gave high shrinkage value but it incorporate with credible intervals for biplots as compared to Bayesian AMMI. PEREZ-ELIZALDE *et al.* (2011) proposed a Bayesian approach for the treatment of problems with AMMI model using Von-Mises Fisher distribution as a prior, aiming two advantages i): Bayesian approach can be applied to GE data sets because it can facilitate in analysis by using prior information related to the experiment under study; ii) the distribution related to any any interest can be obtained by posterior distribution. Further more, credible interval obtained can be used for good interpretation and almost provide biplots similar to classical methods

1.1 Objectives

The objective of the study are as follow:

- To examine the Genotype environment interaction using Bayesian approach.
- Graphical represent of bi-plots using prior information.
- To elaborate that Bayesian models can be easily adapted to GEI.

1.2 Significance of study

Aim of study was to illustrate that Bayesian methodology can be implemented to yield. It was illustrated through credible intervals as well as by biplots that this approach can easily be adapted to explore the $G \times E$, which is often the area of concern for the researcher.

1.3 Structure of study

The thesis is generally structured under five chapters as outlined below:

- Chapter 1; Introduction part includes overall description of research by defining research objectives, and the structure of research and dissertation.
- Chapter 2; This chapter has two parts; first part of this chapter describe about the classical methods used in literature. Second part of this chapter focus on Bayesian methodology .
- Chapter 3; Methodology and material used in this study was disused in this chapter.
- Chapter 4; This chapter presented the main findings of of classical methods and Bayesian approach for wheat crop data . The statistical software [R, 2015](#) was used for analysis.
- The findings of the project were discussed in fifth chapter i.e Summary

Chapter 2: Review of Literature

The information on the topic is not inadequate but the inference of different research are not consistent and exhibit differences reported to the material used and location of experiment . However, some studies from literature are reviewed here which are relevant to the subject.

2.1 Classical approach

A research was conducted to study different stability measures from AMMI model for the analysis of stability. For This purpose a data of two year (2004-2005), which consist of 17 genotypes of chickpea across five environments was take. The experiment was Randomized complete block design with 4 replications. The relationship stability measure was assessed using Spearman's rank correlation. The analysis also shown that stability measures can be categorized into three groups. The first group encompassed, $SIPC_4, EV_4, MASV, D_{Z4}, AV_{(AMGE)}, W_{i(AMMI)}$ and FA_i on the basis of correlation with mean yield. ASV, Da_4, B_i and Za_4 were put in 2nd group for similar criteria. Group 3 contain of FP_i that was not related with yield. Then Principal component analysis was performed to understand relationship among different methods. Bi-plot of PC 1 and PC 2 was made, analysis showed that 84% variation of GE interaction can be explained by first four multiplicative terms. On the basis of high mean yield and according to all measures, it was recommended that genotype G13 (FLIP 97-114) can be adapted widely across environments for good results ([Zali et al., 2012](#)).

For analysis of multi-environment data ([Gauch, 2006](#)) made comparison of merits between two theories which purely base on singular value decomposition. The models were AMMI and genotype main effect and genotype by environment interaction (GGE), principal component analysis was also performed. It was observed that as far

as model choice concern, AMMI was best among them for agriculture analysis because over all variation can be partitioned into three sources by it. where as for predictive accuracy, all methods proved to be equivalently efficient. But it was suggested that to draw useful inference from the data all required model diagnosis.

Genotypes of different genetic structure behave differently, when place in different growing conditions. The different growing responses with respect to environment are called GE (Genotype by Environment). GE interaction is a structure in which the main effects and non-additive interaction of plants breeding or genotypes are studied. There are several techniques; and ANOVA (Analysis Of Variance) is one of such technique to analyze, but for GE the interaction from additive model the effects are non-additive. So AMMI model was proposed for such two way data, AMMI (bi-plots) were used to identify the relationship of interaction. It provided clustering of genotypes on the basis of similar of outputs and identifying trends of genotypes across environment. It was also a useful technique to recognize when GE interaction is provided by insignificant genotype and environment main effects, or where the structure of the interaction is influenced by outliers. Tai's stability statistics may also be used, along with bi-plots analysis, to measure the relative stability, reliability, and ordering of genotypes in a specific regional trial (Shafii *et al.*, 1998).

An experiment was conducted to makes comparison among three different methods SREG, regression analysis and AMMI to explore the $G \times E$. The objective of the study was to the assessment of genotypes which show maximum yield potential and the stability in diverse environments. For this purpose an experiment in which 13 genotypes of wheat were sown in six different locations was conducted in Chile. Analysis revealed that "Pandora-INIA" gave the maximum yield in all environment and was referred as the most stable genotype, it also maintain the quality level. It was also suggest although all method were sufficiently explain the $G \times E$, but among theses these techniques SREG was most appropriate (Castillo *et al.*, 2012).

To examine the yield stability, adaptability of environments and for analyzing

G×E of tobacco, An RCBD experiment was evaluated which comprises 15 hybrids in 8 different environments with three replication for 2006 and 2007. AMMI analysis showed that 88% of sum of squares was explained by the environments. Environment with large sum of squares were indication that they were diverse. From AMMI results it was also noticed that IPCA 1 explained 82% interaction sum of square. Similarly first two IPCA cumulatively explained 91% sum of square for interaction. Results disclosed that hybrids PVH03, K394/NC89 and Coker254/NC89 having smallest interaction, and hybrids ULT109, NC291, Coker254/Coker347 and VE1/Coker347 having highest interaction were found the most consistent and inconsistent hybrids, respectively. Furthermore, Rash for Coker254/K394, NC291 and CC27 non-in drought stress condition were more appropriate and for Rasht, hybrids NC89/Coker347, K394/ Coker347, Coker254/VE1 and ULT109 were more preferable in drought stress condition ([Sadeghi et al., 2011](#)).

The objective of study was to assess whether the presence of GE affects the lint yield and fiber strength or not. For this purpose An MET data which consist of 8 genotypes of cotton in twelve different environments from South Carolina was taken and AMMI analysis was performed. Analysis revealed that GE affects the lint yield and fiber strength. 71% variation of GE for lint yield and 70% variation for fiber strength were explained by first two principal components. Further diagnosis showed that two mega environments in South Carolina cotton areas can be made for lint yield that is; north east areas and southern areas. On the other hand single mega-environment can be representative for whole South Carolina region to test fiber strength and genotypes which can be widely adapted [Campbell et al., 2005](#).

A comparison was made between AMMI and joint regression to explain which method is sufficient to describe genotype-location (GL) effects along with principal component axes's variance which found significant. To asses Repeatability of stability in genotypes in different environments was also main purpose. For analysis different methods which include Euclidean distance about origin for significant PC axes (D), absolute PC values (IPC 1l) and Finlay and Wilkinson method for joint regression were

used. Shukla's stability variance was also implied. This study based on three data sets including two for maize, and one for bread wheat and one for oat. Results revealed that AMMI analysis was more adequate and valuable for explaining the GE interaction in six data sets. D method was proved to be more repeatable than | PC1 | and Shukla's method. As for as ordination of GL effects concerned both methods were found appropriate. It was shown that wheat and maize data set were consistent in ordination for GL interaction PC1, when neither season nor genotypes in common [Annicchiarico, 1997](#).

To assess genotypes which perform well in diverse environment and gave high yield, AMMI model was used by ([Akter *et al.*, 2014](#)) in Bangladesh for rice crop. The experiment layout was RCBD, which consist of 20 genotypes across 5 different places . It was explored the main effects and interaction terms were found significant. AMMI bi-plot between first two PC shown that BRRI 10A/BRRI 10R (G3) was the most stable genotype as it gave maximum yield and BRRI dhan39 (G 12) gave the lowest yield, so it was referred as unstable genotype . It was observed that the variety G3 exhibited high yield in all environments and perform better than other genotypes. It was noticed that $G \times E$ did not affect genotypes BRRI 1A/ BRRI 827R (G1), IR58025A/ BRRI 10R (G2),BRRI 10A/BRRI 10R (G3) and BRRI hybrid dhan1((G4), so these genotypes would be stable in different environments. As far as environments were concerned, Gazipur (E1) and Jessore (E5) gave near zero IPCA scores, this mean these location can be considered stable for the improvement of yield rice crop.

The aim of study was to observe stability and genotype environment interaction. For this purpose a data set of durum wheat was taken from south east Ethiopia for years (2003-2005), which comprises twenty genotypes across 15 environments. The relationship among genotypes and the stable genotypes were identified using different stability parameters. Combined ANOVA and AMMI analysis was performed. In ANOVA except genotype-year interaction other effects were observed significant, whereas in AMMI analysis 4 multiplicative term were significant. Genotype 3 and 4 was recommended across environments as these were identified most stable genotypes. Among different stability measures Sd_i^2 , W_i , Sx_i^2 and ASV showed high rank correlation. So it

was recommended that all are equally beneficial to assess the stable genotypes (Letta, 2007).

A data set which consists of 10 genotypes across seven different environments in Pakistan for two consecutive years 2007-2008 was analyzed by (Mujahid 2011). Analysis of variance was performed and it was observed that approximately 79% Sum of squares (SS) was explained by environments, while genotypes explained 3% SS and 9% was by GE interaction. To assess grouping among environments and among genotypes Cluster analysis was applied. It was shown that environments merge in 4 groups and genotypes in 4 groups. GGE biplot between PC 1 and PC 2 was drawn, Genotype NR-314 and genotype NR-310 behave differently in environments than other genotypes. GE biplot revealed that NR-306 and NR-305 were the most stable genotypes because they gave maximum yield.

To assess the plant production it is important to know how genotypes and environment interact with each other. A trial consists of 21 genotypes on 7 different locations of rye-grass was used to highlight the interaction. Many models were presented by statisticians but a bi-additive model fit well. Bi-plots were used to demonstrate the performances of genotypes in different locations, it was observed that genotype 1 (G1) and location 4 (L4) showed a maximum negative interaction, while genotype (G1) and genotype (G7) were far from origin so they showed a maximum positive interaction; similarly genotype (G2), genotype (G4) and genotype (G9) were poor at location 7 (L7). It was suggested that additive model are better but some areas need some attentions like, Additive model may be considered as extension of generalized linear model, how well is interpretation by confidence region and how precise and valid asymptotic formula is (Denis *et al.*, 1996).

To evaluate GE interaction and determine genotypes stability an experiment was conducted in Ethiopia (Africa) during 2004-2005. Twenty genotypes of wheat were tested in 6 different research stations. The interaction for genotypes and environment was analyzed by using AMMI and GGE biplots. Combined ANOVA showed that

main effects as well as interaction effect were highly significant. GGE biplot analysis revealed that for first year 70% SS and for 2nd year 80% sum of square were explained by PC1 and PC2. Mega-environments were identified for first year by "which-one-where" pattern. It was suggested that a repeatability of which-won-where arrangement over years is the essential and necessary condition for mega-environment delineations and to make recommendations for genotypes (Negash *et al.*, 2013).

Environment effects on genotypes of wheat were assessed by making comparison between AMMI and GGE bi-plot. For analysis of GEI in bread wheat; an experiment was conducted in 2010/2011, and 2011/2012 seasons at the Research Farm of Faculty of Agriculture, Sohag University, Egypt. Ten wheat cultivars differing in to adapt heat in 12 environments were used. The experiment was completely randomized block design which consist of three replicates. It was observed that AMMI and GGE bi-plot models were successful in assessing the performance of genotypes and the choice of best genotypes was almost similar in both of them. On the basis of two analysis AMMI and GGE bi-plot models, G10 (Giza 168), G2 (Sakha 8) and G6 (Sids 1) were identified by high yield and were stable; therefore the G10 (Giza 168) can be considered as an ideal genotype. When this technique does not explain the interaction in grain yield, then another technique called mixed model may produce better results (Mohamed *et al.*, 2013).

To assess the performance of genotypes across diverse environment; an experiment was conducted at different locations in Pakistan to check the stability parameter for yield of grains. Twenty genotypes of spring wheat were evaluated in thirty one locations of Pakistan for the year 2001-02. ANOVA revealed that 98.6% variation for GE interaction was accounted by genotypes. For stability parameter checking they used two methods; Safety-First Rules and cluster analysis. On the basis of Safety First Rules, genotypes such as V-97046, 97B2210 V-98059 and V-97052 were noted to be as stable genotypes. Moreover, the genotypes performing similar response pattern over the environments and vice versa were grouped by using cluster analysis on wheat GE interaction data. twenty genotypes were clustered into 10 groups, whereas 31 environments

were merge into 16 clusters. It was concluded that the Safety-first rule is the best technique due to the reason that this technique explicitly weighs the importance of stability relative to yield, whereas cluster analysis technique is useful to asses that which genotype performed well at specific environment ([Rasul *et al.*, 2007](#)).

Multi-environmental trials (MET) were graphically analyzed by using GGE bi-plot and genotype \times traits bi-plot commonly. A new bi-plot technique covariate-effect bi-plot was used on MET data set of barley. For this purpose an experiment was piloted in North America, which was consisted of one hundred forty five genotypes across 25 environments. A comparison among all methods was done, GGE and covariate effect bi-plot explored that environment can be divided into two meg-environments. Furthermore, 81% pattern of GGE was explained by covariate effect bi-plot. This suggested that indirect selection for trait on basis of yield can effectively reconnoitered by the GGE pattern. In particular for east environments, selection of large kernel weight, good loading resistance, initial heading can be used to improve the yield of barley. While for western environments, yield can be improve only by the selection of yield per se through environments. It was suggested that by using all methods jointly, MET can be analyzed in much better way ([Yan and Tinker, 2005](#)).

To assess the stability and adaptability an experiment was conducted by ([KAYA *et al.*, 2002](#)) in Turkey during year (2000-2001). The experiment was consist of 20 genotypes of wheat crop in six different environments with four replication. The layout of experiment was Randomized Complete Block Design. First, Combined ANOVA was performed for six environments and all effect were found significant. Secondly, AMMI Analysis was performed and it was observed that environmental impact has significant effect on the production of Wheat crop. 100% interaction was explain by first five principal components, whereas PC 1 and PC 2 almost explain 78% interaction.

To examine Stability performance an experiment was evaluated for thirty genotypes of wheat in six different stations. The pattern of experiment was RCBD with three replications. AMMI and GGE analysis was performed for the evaluation of genotypes

and analysis explored that $G \times E$ is highly significant. AMMI stability value (ASV) showed that 14 (Irena \times Veery) have high mean yield so referred as most stable. The GGE revealed that crosses number 11 (Irena \times Chamran) and 17 (S-78-11 \times Chamran) were the most stable combination, and it was recommended these can be used for the making hybrids (Rad *et al.*, 2013).

To study the GE interaction and stability measure, an experiment in Ethiopia was conducted during 2007 and 2008. At sixteen different environments 14 genotypes of pea were appraised, the trial was carried out in RCBD layout with 4 replicates. AMMI analysis and site regression (SREG) bi-plot method were applied for assessment, pooled ANOVA revealed that main effects and interaction effects were significant, two component explain 69% sum of squares for interaction with 52 degree of freedom. The initial five bilinear terms were observed important in AMMI. Except EH02-036-2 and COLL.026/01-4 genotypes no genotype showed better performance than others, as these exhibit top ranking among five out of 16 environment. It was validated that both method can effectively be used for visual comparison and to identify the superior genotypes. It was suggested that indirect selection of environment can be proved effective for the identification of better genotype performance (Tolessa *et al.*, 2013).

To assess that $G \times E$ plays an important role in pasta color which an important trait, an experiment was conducted which included 18 genotypes of wheat crop sown in 13 different sites. Main effects and interaction effects were observed highly significant when combined ANOVA was performed. Ranking changes in genotypes did not show any sign of significance. Among all genotypes, G11 adapted the conditions well enough because it gave maximum grain yield, it was also indicated that the pasta color potential can be improved of semolina. Furthermore, for grain yield as compared to non-crossover type, cross-over type was found more important. Similar pattern was observed for pasta color of semolina. A particular local adaptation arrangement was observed as high GY, TW and semolina yellowness, was identified, also less correlation among these will facilitation in improve of pasta color without effecting the production and quality. (Schulthessa *et al.*, 2013).

Although AMMI model can analyzed $G \times E$ adequately which is based on singular value decomposition (SVD), but problem arise when there are extreme values or outliers which can make data contaminated. As AMMI use least square method can be significantly affected by these contaminations because OLS is sensitive to outliers. (Rodrigues *et al.*, 2015) proposed a robust AMMI (R-AMMI) model to tackle these fragility of classical AMMI model. A simulated as well as two real data sets was used for analysis. It was observed that in classical AMMI OR91 exhibit a significant effect on biplot and shown overlapping in different direction, whereas, R-AMMI biplot despite all influence of OR91 displayed a better visual and made interpretation rather easy. Results explored that R-AMMI can be used to obtain successive principal components, Moreover, similar result and interpretation can be applies on R-AMMI biplots. It was also suggested that precautionary measure should be taken while cleaning detectable measurements.

Treatments and Blocks are two factors for randomized complete block Design (RCBD); if treatments are fixed, best linear unbiased estimation (BLUE) method is better, if treatments are random, best linear unbiased prediction (BLUP) method is preferable because it reduces the treatment means provides less root-mean-square error (RMSE). Practically the variance components are estimated. Prediction obtained through estimated variance component, is called empirical best linear unbiased prediction (EBLUP), but EBULP cannot be reliable when the experiment is small. A simulation was used to assess performance of EBLUP with normally and non-normal random effects and was compared with Bayesian approach. It was observed that EBLUP performance was better as compared to BLUE for RMSE, as well as for non-normally distributed treatment. The Bayesian method provided the smallest RMSE and more precise prediction intervals than other methods (Forkman *et al.*, 2013).

2.2 *Bayesian approach*

Bayesian inference is more useful because it provides easy interpretation of statistical conclusions. The output of analysis is based on posterior distribution, so for unknown parameters it gives the ability to estimate intervals. So this property of Bayesian inference provides flexibility to fit any model for multi parameters (Gelman, Carlin, *et al.*, 2004).

Genotype rank changes across environments, were compared which are termed as crossover interaction (COI). A comparison was made between two bilinear models, the sites regression model (SREG) and shifted multiplicative model (SHMM). Two cultivar, one comprised of 20 genotypes of maize evaluated in fourteen international locations, layout of trail was RCBD with 4 replications. Other data consist of sixty genotypes of wheat in 5 distinguished sites with 2 replication in each site, trail was designed on RCBD layout. For maize dataset, cluster analysis on non-COI for environments was done, whereas wheat data was used for clustering genotypes. It was observed that these methods were valid for clustering different location and genotypes on non-crossover GEI subsets (Jos. Crossa, Yang, *et al.*, 2004).

Estimates of multiplicative interaction can be obtained by Bayesian approach which uses Gibbs sampling with embedded Metropolis-Hasting random walks (Viele *et al.*, 2000).

Principal component analysis (PCA) is dimensions reduction technique of models by rotation of axes. Different noise can be accounted by using an extension of PCA called Bayesian Principal Component Analysis (BPCA). But prior information cannot be utilized by PCA or its extensions. It was showed that BPCA not only estimate the parameters precisely, but also take measurements in much better way. The BPCA algorithm assume that the rank of model is known or it can be estimated, and that the noise follows the Gaussian distribution, but BPCA method is useful even if noise is not

Gaussian. Furthermore, BPCA showed more robustness for errors to estimate the rank of model. The proposed BPCA is useful for to tackle PCA or MLPCA problem such as to estimate prior distribution by using Monte Carlo methods (Nounou *et al.*, 2002).

The problem with the least square estimates of AMMI model was, it did not collaborate the GE interaction of first two components into bi-plots. To over-come this difficulty an alternative Bayesian approach was suggested. For this purpose, a study was carried out using a data set of grain yield, it includes nine genotypes assessed in 20 locations using a RCBD with 4 replications, vague but proper prior was applied. Bayesian bi-plot of bilinear terms was plotted and observed that genotype 1 and 8 perform significantly different from other genotypes and form a group with negative values. Genotype 4, 5, 6, and 9 were on right side and merged into another definite group of genotypes having positive values. Highest posterior density (HPD) interval of 95% and 99% probabilities were formed, and those which did not include null point (0, 0) were referred as highly significant. It was suggested that this new method is better way to assess the bilinear terms which remained under-shadow in other methods (Jos. Crossa, Perez-Elizalde, *et al.*, 2011).

In plant trial AMMI models are widely used to explore GEI, but criteria for the selection of number of multiplicative terms should retain in the model, which can sufficiently explain GEI is debatable issue with AMMI. For this purpose a study was conducted which comprises evaluation of 55 genotypes across nine environments. Randomized complete block design layout was used with three replicates. A comparison between Bayesian AMMI and Bayesian shrinkage AMMI was made; both methods show advantages over one another. Bayesian AMMI gave low shrinkage singular values as compared to Bayesian shrinkage AMMI, but gave useful path for the determination of credible interval. On the other hand, Bayesian Shrinkage AMMI gave principal components which have high shrinkage value as compared to mixed models. It allowed the selection of model vary much same as classical AMMI model. It was suggested that Bayesian shrinkage AMMI can be used for estimation of credible intervals without taking Gaussian assumptions into account which are required in classical methods (Silva

et al., 2015).

Bayesian approach was applied for linear bilinear models inference by using a prior distribution called Von Mises-Fisher distribution. Instead of using orthogonal eigenvectors, orthogonal matrices were used on MCMC samples. Bayes factor was used to observe that which bilinear terms were significant. For analysis; first a simulated data was used. It was consist of 5 rows and 3 columns with 4 replication, for this posterior distribution was estimated and bi-plots were constructed. From the summary of marginal posterior distribution, HPD interval (90% and 95%) indicated that bilinear terms which do not include point (0,0) were significant. It was observe that eigenvectors for $u_{1,1}$, $v_{1,1}$, $v_{2,2}$ and $v_{3,2}$ were the significant. The significance was also shown by using bi-plots that genotype 1 and environments S_1 , S_2 and S_3 were significant.

Further, about the clusters among rows and columns were inspected using dendrogram. Then, a real data set of 12 maize hybrids and of 25 environment for consecutive years was used; Where, first year data were used as prior. Histogram of MCMC samples of marginal posterior and for cumulative proportion of variance explained were shown and indicates that 5 component explain approximately 90% interaction variance. Bayes factor 104.3 indicated that 3 bilinear components are appropriate. Bi-plots indicated that genotype 12 was significant, and environments such as S_5 , S_9 , S_{23} and S_{25} were significant for "GE" interaction variability, joint description was also given; dendrogram of hierarchical clustering algorithm was also shown (PEREZ-ELIZALDE *et al.*, 2011).

In AMMI models the problem of over parameterization arises (there is over parameterization when alteration of any non-constant function of the parameters do not alter the likelihood). It was suggested that this problem can be tackled by defining directly priors on all the parameters without taking into account the constraints by applying an appropriate post processing at the posterior level. This problem can be tractable by standard MCMC algorithm. So genotype environment dataset of 16 Genotypes was used and there were some question needed to be answered, like general performance of genotype, Specific performance of genotype and risk involved with genotype.

Firstly, from the graphical representation it was shown that Genotypes 1-6, have positive main effects, Genotypes 8-14 have negative main effect and Genotypes 15-16 were stable. Secondly, behavior of Genotypes in different environment, It was observed that 1-6 genotypes performed better in poorest environments, 8-14 perform good behavior in the best environment, whereas 15-16 observation were stable in all conditions. Finally it was shown that the risk related to the Genotype that yield will be lower than a specific level, It was shown that Genotype 3 and 16 have probabilities of .16 and .19 have yield less than 4. It was suggested that Bayesian approach can answer these question in much better way. Further, to use effects either fixed or Random can be differentiated by using different prior ([JOSSE *et al.*, 2014](#)).

Chapter 3: Materials and Methods

This section will explain that how this study was designed and methodology strategy was used. Why the methods were selected and how the research had conducted, will also be explained.

3.1 Material

This study used data of wheat for year (2012-13) and (2013-14) were taken from Wheat Research Institute (Ayub Agriculture Research Center Faisalabad, Punjab). Thirty wheat genotypes, listed in (Table 3.2) were evaluated at 13 locations in year 2013 and 2014 listed in (Table 3.1).

Table 3.1: List of locations included in this study

No.	Locations	Location codes	
		Year-2013	Year-2014
1	Okara	L1	S1
2	MMRI	L2	S2
3	Dhakkar	L3	S3
4	Bahawal nagar (BWN)	L4	S4
5	KSK	L5	S5
6	Gujranwala (GRW)	L6	S6
7	Kot Naina	L7	S7
8	Khanewal	L8	S8
9	Multan	L9	S9
10	Vehari	L10	S10
11	Karor	L11	S11
12	Sargodha	L12	S12
13	Piplan	L13	S13

Table 3.2: *List of Wheat genotypes included in this study*

No.	Genotype Name	Genotype Code
1	V-12266	1
2	V-11047	2
3	V-12284	3
4	V-11098	4
5	V-11061	5
6	V-11022	6
7	V-11092	7
8	NW.10.1111-7	8
9	TW11510	9
10	V-12275	10
11	GH	11
12	V-11137	12
13	NN-GAN-3	13
14	V-11138	14
15	V-11046	15
16	NR-411	16
17	V-11365	17
18	V-12265	18
19	Millat-11	19
20	Lasani-08	20
21	Nr-409	21
22	V-11143	22
23	V-11041	23
24	V-11032	24
25	NS-10	25
26	11BT004	26
27	9459-1	27
28	V-12304	28
29	11B2074	39
30	11B2049	30

The experimental layout was randomized complete block design (RCBD) with three replicates. The locations (Table 3.1) where experiment was conducted were different in soil type, further more years differentiate in term of mean seasonal rainfall. Therefore, location were considered as different environments.

Classical approach as well as Bayesian was applied. AMMI model was applied and for Bayesian analysis the multivariate von Mises-Fisher distribution was used as a prior for interaction parameters. First year data were used to elicit the prior to analyze second year. For analysis R software was used (R, 2015).

3.2 Classical Approach

3.2.1 AMMI model

The usual two way analysis of variance model is

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij} \quad (3.1)$$

where

- Y_{ij} = Observed response value of genotype i in environment j ,
- μ = Grand mean
- α_i = effect of genotype i , $i = 1, \dots, r$
- β_j = effect of environment j , $j = 1, \dots, c$
- $(\alpha\beta)_{ij}$ = interaction effect
- ε_{ij} = Error term

Parsimonious modeling of interaction can be considered by the singular value decomposition of $(\alpha\beta)_{ij}$ and by retaining only first few components. This give rise to the usual linear (additive) and bilinear (non-additive) two way model originally introduce by [GOLLOB \(1968 \)](#) and [Mandel \(1971 \)](#) and extensively used in plant breeding trial for assessing stability and adaptation ([Jos. Crossa, Yang, et al., 2004](#)). They expressed that the sum of squares for interaction can be further partitioned in multiplicative components related to eigen values. Such method of analysis of variance is called Additive Main effect and Multiplicative Interaction model (AMMI). Thus AMMI method integrates analysis of variance and principal component analysis in to a unified approach ([Bradu et al., 1978](#), [K. R. Gabriel, 1971](#), [Gauch et al., 1988](#)).

AMMI Can be used to analyze multi-location trials ([J. Crossa et al., 1990](#)). According to [Zobel et al. \(1988 \)](#), considering the three traditional models, analysis of variance fails to determine a significant interaction component, principal component

analysis (PCA) fails to identify and separate the significant genotype and environment main effects. But AMMI analysis reveals a highly significant interaction component that has a clear agronomic meaning and it does not require specific design, except for a two way data structure.

AMMI analysis first fit the additive main effects of genotypes and environments by the usual analysis of variance and the describe the non additive part, genotype-environment interaction, by principal component analysis. The AMMI is used for following three main purposes

1. Model diagnosis: AMMI is more appropriate in the initial Statistical analysis of yield trials, because it provides an analytical tool for diagnosing other models as sub cases when these are better for a particular dataset (Gauch *et al.*, 1988)
2. Genotype by environment interaction explanation: AMMi summarizes patterns and relationships of genotypes and environments (J. Crossa *et al.*, 1990).
3. Improvement in accuracy of yield estimates: To improve the accuracy of yield estimates that is equivalent to increasing the number of replicates by a factor of two to five (Zobel *et al.*, 1988). Such gains may be used to reduce costs by reducing the number of replications, to include more treatments in the experiments, or to improve efficiency in selecting the best genotypes.

The AMMI model with multiplicative terms may be written as

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^t \lambda_k u_{ik} v_{jk} + \epsilon_{ij} \quad (3.2)$$

Where,

- λ_k = The singular value for k-th principal component axis.
- u_{ik} = Element of left singular vector
- v_{jk} = Element of right singular vector
- ϵ_{ij} = Error term

With the side condition that

$$\sum_i u_{ik}^2 = \sum_j v_{jk}^2 = 1$$

and, for $k \neq k'$

$$\sum_i \alpha_{ik} \alpha_{ik'} = \sum_j \beta_{jk} \beta_{jk'} = 0$$

and $t = \min(r, c) - 1$.

The additive parameters are μ , α_i and β_j while, u_{ik} , v_{jk} and λ_k are multiplicative parameters. The u_{ik} are genotype interaction parameters that measure genotype sensitivity to hypothetical environmental factors denoted by environmental interaction parameters v_{jk}

In matrix notation it can be expressed as

$$\mathbf{Y} = \mu \mathbf{1}_r \mathbf{1}_c' + \boldsymbol{\alpha} \otimes \mathbf{1}_c' + \boldsymbol{\beta}' \otimes \mathbf{1}_r + \mathbf{U} \mathbf{D} \mathbf{V}' + \mathbf{E} \quad (3.3)$$

Where,

$$\begin{aligned} \mathbf{Y} &= [Y_{ij}] \\ \boldsymbol{\alpha} &= [\alpha_i] \\ \boldsymbol{\beta} &= [\beta_j] \\ \mathbf{D} &= \text{diag}(\lambda_k, k = 1, \dots, t) \\ \mathbf{U} &= (u_1, \dots, u_t) \\ u_k &= [u_{ik}] \\ \mathbf{V} &= (v_1, \dots, v_t) \\ v_k &= [v_{jk}] \end{aligned}$$

Note that from above model allowing $\boldsymbol{\alpha} = 0$ and $\boldsymbol{\beta} = 0$ other linear by linear model can be obtained.

Formal likelihood ratio tests have been proposed to determine the appropriate number of multiplicative terms that describe the interaction ([Hegemann et al., 1976](#)).

Milliken *et al.* (1989) have also provided comprehensive tables of critical points for the likelihood ratio test statistics. So the General analysis of variance for AMMI model is

The AMMI model combines the analysis of variance for the genotype and environment

Table 3.3: *The General Analysis of Variance for the AMMI Model*

Sources	df
Environment (E)	c-1
Rep(E)	c(b-1)
Genotype (G)	r-1
GE Interaction (GEI)	(c-1)(r-1)
PCA 1	$w_1=c+r-1-(2 \times 1)$
PCA 2	$w_2=c+r-1-(2 \times 2)$
.	.
.	.
.	.
PCA m	$w_n=c+r-1-(2 \times n)$
Exp. Error	c(r-1)(b-1)

effects with principal component analysis of $G \times E$ interaction. it has proven useful for understanding complex $G \times E$ interaction. The results can be graphed in a useful bi-plot that shows both main and interaction effects for both genotypes and environments. Bi-plot analysis of the GE interaction allows for visual inspection and interpretation of the underlying structure and causes of interaction (Zobel *et al.*, 1988, Bradu *et al.*, 1978). The concept of bi-plot was developed by K. R. Gabriel (1971) to graphically display a rank-two matrix. The significance of this concept is that if a two way data can be sufficiently approximated by a rank-two matrix, then it can be graphically displayed and investigated. Bradu *et al.* (1978) explored the use of bi-plot as a diagnostic tool for choosing an appropriate model for the analysis of two way data.

3.2.1.1 Determination of optimal number of interaction components

Non-additive effects are frequently observe in two-way tables, the interpreta-
tion is a problem if replicate observation are present for each cell of the table. Further
more it was pointed out that the non-additivity is often associated with just a few rows
and columns of the table. Hence, good prediction of true trait response in each cell of
two way table can be achieved by truncating the AMMI model.

Dias *et al.* (2003) present two methods outlined by (Krzanowski, 1987), based on 'leave-one-out' procedures that optimize the cross-validation process (i.e. maximizes the number of data points left in the set as each iteration without incurring bias due to re-substitution.), and Gollob method based on F-test. For this study the method suggested by GOLLOB (1968) will be used.

Table 3.4: *The General Analysis of Variance for the AMMI Model*

Sources	df
Environment (E)	c-1
Rep(E)	c(b-1)
Geneotype (G)	r-1
GE Interaction (GEI)	(c-1)(r-1)
PCA 1	$w_1=c+r-1-(2 \times 1)$
PCA 2	$w_2=c+r-1-(2 \times 2)$
.	.
.	.
.	.
PCA s	$w_m= c+r-1-(2 \times n)$
Residual of PCA	$[(c-1)(r-1)-\sum_{t=1}^s w_t]$
Exp. Error	c(r-1)(b-1)

GOLLOB, 1968 suggested using statistics

$$F = \frac{r\hat{\lambda}_k^2}{f_1 MS_E}$$

against an F distribution with $f_1=c+r-(2k)$ and $f_2=c(r-1)(b-1)$ degree of freedom to the k-th multiplicative term of model for significance. Therefore, selection of the optimal model is base on F test for successive terms of interaction. Hence, ANOVA model for truncated AMMI is as below,

3.2.1.2 AMMI Biplots

To generate a bi-plot that can be use in visual analysis of MET data, the singular values have to be partitioned into genotype and environment eigenvectors so that

equation (3.2) can be written in the form

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^t u_{ik}^* v_{jk}^* + \varepsilon_{ij} \quad (3.4)$$

where u_{ik}^* and v_{jk}^* are called interaction PCA axis 'k' scores for genotype i and environment j, respectively. in biplot, genotype i is displayed as a point define by all u_{ik}^* values, environment is displayed as a point defined by all v_{jk}^* values. singular value decomposition is implemented by $u_{ik}^* = \lambda_k^f u_{ik}$ and $v_{jk}^* = \lambda_k^{1-f} v_{jk}$, where f is the partitioning factor for PCA axis k and can ($0 \leq f \leq 1$). Among numerous way to construct biplot , symmetrical scaling (f=.5) is the method used in AMMI (Gauch *et al.*, 1988).

The results of AMMI can be presented graphically in the for of biplot (Shafii *et al.*, 1998, Ebdon *et al.*, 2002, Vargas *et al.*, 1999) in which the genotype and environment scores of first two or three multiplicative terms are represented by a vector in space, with the starting point at origin and end points determined by the scores. The distance between two genotype vector is indicative of the amount of interaction between the genotypes. The cosine of the angle between two genotypes (or environments) vector approximate the correlation between the genotypes or environments with respect to its interaction. The acute angle indicate positive correlation, while parallel vectors (in exactly same directions) representing a correlation of 1. Obtuse angle represent negative correlation, with opposite directions indicating a correlation -1. Perpendicular directions represents a correlation of 0. The relative amount of interaction for a particular genotype over environment can be obtained from orthogonal projection of the environment vectors on the line determined by the direction of corresponding genotype vector. Environmental vectors having same direction as the genotype vectors , have positive interactions where as vector in the opposite directions have negative interaction.

3.2.1.3 AMMI stability value (ASV)

The AMMI model does not make provision for a qualitative stability measures, yet such measures are essential in order to quantify and rank genotypes according to their yield stability. The AMMI stability value (ASV) was calculated using, the formula

suggested by [Purchase \(1997\)](#) to rank genotypes:

$$ASV = \sqrt{\left(\frac{SS\ IPC\ 1}{SS\ IPC\ 2}(IPC\ 1)\right)^2 + (IPC\ 2)^2}$$

Where

ASV = AMMI stability value

SS = Sum of square

IPC1 = Interaction principal component 1

IPC2 = Interaction principal component 2

The genotypes with lowest ASV values were considered the most stable. (SS-IPC1/SS-IPC2) is the weight given to the IPC-1 value by dividing IPC-1 sum of square by IPC-2 sum of squares. The larger the absolute IPC scores, the more specifically adapted a genotype to certain environments. Smaller IPC scores, indicate a more stable genotype across environments.

3.2.1.4 Genotype Selection Index (GSI)

Based on the rank of mean yield (Y_i) across environments and rank of AMMI stability value (ASV_i), selection index call GSI, was calculated for the genotypes using formula used by [Farshadfar, 2008](#):

$$GSI = \text{Rank of ASV} + \text{Rank of } Y_i$$

The least GSI was considered as the most stable with high seed yield

3.2.2 GGE-biplot

The biplot method originated by [K. R. Gabriel, 1971](#)), its use was subsequently expanded by [Kempton, 1984](#)) and [Zobel *et al.*, 1988](#)). The extensive useful-

ness of GE was recently elucidated by (Yan, Hunt, *et al.*, 2000).

The two way table of genotype \times environment data for wheat yield was analyzed using GGE biplot. The GGE biplot analysis produces biplots derived from principal components analysis of the environment centered data (data minus the environment means), which therefore represents the genotype main effect and $G \times E$ interaction. The genotypes and environments are represented by points on a two dimensional plot of principal component (PC) scores (PC 1 and PC 2). The distance of an environment from the biplot origin is measure of its ability to discriminate genotypes, the distance between two environments measures their dissimilarity in discriminating the genotypes, and angle between the environments represents their correlation. Acute angles represent positive correlation, obtuse angles show negative correlation and the right angles represents no correlation between the environments (Yan and Tinker, 2005).

In general, the proximity of genotypes to environments on the biplots is an indication of their similarity, and the proximity of genotypes to environments is an indication of degree of positive interaction with the environments. The GGE biplot provides a range of viewing option to investigate relationship between environments and genotypes, identify mega-environments, examine the representativeness of test environments as selection sites, rank genotypes based on performance in single environment.

3.2.2.1 *Biplot inner product property*

Mathematically, a biplot may be regarded as a graphical display of matrix multiplication. Given a matrix **G** with m rows and r columns and a matrix **E** with r rows and n columns, they can be multiplied to give a third matrix **P** with m rows and n columns. If $r=2$, then matrix **G** can be displayed as m points in a 2-D plot, with the first column as the abscissa (x-axis) and second column the ordinate (y-axis). Similarly, matrix **E** can be displayed as n points in 2-D plot, with the 1st row as a abscissa and 2nd row the ordinate. A 2-D biplot is formed if the two plots are superimposed, Which would contain m+n points. An interesting property of this biplot is that it not only displays matrices **G** and **E**, but also implicitly displays the $m \times n$ values of matrix **P**, because

each element of \mathbf{P} can be visualized as;

$$P_{ij} = x_i x'_j + y_i y'_j = |\mathbf{g}_i| |\mathbf{e}_j| \cos \theta_{ij} \quad (3.5)$$

Where (x_i, y_i) are the coordinates for row i and x'_j, y'_j are coordinates for column j . $|\mathbf{g}_i|$ is the vector length for row i and $|\mathbf{e}_j|$ is the vector length for column j . θ_{ij} is the angle between vectors of row i and column j .

Above equation is referred to as the inner-product property of the biplot. It is the most important property of a biplot. It not only allows each element of matrix \mathbf{P} to be estimated but also constitute the basis for visualizing the pattern in matrix \mathbf{P} , including ranking the rows relative to any column, ranking the columns relative to any row, comparing any two rows relative to individual columns, identifying the rows with largest (or smallest) value for each column, or vice versa.

3.2.2.2 *Data centering prior to singular value decomposition*

In a GE two-way table \mathbf{Y} , the value of each cell can be recorded as mixed effect of the grand mean(μ) modified by the genotype (row) main effect (α_i), the environment(columns) main effect (β_j), and the specific genotype(row) by environment (column) interaction $(\alpha\beta)_{ij}$, plus any random error ϵ_{ij} :

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij} \quad (3.6)$$

The matrix \mathbf{P} that is subject to SVD can be any part of \mathbf{Y} , resulting indifferent models, ignoring random error:

$$p_{ij} = Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad (3.7)$$

$$p_{ij} = Y_{ij} - \mu = \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad (3.8)$$

$$p_{ij} = Y_{ij} - \mu - \alpha_i = \beta_j + (\alpha\beta)_{ij} \quad (3.9)$$

$$p_{ij} = Y_{ij} - \mu - \beta_j = \alpha_i + (\alpha\beta)_{ij} \quad (3.10)$$

$$p_{ij} = Y_{ij} - \mu - \alpha_i - \beta_j = (\alpha\beta)_{ij} \quad (3.11)$$

3.2.2.3 *Data scaling and singular values decomposition and Singular values partitioning*

The GGE biplot model can be generally presented as:

$$P_{ij} = (Y_{ij} - \mu - \beta_j)/s_j = (\alpha_i + (\alpha\beta)_{ij})/s_j \quad (3.12)$$

Where s_j is scaling factor. Thus there can be different GGE models, depending on the definition of s_j . When s_j is the standard deviation for column j, the data is said to "standardized" because all columns are give same weight. when s_j is the standard error in environment j, any heterogeneity among the environments will be removed. Replicated data are essential for estimating standard error in each environment. Data standardization is essential for biplot analysis of two-way tables in which the columns are of different units or scales. GGE biplot allows each of the above models to be scaled in various ways.

The practical application of a biplot in data analysis was started most clearly by the founder of biplot ([K. R. Gabriel, 1971](#)): any two way table can be graphically analyzed using a 2-D biplot as soon as it can be sufficiently approximated by a rank-2 matrix. Give a genotype by environment two way table \mathbf{P} of g genotypes and e environments, biplot analysis started with its decomposition into three matrices \mathbf{G} , \mathbf{L} and \mathbf{E} via SVD;

$$\mathbf{P}_{(g \times e)} = \mathbf{G}_{(g \times r)} \mathbf{L}_{(r \times r)} \mathbf{E}'_{(e \times r)} \quad (3.13)$$

Matrix \mathbf{G} has g row and r columns; it characterizes the m genotypes. Matrix \mathbf{E} has r rows and e columns; it characterizes the e environments. Matrix \mathbf{L} is a diagonal matrix containing r singular values. in summation notation, SVD decomposes \mathbf{P} into r principal components (PC), each a genotype vector (u_i), an environment vector (v_j) and singular values (λ):

$$P_{ij} = \sum_{k=1}^r \lambda_k u_{ik} v_{jk} \quad (3.14)$$

Where r is the rank of the two way table. Another requirement of SVD is that $\mathbf{G}'\mathbf{G} = \mathbf{I}_r = \mathbf{E}'\mathbf{E}$, where \mathbf{I}_r is the r by r identity matrix. The singular values must be partitioned into genotype and environment scores before a biplot can be constructed.

$$P_{ij} = \sum_{k=1}^r u_{ik}^* v_{jk}^* = \sum_{k=1}^r (\lambda_k^f u_{ik}) (\lambda_k^{1-f} v_{jk}) \quad (3.15)$$

Where f is the partitioning factor, which can be between 0 and 1. Among number of ways of singular value partitioning, column-metric preserving, row-metric preserving and symmetrical method Gabriel 2002, Yan 2002 are mostly used and are discussed below.

3.2.2.4 Column-Metric preserving and associated interpretation

When $f=0$, the singular values are entirely partitioned into the column (here environment) eigenvectors referred to as column-metric preserving or environment-focused scaling. Since $\mathbf{E}^* = \mathbf{E}\mathbf{L}$, we have $\mathbf{E}^*\mathbf{E}^{*'} = \mathbf{P}\mathbf{P}'$, which is sum of squares and cross product matrix of \mathbf{P} . if \mathbf{P} is column-centered then $\mathbf{P}\mathbf{P}'$ is $(r-1)$ times of covariance matrix. Therefore partitioning is appropriate for studying the relationship among column factors.

3.2.2.5 Row-Metric preserving and associated interpretation

When $f=1$, the singular values are entirely into row eigenvectors, which is referred as row preserving or genotype focused scaling. Since $\mathbf{G}^* = \mathbf{G}\mathbf{L}$, we have $\mathbf{G}^*\mathbf{G}^{*'} = \mathbf{P}\mathbf{P}'$. Therefore, this partitioning recovers the Euclidean distance among row factor (here genotype), is appropriate for visualizing the similarity/dissimilarity among row factors. Because all of Singular values are partitioned into genotypes scores, the range of the genotype scores are likely to be many times greater than that of the environment scores. As a result, the environments in biplot are likely to be crowded relative to the genotypes.

3.2.2.6 *Symmetrical partitioning and associated interpretation*

When $f=0.5$, it is called symmetrical scaling. This type of scaling has the unique property that genotype scores have the same unit for both PCA axis 1 and PCA axis 2, which is square root of original yield. This property makes it possible to visualize the relative magnitude of genotype variation and environment variation for both axes.

3.3 *Bayesian Approach*

As discussed earlier classical methods offer opportunities for GGE or GE in modeling and can explain it very well but there are some limitations of these methods.

- principal component analysis (PCA) fails to identify and separate the significant genotype and environment main effects ([Zobel *et al.*, 1988](#))
- Fixed effect AMMI model have no inferential statistics attached to the interaction parameters used to build the biplot ([Yang *et al.*, 2009](#)).
- A method to obtain Confidence regions in the biplot representation in frequentist frame work was proposed by ([Denis *et al.*, 1996](#)). It was based on asymptotic consideration; however, these confidence regions are not implemented for models with more than two bilinear terms.
- The AMMI model applies singular value decomposition (SVD) to the residuals of a specific linear model, to decompose the genotype-by-environment interaction into a sum of multiplicative terms. However, SVD being a least squares method is highly sensitive to contamination and the presence of even a single outliers may draw the leading principal component towards itself resulting in possible mis-interpretations and in turn lead to bad practical decisions ([Rodrigues *et al.*, 2015](#)).

To overcome these problems ([PEREZ-ELIZALDE *et al.*, 2011](#)) suggested a Bayesian

treatment of AMMI models in the frame work of GE data. Two main points motivate this approach

- Firstly, a Bayseian strategy offers the possibility to incorporate prior information on the phenomenon under study (experts knowledge, historical data, etc.).
- Secondly, distributions of any quantity of interest are available through the prior distribution. These distribution may be straight-forwardly used, for instance to, to derive credible areas in the biplot representation, which can be helpful for the interpretation.
- Another strong point of the Bayesian approach is that it can be straightforwardly used on incomplete datasets. Consequently, an interesting research perspective of our work is to assess our proposal in relation to unbalanced data. Many methods are available to deal with missing values in bilinear models. A common approach to obtain point estimates for parameters consists in using alternating weighted least squares algorithms or iterative imputation algorithms (K. Ruben. Gabriel *et al.*, 1979, Kiers, 1997, Gauch *et al.*, 1990).
- Estimating the variability of the parameters in a missing data framework is more difficult, and some authors Adams *et al.* (2002) and Josse *et al.* (2011) suggested approaches based on bootstrap simulations. Using a Bayesian point of view will directly provide credible regions for the parameters through the posterior distributions, which is very appealing. In addition, credible regions will also be available for missing entries. We mention that draw a parallel between the way missing values are handled and the over-parameterization issue.

The basis of all bayesian statistics is Bayes's theorem, which is way back originated by (Bayes, 1763).

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

Once the data has been observed, the likelihood function, or simply the likelihood, is constructed. The likelihood is the joint probability function of the data, but

viewed as a function of the parameters, treating the observed data as fixed quantities. Assuming that the data values, $y = (y_1, \dots, y_n)$ are obtained independently, the likelihood function is given by

$$L(\theta|y) = p(y_1, \dots, y_n|\theta) \prod_{i=1}^n p(y_i|\theta)$$

In the Bayesian framework, all of the information about θ coming directly from the data is contained in the likelihood. Values of the parameters that correspond with the largest values of the likelihood are the parameters that are most supported by the data (Glickman *et al.*, 2007).

To obtain the posterior distribution, $p(\theta|y)$, the probability distribution of the parameters once the data have been observed, we apply Bayes' theorem:

$$P(\theta|y) = \frac{p(\theta)(\theta|y)}{\int p(\theta)p(y|\theta)d\theta}$$

In this study von Mises-Fisher distribution is used as prior and given below.

3.3.1 Von Mises-Fisher distribution

The set $r \times k$ orthonormal matrices is called the Stiefel manifold, which is denoted as $v_{k,r}$. A probability distribution on $v_{k,r}$, whose density has exponential form with linear and quadratic terms, is the matrix Bingham-Von Mises distribution. The density function is given by

$$p(X|\mathbf{A}, \mathbf{B}, \mathbf{C}) \propto \text{etr}(\mathbf{C}' \mathbf{X} + \mathbf{B} \mathbf{X}' \mathbf{A} \mathbf{X})$$

Where \mathbf{A} and \mathbf{B} may be Assumed symmetric and diagonal matrices, respectively. A random variable \mathbf{X} with Von Mises Fisher distribution is denoted as $\mathbf{X} \sim BFM(\mathbf{A}=\mathbf{0}, \mathbf{B}=\mathbf{0}, \mathbf{C})$ (Khatri *et al.*, 1977). The normalization of Von Mises-Fisher density is given by the hyper-geometric function of a matrix argument $F_1(\frac{1}{2}\gamma, \frac{1}{4}D_\phi^2)$, Where D_ϕ is the diagonal matrix of singular values of \mathbf{C} (Herz, 1955, James, 1964)

3.3.2 Likelihood Function

The likelihood function for parameters of model (3.3) is

$$L(\mu, \alpha, \beta, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau | \mathbf{Y}) \propto \tau^{\frac{nrc}{2}} \exp\left\{-\frac{\tau}{2} [n \operatorname{tr}(\mathbf{E}\mathbf{E}') + (n-1)\operatorname{tr}(\mathbf{S}\mathbf{S}')]\right\} \quad (3.16)$$

where

$$\tau = 1/\sigma^2$$

$$\mathbf{S} = \{\sqrt{s_{ij}^2}\}$$

$$s_{ij}^2 = \sum_{l=1}^n \frac{(\bar{Y}_{ij} - Y_{ijl})^2}{n-1}$$

$$\mathbf{E} = \mathbf{Y} - \mu \mathbf{1}_r \mathbf{1}_c' - \alpha \otimes \mathbf{1}_c' - \beta' \otimes \mathbf{1}_r - \mathbf{U}\mathbf{D}\mathbf{V}'$$

3.3.3 Prior Distribution

For assessing the prior distributions of the unknowns, we used conditional conjugate prior distributions such that the posterior distribution is proper and can be used to incorporate valuable prior information from experimenter expertise or from information generated by previous trials. Note that, since the matrices \mathbf{U} and \mathbf{V} are orthonormal and \mathbf{D} is diagonal,

$$\begin{aligned} \operatorname{tr}(\mathbf{1}_r \mathbf{1}_c' \mathbf{V}\mathbf{D}\mathbf{U}') &= 0 \\ \operatorname{tr}((\alpha \otimes \mathbf{1}_c') \mathbf{V}\mathbf{D}\mathbf{U}') &= 0 \\ \operatorname{tr}((\beta' \otimes \mathbf{1}_r) \mathbf{V}\mathbf{D}\mathbf{U}') &= 0 \\ \operatorname{tr}((\mathbf{U}\mathbf{D}\mathbf{V}')(\mathbf{U}\mathbf{D}\mathbf{V}')') &= \operatorname{tr}(\mathbf{D}'\mathbf{D}) = \sum_{k=1}^t \lambda_k^2 \\ \operatorname{tr}\{(-2\mathbf{Y} + \mathbf{U}\mathbf{D}\mathbf{V}')(\mathbf{U}\mathbf{D}\mathbf{V}')'\} &= \operatorname{tr}\{(\mathbf{D} - \mathbf{U}'\mathbf{Y}\mathbf{V})'(\mathbf{D} - \mathbf{U}'\mathbf{Y}\mathbf{V}) - (\mathbf{U}'\mathbf{Y}\mathbf{V})'(\mathbf{U}'\mathbf{Y}\mathbf{V})\} \end{aligned} \quad (3.17)$$

Thus it can be shown from (3.3) that, given $\theta = (\mu, \alpha, \beta)$ and τ , the conditional likelihood function for the matrices $(\mathbf{U}, \mathbf{D}, \mathbf{V})$ is

$$\begin{aligned} L(\mathbf{U}, \mathbf{V}, \mathbf{D} | \theta, \tau, \mathbf{Y}) &= L(\mathbf{U}, \mathbf{V}, \mathbf{D} | \tau, \mathbf{Y}) \propto \exp\left\{-\frac{n\tau}{2} \text{tr}((-2\mathbf{Y} + \mathbf{U}\mathbf{D}\mathbf{V}')(\mathbf{U}\mathbf{D}\mathbf{V}')')\right\} \\ &= \text{etr}\left\{-\frac{n\tau}{2} \text{tr}((-2\mathbf{Y} + \mathbf{U}\mathbf{D}\mathbf{V}')(\mathbf{U}\mathbf{D}\mathbf{V}')')\right\} \end{aligned} \quad (3.18)$$

where 'etr' is the exponential for the trace. The conditional likelihoods for \mathbf{U}, \mathbf{V} and \mathbf{D} are

$$L(\mathbf{U}, \mathbf{V}, \mathbf{D} | \tau, \mathbf{Y}) \propto \text{etr}\{n\tau \mathbf{Y}\mathbf{V}\mathbf{D}\mathbf{U}'\} \quad (3.19)$$

$$L(\mathbf{U}, \mathbf{V}, \mathbf{D} | \tau, \mathbf{Y}) \propto \text{etr}\{n\tau \mathbf{Y}'\mathbf{U}\mathbf{D}\mathbf{V}'\} \quad (3.20)$$

$$\begin{aligned} L(\mathbf{U}, \mathbf{V}, \mathbf{D} | \tau, \mathbf{Y}) &\propto \text{etr}\left\{-\frac{n\tau}{2} (\mathbf{D} - \mathbf{U}'\mathbf{Y}\mathbf{V})'(\mathbf{D} - \mathbf{U}'\mathbf{Y}\mathbf{V})\right\} \\ &\propto \text{etr}\left\{-\frac{n\tau}{2} \sum_{k=1}^t (\lambda_k - l_k)^2\right\}, \lambda_1 > \lambda_2 > \dots > \lambda_t \end{aligned} \quad (3.21)$$

respectively, where

$$(l_1, \dots, l_t) = \text{diag}(\mathbf{U}'\mathbf{Y}\mathbf{V}).$$

From expression (3.19) it follows that a conditional conjugate prior for \mathbf{U} is

$$\pi(\mathbf{U} | \tau) \propto \text{etr}\{\tau n_0 \mathbf{Y}_0 \mathbf{V}_0 \mathbf{D}_0 \mathbf{U}'\} \quad (3.22)$$

where \mathbf{Y}_0 could be interpreted as the matrix of prior cell average such that

$$\mathbf{Y}_0 = \mu_0 \mathbf{1}_r \mathbf{1}_c' + \alpha_0 \otimes \mathbf{1}_c' + \beta_0' \otimes \mathbf{1}_r + \mathbf{U}_0 \mathbf{D}_0 \mathbf{V}_0'$$

that is, \mathbf{D}_0 is diagonal matrix of prior singular values, and \mathbf{U}_0 and \mathbf{V}_0 are orthonormal matrices such that $\mathbf{U}_0 \mathbf{D}_0 \mathbf{V}_0'$ is the singular value decomposition (SVD) of $\mathbf{Z}_0 = \mathbf{Y}_0 - \mu_0 \mathbf{1}_r \mathbf{1}_c' - \alpha_0 \otimes \mathbf{1}_c' - \beta_0' \otimes \mathbf{1}_r$, where μ_0, α_0 and β_0 are prior values for linear effects.

Similarly, from (3.20), a prior for \mathbf{V} is

$$\pi(V|\tau) \propto \text{etr}\{\tau n_0 \mathbf{Y}'_0 \mathbf{U}_0 \mathbf{D}_0 \mathbf{V}'\} \quad (3.23)$$

Both (3.22) and (3.23) are von Mises–Fisher distributions. From (3.21) for each one of the elements $\lambda_1 > \lambda_2 > \dots > \lambda_k$ on the diagonal of \mathbf{D} , the conditional conjugate prior distribution are left truncated normal with marginal densities of the form

$$\pi(\lambda_k|\tau) = \{1 - \Phi(\sqrt{n_0\tau}(\lambda_{k+1} - l_k^0))\}^{-1} N(\lambda_k | l_k^0 (n_0\tau)^{-1}) \quad (3.24)$$

For linear terms $\boldsymbol{\theta} = (\mu, \boldsymbol{\alpha}, \boldsymbol{\beta})$ of model (3.1), a conditional conjugate prior is a (1+r+c) multivariate normal distribution with mean $\boldsymbol{\theta}_0 = (\mu_0, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and singular block diagonal covariance matrix

$$(n_0\tau)^{-1} \begin{bmatrix} (r_0 c_0)^{-1} & 0 & 0 \\ 0 & c_0^{-1} \mathbf{K}_r \mathbf{K}_r' & 0 \\ 0 & 0 & r_0^{-1} \mathbf{K}_c \mathbf{K}_c' \end{bmatrix}$$

where \mathbf{K}_w is a matrix such that $\mathbf{K}_w' \mathbf{K}_w = \mathbf{I}_{w-1}$ and $\mathbf{K}_w' \mathbf{K}_w = \mathbf{I}_w - \frac{1}{w} \mathbf{J}_w$, where \mathbf{J}_w is a $w \times w$ matrix with all its elements equal to one. Because of the restrictions $\boldsymbol{\alpha}' \mathbf{1}_r = 0$ and $\boldsymbol{\beta}' \mathbf{1}_c = 0$, the distribution characterized by the covariance matrix above is a singular multivariate normal distribution that does not have a density. For a prior density we need to choose a one to one transformation like $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = (\mathbf{K}_r' \boldsymbol{\alpha}, \mathbf{K}_c' \boldsymbol{\beta})$.

Let $\boldsymbol{\theta}^* = (\mu, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$; then the prior density of $\boldsymbol{\theta}^*$ is

$$\begin{aligned} \pi(\boldsymbol{\theta}^*|\tau) &\propto |\Sigma_0|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0^*)' \Sigma_0^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\theta}_0^*)\right\} \\ \Sigma_0 &= (n_0\tau)^{-1} \begin{bmatrix} (r_0 c_0)^{-1} & 0 & 0 \\ 0 & c_0^{-1} \mathbf{I}_{r-1} & 0 \\ 0 & 0 & r_0^{-1} \mathbf{I}_{c-1} \end{bmatrix} \end{aligned} \quad (3.25)$$

which is the density of a (1+ r-1+c-1) multivariate normal distribution with mean,

$$\boldsymbol{\theta}_0^* = (\mu_0, \boldsymbol{\alpha}_0^*, \boldsymbol{\beta}_0^*) = (\mu, \mathbf{K}_r' \boldsymbol{\alpha}_0, \mathbf{K}_c' \boldsymbol{\beta}_0)$$

$$\Sigma_0 = \text{covariance matrix}$$

The joint likelihood (3.16) suggests that conjugate prior for τ is gamma distribution with parameters $a/2$ and $s_0^2/2$; that is

$$\pi(\tau) \propto \tau^{\frac{a}{2}-1} \exp\left\{-\frac{as_0^2}{2}\tau\right\} \quad (3.26)$$

or equivalently, $as_0^2\tau \sim \chi_a^2$; thus a and s_0^2 may be considered as prior values for sample size and variance, respectively. Finally, the joint prior for $(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau)$ is :

$$\pi(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau) = \pi(\boldsymbol{\theta}^*|\tau)\pi(\mathbf{U}|\tau)\pi(\mathbf{D}|\tau)\pi(\mathbf{V}|\tau)\pi(\tau) \quad (3.27)$$

The proposed prior has practical advantages and is flexible enough to incorporate prior uncertainty about unknown parameters. On the other hand, the main disadvantage of the prior used by [Jarquin et al. \(2011\)](#) for implementing their Bayesian approach was the elicitation of the distribution of each element on the matrices given by the SVD decomposition of the interaction.

In our proposal, the incorporation of prior information is straightforward and intuitive, as it only needs to express our beliefs in the prior cell averages \mathbf{Y}_0 and prior linear effects $\boldsymbol{\theta}_0$. Then $\mathbf{U}_0, \mathbf{V}_0$ and \mathbf{D}_0 are obtained from the SVD decomposition of \mathbf{Z}_0 , under the restrictions $\mathbf{U}_0' \mathbf{1}_r = 0$ and $\mathbf{V}_0' \mathbf{1}_c = 0$. The prior distribution of the linear effects is completely specified by giving a belief $\boldsymbol{\theta}_0^*$ about $\boldsymbol{\theta}^*$. This prior belief may be expressed as a function of \mathbf{Y}_0 ; for example, $\boldsymbol{\theta}_0^*(\mathbf{Y}_0) = (\frac{\mathbf{1}' \mathbf{Y}_0 \mathbf{1}_c}{r_0 c_0}, \frac{\mathbf{K}' \mathbf{Y}_0 \mathbf{1}_c}{c_0}, \frac{\mathbf{K}' \mathbf{Y}_0 \mathbf{1}_r}{r_0})$. Also, it is important to note that a vague prior for ' τ ' implies diffuse priors for all the other parameters. Then, an objective or default Bayesian analysis could be performed by setting small values to the hyper-parameter a in the prior for τ given by (3.26). There for , we may summarize our prior information by giving, a prior, a prediction of the two-way array of means \mathbf{Y}_0 and a measure of our prior uncertainty s_0^2 given a prior sample size ' a '.

3.3.4 Posterior Distribution And GIBBS Sampler

The joint posterior distribution is obtained by combining the likelihood function that is

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

so here the joint posterior distribution is obtained by combining the likelihood function (3.16) and the prior distribution (3.27),

$$\pi(\theta^* | \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau | \mathbf{Y}) \propto L(\theta^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau | \mathbf{Y}) \pi(\theta^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau)$$

where $L(\theta^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau | \mathbf{Y})$ is a re-parameterization of $L(\theta, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau | \mathbf{Y})$. The marginal posterior distribution, which involves high dimensional integration on complex spaces, is needed for marginal inference about the unknowns. In order to use a Markov Chain Monte Carlo (MCMC) method through the Gibbs sampler, the full conditional posterior distributions, which are known except for the proportionality constants, are needed. These distributions are computed by considering the joint posterior as a function only of a variable when fixing the others. Thus, the conditional posterior for θ^* is

$$\pi(\theta^* | \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau, \mathbf{Y}) \propto L(\theta^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau | \mathbf{Y}) \pi(\theta^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau)$$

Knowing that the conditional likelihood function of $(\mathbf{U}, \mathbf{V}, \mathbf{D})$ does not depend on θ^* , and that given τ , the prior for θ^* is independent of $(\mathbf{U}, \mathbf{V}, \mathbf{D})$, then

$$\begin{aligned} \pi(\theta^* | \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau, \mathbf{Y}) &= \pi(\theta^* | \tau, \mathbf{Y}) \propto L(\theta^* | \tau, \mathbf{Y}) \pi(\theta^* | \tau) \\ &\propto L(\theta^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau | \mathbf{Y}) \pi(\theta^* | \tau) / L(\mathbf{U}, \mathbf{V}, \mathbf{D} | \tau, \mathbf{Y}) \\ &\propto \exp\left\{-\frac{n\tau}{2} \text{tr}(\mathbf{Z}\mathbf{Z}')\right\} \pi(\theta^* | \tau) \end{aligned}$$

where,

$$\mathbf{Z} = \mathbf{Y} - \mu \mathbf{1}_r \mathbf{1}_c' - \alpha \otimes \mathbf{1}_c' - \beta' \otimes \mathbf{1}_r |_{\alpha = \mathbf{K}_r \alpha^*, \beta = \mathbf{K}_c \beta^*}$$

it can be shown that the conditional posterior of θ^* is multivariate normal with density

:

$$\begin{aligned}
\pi(\boldsymbol{\theta}^*|\boldsymbol{\tau}, \mathbf{Y}) &= N_{r+c-1}(\boldsymbol{\theta}^*|\boldsymbol{\theta}_n^*, \boldsymbol{\Sigma}_n^*) \\
\text{covariance matrix} &= \boldsymbol{\Sigma}_n^* = (\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_n^{-1}) \\
\text{mean} &= (\boldsymbol{\theta}^* = (\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_n^{-1}) \times (\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_0^* + \boldsymbol{\Sigma}_n^{-1}\hat{\boldsymbol{\theta}}_n^*))
\end{aligned}$$

and with

$$\hat{\boldsymbol{\theta}}_n^* = \left(\frac{\mathbf{1}'\mathbf{Y}\mathbf{1}_c}{rc}, \frac{\mathbf{K}'\mathbf{Y}\mathbf{1}_c}{c}, \frac{\mathbf{K}'\mathbf{Y}'\mathbf{1}_r}{r} \right)$$

and

$$\boldsymbol{\Sigma}_n = (n\boldsymbol{\tau})^{-1} \begin{bmatrix} (rc)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & c^{-1}\mathbf{I}_{r-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & r^{-1}\mathbf{I}_{c-1} \end{bmatrix}$$

We may use the conditional likelihoods given by (3.19)–(3.21) to calculate the conditional posterior for \mathbf{U}, \mathbf{V} and \mathbf{D} ; that is,

$$\begin{aligned}
\pi(\mathbf{U}|\mathbf{V}, \mathbf{D}, \boldsymbol{\tau}, \mathbf{Y}) &\propto L(\mathbf{U}|\mathbf{V}, \mathbf{D}, \boldsymbol{\tau}, \mathbf{Y})\pi(\mathbf{U}|\boldsymbol{\tau}) \\
&\propto \text{etr}\{\boldsymbol{\tau}(n_0\mathbf{Y}_0\mathbf{V}_0\mathbf{D}_0 + n\mathbf{YVD})\mathbf{U}'\} \quad (3.28)
\end{aligned}$$

$$\begin{aligned}
\pi(\mathbf{V}|\mathbf{U}, \mathbf{D}, \boldsymbol{\tau}, \mathbf{Y}) &\propto L(\mathbf{V}|\mathbf{U}, \mathbf{D}, \boldsymbol{\tau}, \mathbf{Y})\pi(\mathbf{V}|\boldsymbol{\tau}) \\
&\propto \text{etr}\{\boldsymbol{\tau}(n_0\mathbf{Y}'_0\mathbf{U}_0\mathbf{D}_0 + n\mathbf{Y}'\mathbf{UD})\mathbf{V}'\} \quad (3.29)
\end{aligned}$$

$$\begin{aligned}
\pi(\mathbf{D}|\mathbf{U}, \mathbf{V}, \boldsymbol{\tau}, \mathbf{Y}) &\propto L(\mathbf{D}|\mathbf{U}, \mathbf{V}, \boldsymbol{\tau}, \mathbf{Y})\pi(\mathbf{D}|\mathbf{U}, \mathbf{V}, \boldsymbol{\tau}) \\
&\propto \exp\left\{-\frac{n\boldsymbol{\tau}}{2} \sum_{k=1}^t (\lambda_k - l_k)^2\right\} \prod_{k=1}^t \pi(\lambda_k|\boldsymbol{\tau}) \\
&\propto \prod_{k=1}^t \left\{1 - \Phi\left(\sqrt{\frac{\boldsymbol{\tau}-1}{n_0+n}}\left(\lambda_{k+1} - \frac{n_0 l_k^0 + n l_k}{n_0+n}\right)\right)\right\}^{-1} \\
&\times N\left(\lambda_k \middle| \frac{n_0 l_k^0 + n l_k}{n_0+n}, \frac{\boldsymbol{\tau}^{-1}}{n_0+n}\right)
\end{aligned}$$

$$\lambda_1 > \lambda_2 > \dots > \lambda_t > \lambda_{t+1} = 0$$

Finally, the conditional posterior for the percision of τ is a gamma with density

$$\begin{aligned}\pi(\tau|\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \mathbf{Y}) &\propto L(\boldsymbol{\theta}^*, \mathbf{U}, \mathbf{V}, \mathbf{D}, \tau|\mathbf{Y})\pi(\tau) \\ &\propto Ga(\tau|\frac{a_n}{2}, \frac{b_n}{2})\end{aligned}$$

where

$$\begin{aligned}a_n &= a + nrc \\ b_n &= as_0^2 + n(tr(\mathbf{E}\mathbf{E}')) + (n-1)tr(\mathbf{S}\mathbf{S}')$$

3.3.4.1 Gibbs Sampler

The Gibbs Sampler is implemented by sequentially drawing simulated samples from the full conditional posterior distributions; thus we may proceed with the following algorithm. Let 's' be the desired length of the Markov chain to be simulated. Let $\mathbf{U}^{(0)}, \mathbf{V}^{(0)}$ and $\mathbf{D}^{(0)}$ be the initial values of the simulated Markov chain. For $i = 0, \dots, s$ simulate

$$\begin{aligned}\tau^{(i+1)} &\sim \pi(\tau|\boldsymbol{\theta}^{*(i)}, \mathbf{U}^{(i)}, \mathbf{D}^{(i)}, \mathbf{Y}) \\ \boldsymbol{\theta}^{*(i+1)} &\sim \pi(\boldsymbol{\theta}^*|\tau^{(i+1)}, \mathbf{Y}) \\ \mathbf{U}^{(i+1)} &\sim \pi(\mathbf{U}|\tau^{(i+1)}) \\ \mathbf{D}^{(i+1)} &\sim \pi(\mathbf{D}|\tau^{(i+1)}) \\ \mathbf{V}^{(i+1)} &\sim \pi(\mathbf{V}|\tau^{(i+1)})\end{aligned}$$

After a burn-in period, we assume that the generated samples arise from the stationary distribution, i.e., joint posterior distribution $\pi(\boldsymbol{\theta}^*, U, V, D, \tau|Y)$. For reversibility of the Markov chain, we may permute the order of simulation, but in what follows we use the order indicated above. Some standard convergence diagnostic tools may be used to determine an effective sample size; in the examples below, criteria for convergence of simulated Markov chains from [Raftery *et al.* \(1995\)](#) and [Gelman and Rubin \(1992\)](#) were used.

Chapter 4: Results and Discussion

In this chapter result of two different method are given. First section contain the results of classical methods while in second second section Bayesian analysis was performed.

4.1 Classical approaches analysis

In this section Combine ANOVA, AMMI model, Cluster analysis and GGE bi-plot methods are applied for the diagnosis of wheat crop data which comprises of 30 genotypes tested at 13 locations for two year.

4.1.1 Combined Analysis of Variance

The combined analysis of variance (ANOVA) of thirty genotypes in 13 environment for 2 years was performed and output is provided in (Table 4.1). Analysis showed that genotype and environment were highly significant, which mean these effect have significant differences among them ($p < 0.01$). Significance of $G \times E$ interaction ($p < 0.01$) indicated that genotypes have different responses in different environments. Environment, genotypes and $G \times E$ explained 84.3%, 3.5% and 8.6 % variation of sum of square.

Table 4.1: Combined analysis of variance (ANOVA)

Sources	DF	SS	MS	p-value	% Sum of square
Environment (E)	25	2226401709	89056068	0.00	84.3
Rep(ENV)	52	3653099	70252	0.1861	
Geneotype (G)	29	92402025	3186277	0.00	3.5
$G \times E$	725	227172934	313342	0.00	8.6
Residual	1508	90129189	59767		

From Combined ANOVA provided in (Table 4.1), it is evident that $G \times E$ have

significant effect hence the identification of superior genotypes across environment is not possible on the basis of just performance of mean yield. Furthermore, ANOVA unable to partition the interaction in various component, so other stability measures are required for to explore $G \times E$.

4.1.2 AMMI analysis and bi-plot representation

As (Table 4.1) indicated that main effect as well as interaction effect were significant, so $G \times E$ is partitioned into bilinear terms. The IPCA are arrange in ascending order on their importance give in table (Table 4.2). $G \times E$ interaction component significance was measure using the Gollob's F-test at 0.01 probability. First four principal components explained 59.2% variation of GEI sum of squares. Hence, AMMI-4 model can be regarded as best fit model for the wheat crop data.

Table 4.2: Combined analysis of variance (ANOVA) according to the AMMI model and Gollob's test of interaction PC's

Sources	DF	SS	MS	p-value	$G \times E$ (%)	Cumulative(%)
Environment (E)	25	2226401709	89056068			
Rep(ENV)	52	3653099	70252			
Genotype (G)	29	92402025	3186277			
$G \times E$	725	227172934	313342			
IPCA 1	53	53024110	1000455	0.00	23.3	23.3
IPCA 2	51	31803670	623601	0.00	14.0	37.3
IPCA 3	49	27472720	560667	0.00	12.1	49.4
IPCA 4	47	22299260	474452	0.00	9.8	59.2
IPCA Residual	525	92573174	176329.8	0.00		
Residual	1508	90129189	59767			

IPCA = Interaction Principal Component Axis

Although, first four component were found significant, but to avoid from complexity first two principal component can be used to analyze result using AMMI bi-plot. For predictive model two principal components for interaction were adequate. Whereas, other components mostly contain the non-predictive variation. Therefore, the interaction of thirty genotypes of wheat along with 26 environments was by first two component, as [KAYA et al., 2002](#) suggested that predictive criterion choose such model which consist of two interaction principal components.

First IPC axis explained 23.3% variation while 14.0% was captured by the second principal component for interaction sum of square. First two PC cumulatively explain 37.3% variation for interaction sum of squares with 104 degree of freedom out of 725 was an indication that sufficient percentage of $G \times E$ explained. So further diagnosis can be done by graphing a biplot using first two IPCA.

4.1.2.1 AMMI bi-plot

The biplot of PCA-1 versus the PC1-2 containing 37.3% variation is displayed in (Figure 4.1). Important information related to the interaction can obtained from the direction of genotypes and environments. From plot, it can be observe that 11BT004 has a positive interaction with S10(Vehari-14) and L10(Vehari-13) and a negative interaction with S13(Piplan-14) and S13(piplan-13). Genotypes NR-411, 11BT004 and V-12304 having high PC1 and PC2 score shown that they have significant effect in interaction.

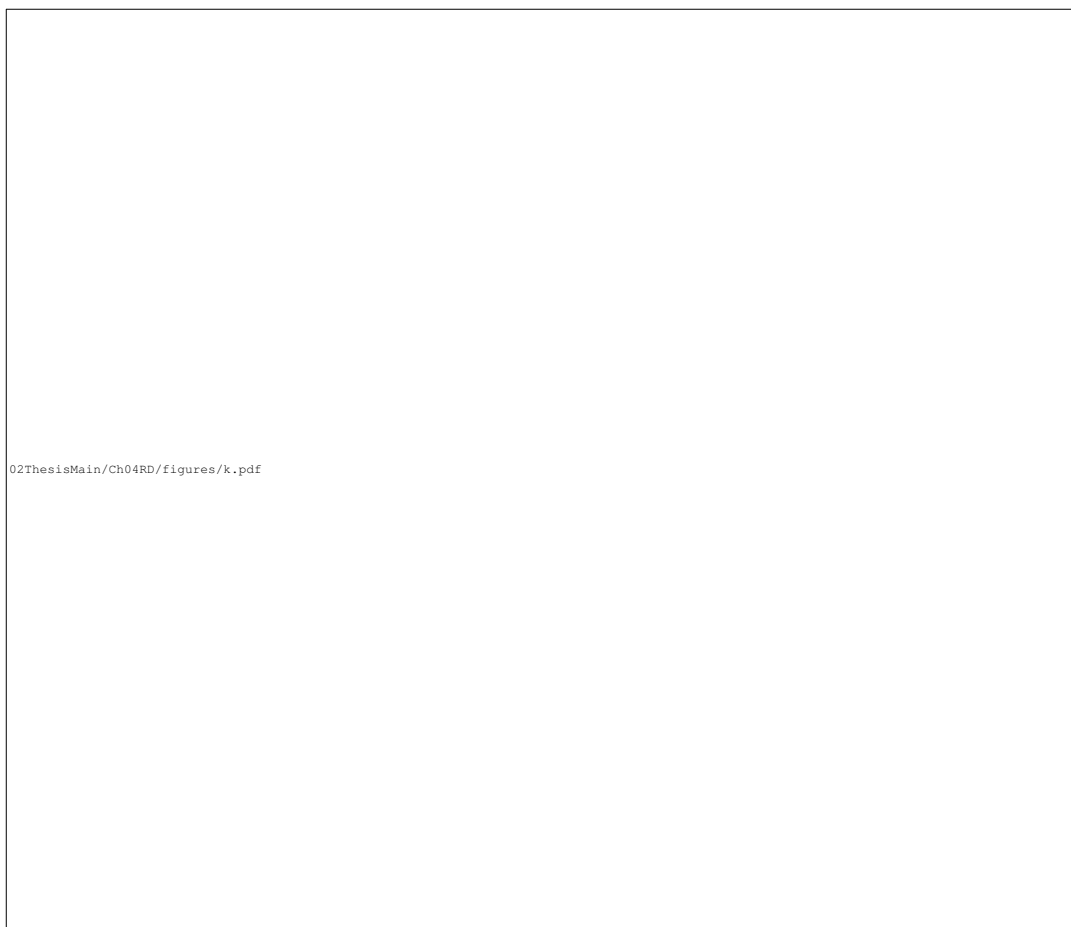


Figure 4.1: *Biplot of interaction principal component analysis (PCA) axis-1 versus axis-2 of wheat data for 30 genotypes in 26 environments*

The second biplot was of IPCA axis-1 versus the mean yield of environments and genotypes [Figure 4.2]. The low yielding genotypes were located in upper left and environments are shown in lower left quadrant. Among environments; S4(BWN-14), S9(Multan-14) and S10(Vehari-14) were higher yielding environments, whereas S13(Piplan-14) and S8(Khanewal-14) were the most favorable environments. Among 26 environments L6(Gujranwala-13) was low yielding and was the most unfavorable environment.

Figure 4.2: *Biplot of interaction principal component analysis (PCA) axis-1 versus mean yield of wheat data for 30 genotypes in 26 environments*

4.1.2.2 AMMI Stability Value (ASV)

ASV is the distance from zero in a two dimensional scatter-gram of IPCA1 scores against IPCA2 scores. As in GEI the contribution of PC-1 is much more than that of PC-2, so more weight was give according to the relative contribution of PC's (Farshadfar, 2008). The distance from zero was obtain by applying Pythagoras theory (Purchase, 1997).

Table 4.3: Mean yield genotypes, scores for the IPC 1 and IPC 2, AMMI stability (value ASV), rank of genotypes based on ASV and genotype selection index for 30 genotypes

Genotype	Mean yield (Y_i)	Rank (Y_i)	ASV	Rank of ASV	GSI
V-12266	3307.80	6	13.32	15	21
V-11047	2850.33	27	10.11	9	36
V-12284	3230.41	12	7.41	8	20
V-11098	3493.20	2	0.22	1	3
V-11061	3218.96	13	14.68	16	29
V-11022	3261.07	10	18.67	20	30
V-11092	3032.99	23	21.00	21	44
NW.10.1111-7	3059.22	22	0.41	2	24
TW11510	3023.16	25	6.03	6	31
V-12275	3148.69	16	11.32	11	27
GH	2853.74	26	3.98	3	29
V-11137	3157.16	15	18.65	19	34
NN-GAN-3	3232.26	11	15.69	17	28
V-11138	3310.37	5	5.89	5	10
V-11046	3264.55	8	4.61	4	22
NR-411	3274.53	7	23.61	23	30
V-11365	3084.32	21	13.28	14	35
V-12265	3135.73	17	10.62	10	27
Millat-11	3121.67	18	12.90	13	31
Lasani-08	3376.89	4	25.43	25	29
Nr-409	3262.65	9	26.66	26	35
V-11143	3401.76	3	29.60	29	32
V-11041	3181.39	14	18.28	18	32
V-11032	3031.73	24	6.77	7	31
NS-10	2839.56	28	40.95	30	58
11BT004	3115.61	20	27.57	27	47
9459-1	2809.60	29	12.54	12	41
V-12304	3591.69	1	12.54	22	23
11B2074	2720.89	30	24.91	24	24
11B2049	3116.34	19	27.58	28	47

AMMI stability value, mean and genotype selection index are given in (Table 4.3). Stability analysis showed that mean yield have range from 2720.9 to 3591.69 kg/plot and ASV ranged from 0.22-40.95 for wheat genotypes. In ASV as V-11098 followed by , NW.10.1111-7 and "GH" gave the lowest ASV values indicated these were the most stable genotypes, while NS-10 was recognized the most unstable as it gave the highest ASV.

4.1.2.3 Genotype selection index (GSI)

It is not necessary that stable genotype also give the highest yield so stability per *Se* should not be only selection parameter. As ASV take into account both stability and yield criteria , so according to ASV given in (Table 4.3) V-11098 was chosen as the most desirable genotype for selection based on stability high yielding performance; whereas, NS-10 having highest GSI was least desirable.

4.1.3 GGE-biplot

Visual results using GGE biplot of 30 wheat genotypes in 26 different environments are plotted in Figures 4.3 to 4.8. GGE analysis Revealed that, IPCA 1 explained 33.91% of $G \times E$ variation, where as $G+GE$ variation was explained by IPCA2 was 16.02%. Over all GGE biplot explained 49.93% by both axis.

In [Fig.4.3] GGE-biplot based on genotype focused scaling is represented for the identification where the genotypes are located. The genotypes V-12304, V-11143, Nr-409, NR-411, Lasani, V-11022 , V-11061 , V-11098, gen-14, V-12266 were above average, while other genotypes were below average. These above average genotypes can be divided into two groups; V-12266, V-11138, V-11046, and V-11098 were declared stable because they have low PC2 score and others were considered unstable genotypes (having high PC2 score).

Figure 4.3: *GGE-biplot based on genotype-focused scaling for genotypes. Details of genotypes are give in Table 3.1*

GGE Biplot based on environment-focused scaling was plotted to explore relationship among environments, it was shown that all environment exhibited similar direction with respect to PC 1. The angles between environments shown that, they can be divided in four mega environments. S10(Vehari-13) and L10(Vehari-13) formed a group [Fig.4.4], while S12(Sargodha-14), L12(Sargodha-13), S13(Piplan-14) and L13(Piplan13) formed other one. remaining genotypes further were grouped in two mega-environments. Standard deviation related within environments can be approximated by the length of vector. Among 26 environments S13(Piplan-14) and S10(Vehari-13) were found most discriminant environments, where as, L12(Sargodha-13) and L6(Gujranwala-13) were least discriminant among all environments.

Figure 4.4: *GGE-biplot based on environment-focused scaling for environments. Details of environments are give in Table 3.2*

To assess where the genotypes is best suited, genotypes were plotted against environments for the visualization of which-won-where pattern [Fig.4.5]. In this biplot, a polygon was formed by joining the six vertex of genotypes V-12304, NS-10, 11B2074, V-11047, 11B2049 and V-11143; which were the most responsive genotypes. Corners also gave the idea that Genotype V-11143 and V-12304 were low yielding. From the polygon view were close to ideal genotypes suggested by [Yan and Tinker, 2005](#).

Figure 4.5: *Polygon views of GGE-biplot based on symmetrical scaling for the which-won-where pattern for genotypes and environments*

A GGE biplot viewing the Average environment coordination (AEC) based on genotype focused scaling is portrayed in [Fig.4.6]. The average-environment axis (AEA) points to higher average seed yield; while average-environment coordinate (AEC) which is perpendicular to AEA points to greater variability in either direction. Results revealed that 11B2074 have the highest yield where as V-12304 gave the lowest. Genotypes; V-11143, NS-10, 11BT004 and 11B2049 having large AEC projection indicated that these were least stable and genotypes V-11098, V-11047, NW.10.1111-7, V-11046, GH, V-11365 and 9459-1 were proved to be most stable. Hence on basis of both mean yield and stability genotypes 9459-1, NW.10.1111-7, GH, V-11047 were recommended

as the most favorable genotypes.

Figure 4.6: *Average environment coordination (AEC) views of the GGE-biplot based on gen-focused scaling for mean performance and the stability of genotypes*

GGE biplot also facilitate in assessment of ideal genotype showing concentric circles, as (Farshadfar, 2008) suggested ideal genotype (high mean yield and the most stable one) should be located in the center. No genotype was located in center [Fig.4.7], so V-11098 was considered as the most desirable genotype across all environments as it lie on first concentric circle as The closer a genotype to the ideal one is the more valuable it is. Although such an "ideal" genotype may not exist in reality, it can be used as a reference for genotype evaluation (Yan and Tinker, 2005). Genotypes 11B2074 and NS-10 although gave the high yield located at far away from center so these were recommended the most undesirable genotypes among all genotypes.

Figure 4.7: *GGE-biplot based on genotype-focused scaling for comparison of the genotypes with ideal genotype*

The ability of environments to discriminate for genotypes are visualized using GGE-biplot presented in concentric circles based on environment focused scaling in [Fig.4.8] . Among environments none of it located at center so not a single environment can be declared ideal. Environment S9(Multan-14) is located at second concentric circle so it may be regarded as favorable environment. Environments L12(Sargodha), S10(Vehari-14) and S13(Piplan) were distinctly different from other environments. As these environments were far apart located from center and exhibit low yield potential and are declared as unfavorable environments.

Figure 4.8: *GGE-biplot based on environment-focused scaling for comparison of the environments with ideal environment*

4.1.4 Cluster Analysis

For assessment of group an similarities among different genotypes, Hierarchical cluster analysis using complete linkage for 30 genotypes was performed using Eu-

clidean distance given in [Fig.4.9]. Analysis explored that genotypes make small grouping then merge into 3 mega-groups, genotypes V-11098 and Lasani-08 formed a group indicating that these have similarity between them. Similar trends can be observed in other groups

Figure 4.9: Cluster of 30 genotypes for wheat crop data

Similarities among environment are assessed by plotting cluster for environments using complete linkage method. From [Fig.4.10] given below it can be observed that environment are divided into four major groups. Environment S13(Piplan-14) form group with S8(Khanewal-14), S9(Multan-14) and S10(Vehari-14). It is noticeable that S6(Gujranwala-14) form group with all environments of year 2013.

Figure 4.10: Cluster of 13 environments for wheat crop data

4.2 Bayesian analysis

A multi-environmental data of two consecutive years was used for analysis which consist of 30 genotypes of wheat crop evaluated in 13 different environments. The layout of experiment was RCBD with three replication for all genotypes as well as for years. To elicit prior, first year data was used and Von-MISES Fisher distribution was applied as proper prior suggested by (PEREZ-ELIZALDE *et al.*, 2011).


After obtaining large MCMC sample using joint posterior distribution, parameters were estimated using sample mean. Histogram and traces of μ and first two singular values (λ_1, λ_2) are presented in [Fig.4.11]. These trace and histogram exhibit that all were slightly positively bell shaped and have acceptable shape with marginal posterior distribution. The Bayesian AMMI estimate for λ_1 was (2077.26) and for λ_2 was 1292.73 and standard deviation for λ_1 and λ_2 were 423.2 and 381.9, respectively. The posterior densities of remaining singular values show a tendency to move toward zero, complete posterior densities for singular values are given in (Table 4.6).

The histogram of MCMC sample along with marginal posterior singular values

$(\lambda_{11}-\lambda_{12})$ are shown in [Fig.4.12]. The figure shows a similar pattern first 2,3 posterior singular values have slightly positive bell shape, where as later values tend to have a direction towards zero. Cumulative proportion of variation for posterior densities by eigenvalues explored that about 90% interaction variance can be explained by four components [Fig.4.12]. It can be evaluated that first component explain about 60% interaction variation, and second component approximately 17%. It shows that two bilinear term can retained in the model as first two singular values can explain 80% cumulative variation. These results of two significant bilinear components are an evidence with those normally present in the analyses of genotype by environment trials where the requirement is to have more than one bilinear term in the model to handle complexity of the $G \times E$ (PEREZ-ELIZALDE *et al.*, 2011).

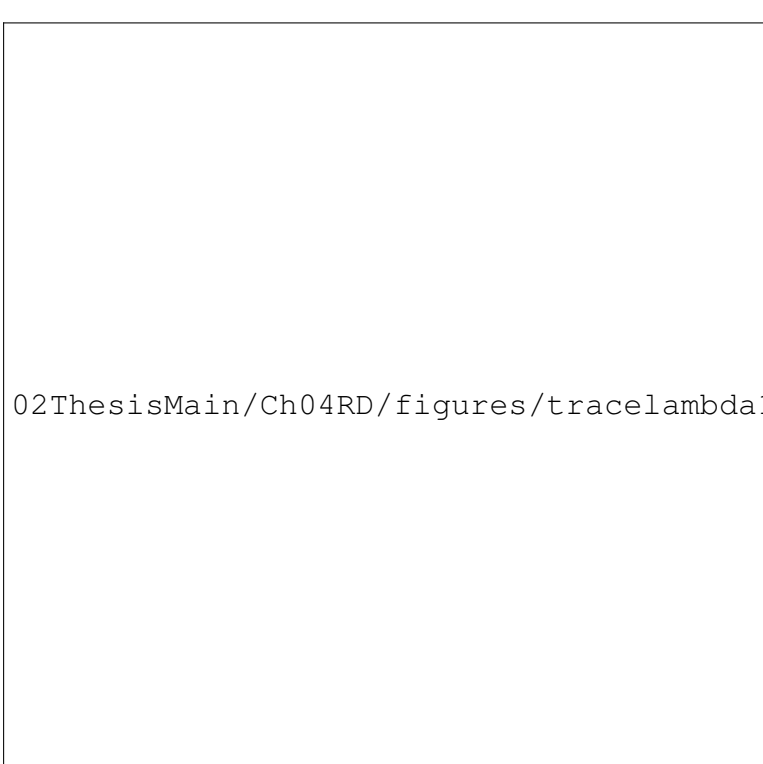
4.2.1 Uncertainty and confidence regions of first two bilinear terms in Bayesian AMMI

In general, the values of first bilinear terms u_{i1} and v_{j1} were, either positive or negative, larger than the values than those of other bilinear terms such as of u_{i2} and v_{j2} for mean , whereas the standard deviations (SD) exhibit different pattern as SD, for u_{i1} and v_{j1} were lower than those of u_{i2} and v_{j2} . Further more, HPD regions have narrow length for the first bilinear components of genotypes and environments (u_{i1} and v_{j1}) as compared to the second bilinear components, u_{i2} and v_{j2} [see appendix] . These results are indication that discrimination of genotypes and environment is done more precisely by first bilinear component than those of second bi-leaner component. In general, this is similarity with the classical approach named Principal component analysis (PCA) where the variability explained by first component is always more than second principal component. The summary is, there was less uncertainty in first bilinear component than second bilinear component.



02ThesisMain/Ch04RD/figures/tracemu.pdf

***((a))** fig 1*



02ThesisMain/Ch04RD/figures/tracelambda1.pdf

***((b))** fig 2*



((a)) fig 1



((b)) fig 2

Figure 4.12: Wheat Crop data of 30 genotypes in 13 environments:(a) Posterior densities and 0.95 HPD regions of the singular values, $\lambda_1, \dots, \lambda_{11}$ (C.1-C.12); (b) posterior densities and 0.95 HPD regions of the cumulative proportion of variance $\phi_t = \frac{\sum_{k=1}^t \lambda_k^2}{\sum_{k=1}^{\min(r,c-1)} \lambda_k^2}$, $t = 1, \dots, \min(r, c)-2$.

4.2.2 *Credible intervals for genotypes*

The posterior, mean, SD and HPD intervals for probability 0.90 and 0.95 are provided in (Table 4.4). The posterior Sd of main effects α_i have range from 100.34 to 103.26. In general, the values of ui1 and vj1 were, in absolute terms, larger than the values of ui2 and vj2, whereas the standard deviations (SD) of ui1 and vj1 were smaller than those of ui2 and vj2. Consequently, the lengths of the HPD regions were narrower for the first bilinear components of genotypes and environments (ui1 and vj1) than for the second bilinear components, ui2 and vj2 (data not shown). In summary, there is more posterior uncertainty in the second bilinear components for genotypes and sites than in the first bilinear components. Further illustration is provided in (Table 4.4).

4.2.3 *Credible intervals for environments*

The posterior, mean, SD and HPD intervals for probability .90 and .95 for environment main effect are provided in (Table 4.6). The posterior SD of main effects β_j have range from 64.80 to 66.36.

4.2.4 *Credible Regions of the First Two Bilinear Terms of the Linear Bilinear Model*

The bilinear terms which do not include the null point (0, 0) for the bivariate 0.90 and .95 HPD region are referred as significant terms, it mean these terms significantly contribute in explanation of interaction variation (PEREZ-ELIZALDE *et al.*, 2011).

In (Table 4.6) the posterior mean of 12 singular values and their .95 and .90 Highest posterior density (HPD) are give. It can be seen that the first two eigenvalues have high value where as remaining values tends to moves towards zero. Along with singular values, eigenvectors values for those genotypes and environments are give which do not have null points for .90 HPD and .95 HPD. The score of environment

Table 4.4: Posterior summary (Mean), standard deviation (SD), quartiles ($q_{0.25}$, $q_{0.50}$, and $q_{0.75}$), 0.90 HPD, and 0.95 HPD intervals computed with 20,000 approximately independent samples simulated from the posterior distribution of 30 linear effects of genotype for wheat crop data

Para	Mean	SD	$q_{0.25}$	$q_{0.50}$	$q_{0.75}$.90 HPD interval		.95 HPD interval	
						lower	Upper	lower	upper
μ	3149.47	19.11	3136.41	3149.33	3162.22	3117.90	3180.29	3112.91	3187.78
α_1	144.47	102.01	76.88	144.77	213.02	-24.99	308.58	-48.61	350.92
α_2	-299.67	102.29	-368.71	-300.44	-230.39	-468.43	-132.72	-491.40	-93.06
α_3	80.48	100.34	12.61	80.22	147.79	-87.77	241.36	-113.71	280.58
α_4	343.76	101.80	274.76	343.74	411.53	184.22	519.40	145.95	543.93
α_5	68.06	102.55	-0.988	68.10	138.96	-93.04	240.88	133.03	264.83
α_6	111.16	101.75	42.36	111.66	179.85	-55.01	278.38	-91.35	305.81
α_7	-116.49	102.68	-185.74	-117.63	-47.77	-278.46	61.121	-311.26	91.21
α_8	-89.62	102.16	-159.11	-90.23	-21.23	-254.75	79.31	-285.64	113.89
α_9	-127.20	101.35	-196.58	-127.14	-57.99	-292.96	37.66	-332.64	63.68
α_{10}	-2.16	102.30	-69.51	-2.22	65.91	-169.02	164.47	-211.16	190.67
α_{11}	-296.26	101.48	-365.17	-296.33	-228.84	-468.71	-136.58	-490.76	-97.13
α_{12}	6.65	101.60	-61.12	7.11	75.31	-163.21	171.12	-195.24	204.97
α_{13}	82.29	102.6	13.19	81.11	151.35	-87.60	250.91	-117.20	282.40
α_{14}	161.84	103.42	93.04	162.78	230.93	-3.26	336.82	-40.87	360.99
α_{15}	114.64	101.61	46.99	115.30	182.10	-52.28	285.2	-87.42	311.87
α_{16}	125.49	101.25	58.68	124.24	193.01	-32.91	300.62	-65.19	330.20
α_{17}	-64.89	101.89	-134.56	-65.52	4.5839	-224.59	108.64	-262.62	132.10
α_{18}	-12.88	103.60	-83.14	-12.94	56.77	-181.72	154.50	-216.43	190.61
α_{19}	-29.45	102.42	-99.18	-29.20	40.05	-200.18	134.87	-244.40	156.10
α_{20}	227.33	100.42	160.09	227.98	295.82	69.60	396.76	31.49	424.60
α_{21}	112.95	102.55	43.99	112.25	181.99	50.03	286.92	-86.49	312.11
α_{22}	252.68	102.25	182.49	253.66	322.34	93.08	426.27	49.43	447.91
α_{23}	32.70	102.92	-36.10	31.65	102.47	-136.018	199.39	-167.03	233.40
α_{24}	-118.83	102.53	-187.85	-119.65	-49.46	-285.57	52.50	-318.81	81.03
α_{25}	-311.46	102.30	-380.53	-310.35	-243.37	-482.56	-145.76	-507.53	-109.09
α_{26}	-35.89	102.61	-105.44	-36.89	33.58	-197.72	138.63	-245.10	156.20
α_{27}	-340.15	101.36	-408.19	-339.38	-270.92	-502.16	-168.85	-536.61	-141.80
α_{28}	441.91	102.22	374.73	442.90	511.35	276.88	612.36	243.23	648.14
α_{29}	-429.16	103.26	-498.31	-428.42	-359.38	-602.33	-262.07	-628.11	-222.16
α_{30}	-32.3	101.5	-100.46	-31.99	36.55	-201.45	131.06	-227.45	171.49

S13 ($v_{13,1}$) and genotype 25 ($u_{25,1}$) score do not have null point for both HPD intervals, indicates that they have significant contribution in $G \times E$.

Table 4.5: Posterior summary (Mean), standard deviation (SD), quartiles ($q_{0.25}$, $q_{0.50}$, and $q_{0.75}$), 0.90 HPD, and 0.95 HPD intervals computed with 20,000 approximately independent samples simulated from the posterior distribution of 13 linear effect of environments wheat crop data

Para	Mean	SD	$q_{0.25}$	$q_{0.50}$	$q_{0.75}$.90 HPD Interval		.95 HPD interval	
						lower	Upper	lower	upper
β_1	-62.10	66.36	-106.74	-62.45	-17.49	-171.05	45.249	-188.46	69.61
β_2	-440.83	65.57	-485.00	-440.82	-395.33	-548.69	-335.02	-567.72	-312.09
β_3	-137.94	65.52	-182.41	-138.00	-93.69	-242.82	-28.64	-275.31	-18.64
β_4	28.50	65.90	-16.00	28.62	73.100	-77.38	137.87	-101.63	155.95
β_5	-429.95	65.85	-474.58	-430.60	-384.73	-533.56	-318.72	-558.36	-303.55
β_6	-980.62	65.59	-1024.24	-981.34	-936.84	-1084.69	-868.05	-1104.70	-845.19
β_7	-250.62	66.00	-294.87	-251.14	-205.93	-356.828	-140.71	-383.77	-126.10
β_8	1028.13	65.89	983.95	1028.66	1071.78	918.98	1135.92	899.01	1154.80
β_9	745.15	65.72	700.67	745.42	789.45	634.47	849.65	616.27	871.52
β_{10}	600.14	66.03	556.00	600.40	644.77	494.81	711.38	465.49	725.05
β_{11}	-131.97	65.28	-176.06	-131.55	-88.76	-236.91	-22.04	-261.44	-4.74
β_{12}	-539.01	65.08	-581.73	-539.37	-496.44	-650.78	-434.14	-662.55	-406.02
β_{13}	571.13	64.80	527.11	570.95	614.77	464.81	677.83	444.81	696.26

Table 4.6: Posterior summary (Mean), standard deviation (SD), quartiles ($q_{0.25}$, $q_{0.50}$, and $q_{0.75}$), 0.90 HPD, and 0.95 HPD intervals computed with 20,000 approximately independent samples simulated from the posterior distribution of the residual variance (σ), all singular values ($\lambda_1, \dots, \lambda_{12}$) and the right and the left singular vector elements of genotypes and environments, respectively, whose 0.90 HPD and 0.95 HPD intervals do not contain the null value (0, 0). Data for wheat yield

Para	Mean	SD	$q_{0.25}$	$q_{0.50}$	$q_{0.75}$.90 HPD Interval		.95 HPD interval	
						lower	Upper	lower	upper
λ_1	2077.26	423.20	790.88	2084.27	2369.18	1369.07	2748.10	1240.48	2888.35
λ_2	1292.73	381.89	1036.29	1303.84	1553.56	675.14	1920.03	533.34	2018.84
λ_3	903.23	347.96	664.77	909.93	1148.50	316.96	1470.83	216.29	1579.14
λ_4	597.50	306.67	371.38	588.14	814.40	24.49	1022.25	0.1407	1110.15
λ_5	379.46	254.98	173.69	348.85	556.13	0.0037	734.44	0.0037	836.66
λ_6	236.08	199.88	66.70	191.74	359.06	0.0015	528.47	0.0015	621.69
λ_7	140.88	150.84	22.27	87.94	213.78	0.0001	365.08	0.0001	458.43
λ_8	79.15	106.10	6.91	34.94	109.76	0.00	225.29	0.00	314.97
λ_9	43.05	70.21	2.27	13.20	51.13	0.00	128.98	0.00	190.38
λ_{10}	22.06	42.92	0.71	4.81	22.37	0.00	65.143	0.00	108.12
λ_{11}	11.51	26.51	0.22	1.68	9.66	0.00	31.60	0.00	58.90
λ_{12}	5.88	15.83	0.073	0.62	3.94	0.00	15.14	0.00	30.46
u_{25_1}	0.23	0.141	0.14	0.244	0.33	0.015	0.47	-0.041	0.501
$v_{13,1}$	-0.63	0.11	-0.71	-0.64	-0.55	-0.81	-0.45	-0.83	-0.39

Biplot for the bivariate 0.90 HPD and 0.95 HPD regions of column score is shown in Figure [Fig.4.13]. Visual analysis shown that S10 form a group of distinct group of environment which have positive value, while S13 showed a negative correlation with S10. some degree of similarity was also shown between S13 and S9, S10 overlap with S12, S8 and S4, it means they form a homogeneous group. whereas, S2, S3, S1, S5, S7 and S11 form a group which were not as significant as others.

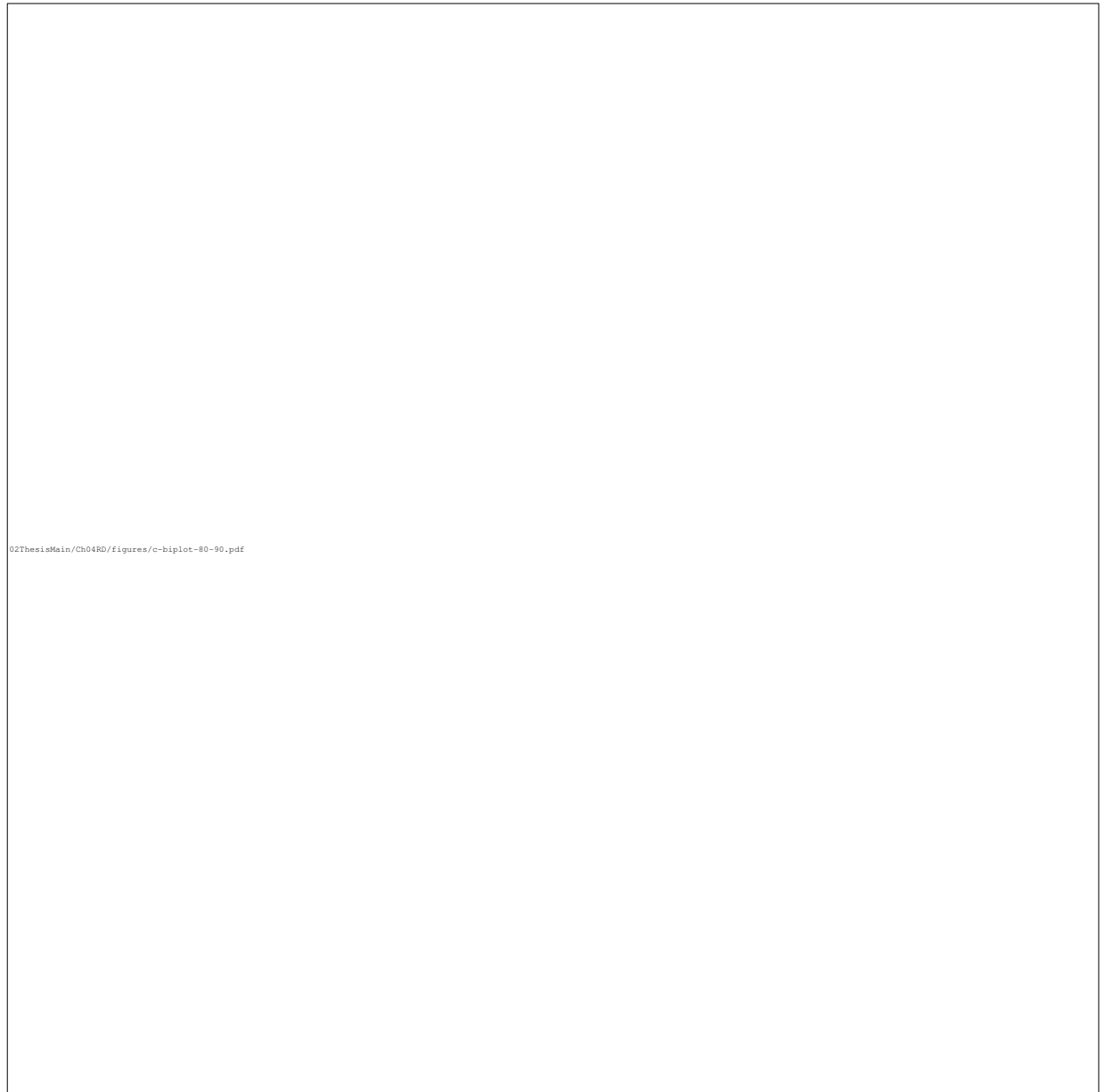


Figure 4.13: Wheat crop data of 30 genotypes in 13 environments: plot of the bivariate column scores $\mathbf{V}'\mathbf{D}^{1/2}$ and the bivariate 0.95 (gray external contour) and 0.90 (gray internal contour) HPD regions

In [Fig.4.14] bi-plot for row score is shown, among all genotypes 25($u_{25,1}$) was

far away showing positive value, followed by genotypes 20, 29 and 26 overlapping each other show some similarities among them. These genotypes form a distinct group. Genotype 25 is only which do not contain null point (0,0) and significantly contributed to the interaction. The remaining genotypes form an other group in first bilinear components are not different from zero.



Figure 4.14: Wheat crop data of 30 genotypes in 13 environments: plot of the bivariate row scores $\mathbf{U}'\mathbf{D}^{1/2}$ and the bivariate 0.95 (gray external contour) and 0.90 (gray internal contour) HPD regions

As biplots are based on 0.90 and 0.95 HPD interval have shown overlapping so

some significant environment may not be clearly portrayed. A hierarchical clustering based on posterior mean of Euclidean distance between the rows having interaction with complete linkage method was performed [Fig.4.15]. It can be seen that S10 form a single group then merge into other groups, similarly S9 and S13 form a cluster is an evidence that there is some similarity between them.

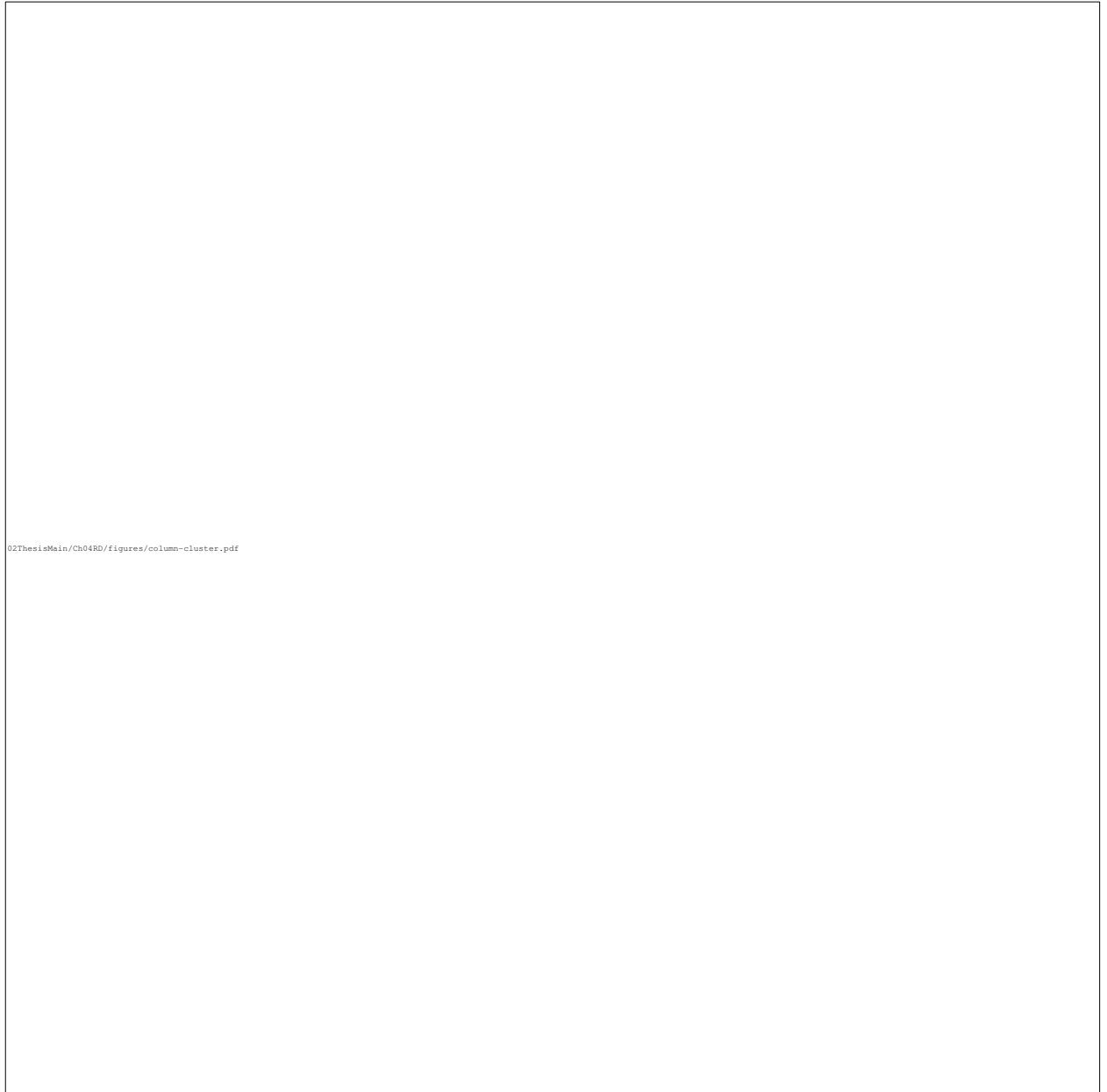


Figure 4.15: *Dendrogram of the 13 environments using the first two right singular vectors*

A clusters analysis using posterior mean for genotypes using first two left singular vectors was also performed shown in [Fig.4.16]. Genotypes merge into 4 major

clusters, gen-20, gen-29 and gen-25 form a cluster showing these have some similar characteristics among them. This technique for the identification of homogeneous genotypes and environment subsets is analogous to that provided by (Burguenoa *et al.*, 2008)

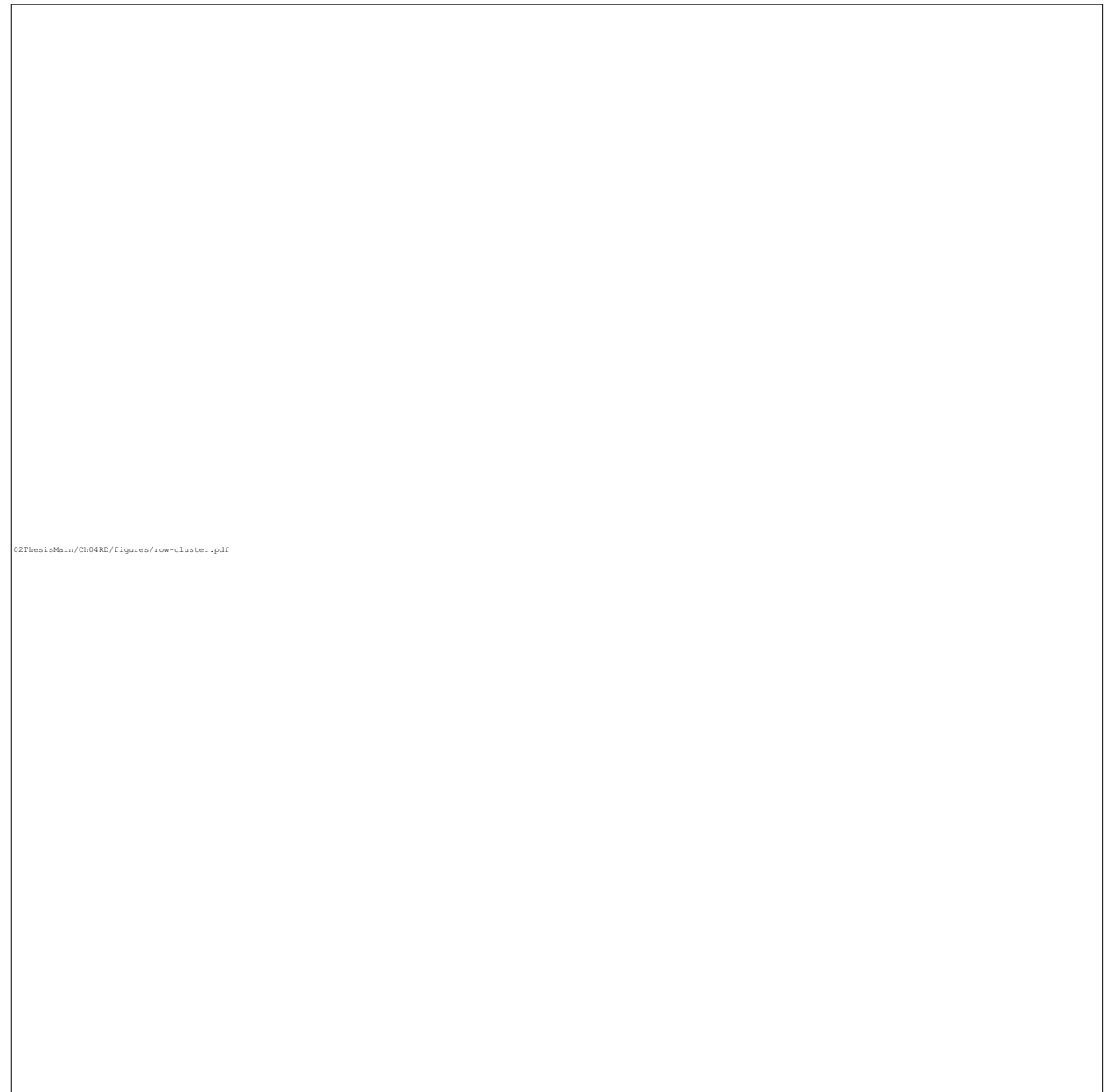


Figure 4.16: Dendrogram of the 30 genotypes using the first two left singular vectors

The graphical comparison between the biplots obtained using ordinary least square method (OLS) and Bayesian is given in [Fig.4.17]. Both methods provide almost similar results, it is also an indication that Bayesian can be adapted for the analysis of trials which have $G \times E$ interaction.

02ThesisMain/Ch04RD/figures/comparison.pdf

Figure 4.17: *Comparison between the biplots of Classical AMMI using ordinary least square method and Bayesian approach for 30 genoypes and 13 environments of wheat crop data*

OLS biplot have more spread-ed observation as compared to Bayesian biplot. Both revealed that genotype 25 and environment 13 (Pilpan) are highly significant and explain the variability of interaction.

Chapter 5: Summary

The objective of the study were (a): To examine the Genotype environment interaction using Bayesian approach; (b): Graphical represent of bi-plots using prior information; (c): To elaborate that Bayesian models can be easily adapted to GEI. For this purpose an experiment was conducted, which consist of 30 genotypes of wheat in thirteen different location of Punjab for two consecutive years. The experiment was arranged in RCBD layout with three replication.

Firstly combined analysis of variance was and it was observed that genotype and environment main effect as well as interaction was found significant. Then AMMI ANOVA suggested by [GOLLOB, 1968](#) was performed, analysis revealed that PC1 explain 23.4 and Pc2 explain 14.7 variation for interaction sum of square.

Bayesian approach using von-Mises Fisher distribution was applied to elicit prior for first year data. Credible Regions was obtained and bilinear terms which do not contain null points (0, 0) were refereed significant. Bilplots using first and second bilinear terms were drawn and observed that bilinear terms $u_{25,1}$ for genotype NS-10 genotype and $v_{13,1}$ for environment S13 (Piplan-14) have significant effect on interaction. From result it is concluded that Bayesian bilinear model can be applied to yield trial data because of following advantages.

- Confidence regions for genotype and environments for parameters and for their interaction can be obtain naturally.
- This approach provide facilitation to identify genotypes and environments that have impact on significant interaction; furthermore, genotypes and environments having similar responses can also be recognized.
- It can deal with unbalanced data (most of the time present in yield trials) naturally.
- Prior information related to genotypes, environments and for interaction can be incorporated by using Bayesian approach for means as well as variances.

- This approach can be adopted for cells having unequal size.

Although there is plenty of literature available related to Bayesian frame-work but a little work has been done on Bayesian inference for yield trials. It is suggested this methodology can be carried out other bilinear model by minimizing some restrictions and putting some parameters equal to zero. This approach will prove a good alternative of Classical method for Analyzing Genotype by environment interaction, which is often a major area of concern for yield experimenters.

Bibliography

- Adams, E., W. B. V. C., R. PG and M. DL. 2002. Principal Component Analysis of Dissolution Data with Missing Elements. *International Journal of Pharmaceutics* 234: 169–178.
- Akter, A., J. H. M., U. K. M., I. MR, K. Hossain and M. R. M. 2014. AMMI Biplot Analysis for Stability of Grain Yield in Hybrid Rice. *Rice Research* 2: 2.
- Allard, R. W. and A. D. Bradshaw. 1964. Implications of Genotype-Environmental Interactions in Applied Plant Breeding. *Crop Science* 4: 503–508.
- Annicchiarico, P. 1997. Joint regression vs AMMI analysis of genotype-environment interactions for cereals in Italy. *Euphytica* 94: 53–62.
- Bayes, T. 1763. An Essay Towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, 53: 370–418.
- Becker, H. C. and J. Leon. 1988. Stability Analysis in Plant Breeding. *Plant Breeding* 101: 1–23.
- Bradu, D. and K. R. Gabriel. 1978. The Biplot as a Diagnostic Tool for Models Of Two-Way Tables. *Technometrics* 20: 47–68.
- Burguenoa, J., J. Crossa, P. L. Corneliusb and R.-C. Yang. 2008. Using Factor Analytic Models for Joining Environments and Genotypes without Crossover Genotype Environment Interaction. *Crop Science* 4: 1291–1305.
- Campbell, B. T. and M. A. Jones. 2005. Assessment of genotype environment interactions for yield and fiber quality in cotton performance trials. *Euphytica* 144: 69–78.
- Castillo, D., I. Matus, A. del Pozo, R. Madariaga and M. Mellado. 2012. ADAPTABILITY AND GENOTYPE \times ENVIRONMENT INTERACTION OF SPRING WHEAT CULTIVARS IN CHILE USING REGRESSION ANALYSIS, AMMI, AND SREG. *CHILEAN JOURNAL OF AGRICULTURAL RESEARCH* 72: 167–174.
- Crossa, J., H. G. Gauch and R. W. Zobel. 1990. Additive Main Effects and Multiplicative Interaction Analysis of Two International Maize Cultivar Trials. *Crop S* 30: 493–500.
- Crossa, J., S. Perez-Elizalde, D. Jarquin, J. M. Cotes, K. Viele, G. Liu and P. L. Cornelius. 2011. Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction Model. *CROP SCIENCE* 51: 1458–1469.

- Crossa, J., R.-C. Yang and P. L. Cornelius. 2004. Studying Crossover Genotype \times Environment Interaction Using Linear-Bilinear Models and Mixed Models. *Journal of Agricultural, Biological, and Environmental Statistics* 9: 362–380.
- Denis, J.-B. and J. C. Gower. 1996. Asymptotic Confidence Regions for Biadditive Models: Interpreting Genotype- Environment Interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 45: 479–493.
- Dias, C. T. dos and W. J. Krzanowski. 2003. Model Selection and Cross Validation in Additive Main Effect and Multiplicative Interaction Models. *CROP SCIENCE* 43: 865–873.
- Ebdon, J. S. and H. G. Gauch. 2002. Additive Main Effect and Multiplicative Interaction Analysis of National Turfgrass Performance Trials. *Crop Science* 42: 497–506.
- Farshadfar, E. 2008. Incorporation of AMMI stability value and grain yield in a single non-parametric index (GSI) in bread wheat. *Pak J Biol Science* 11: 1791–1796.
- Forkman, J. and H.-P. PIEPHO. 2013. Performance of empirical BLUP and Bayesian prediction in small randomized complete block experiments. *The Journal of Agricultural Science* 151: 381–395.
- Gabriel, K. R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58: 453–467.
- Gabriel, K. R. and S. Zamir. 1979. Lower Rank Approximation of Matrices by Least Squares with Any Choice of Weights. *Technometrics* 21: 489–498.
- Gauch, H. G. 2006. Statistical Analysis of Yield Trials by AMMI and GGE. *Crop Science* 46: 1488–1500.
- Gauch, H. G. and R. W. Zobel. 1988. Predictive and postdictive success of statistical analyses of yield trials. *THEORETICAL AND APPLIED GENETICS* 76: 1–10.
- Gauch, H. G. and R. W. Zobel. 1990. Imputing Missing Yield Trial Data. *Theoretical and Applied Genetics* 79: 753–761.
- Gauch, H. G. and R. W. Zobel. 1997. Identifying Mega-Environments and Targeting Genotypes. *Crop Science* 37: 311–326.
- Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin. 2004. *Bayesian Data Analysis*. CHAPMAN & HALL/CRC.
- Gelman, A. and D. B. Rubin. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 4: 457–472.
- Glickman, M. E. and D. A. van Dyk. 2007. Basic Bayesian Methods. *Topics in Biostatistics* 2007: 319–338.
- GOLLOB, H. F. 1968. A Statistical Model Which Combines Features of Factor Analytic and Analysis of Variance. *Psychometrika* 33: 73–115.

- Hegemann, V. and D. E. Johnson. 1976. On Analyzing Two-Way AoV Data with Interaction. *Technometrics* 18: 273–281.
- Herz, C. S. 1955. Bessel Functions of Matrix Argument. *Annals of Mathematics* 61: 474–523.
- Hoff, P. D. 2007. Model Averaging and Dimension Selection for the Singular Value Decomposition. *Journal of the American Statistical Association* 102: 675–684.
- James, A. T. 1964. Distributions of Matrix Variates and Latent Roots Derived From Normal Samples. *Annals of Mathematical Statistics* 35: 475–501.
- Jarquín, D., J. Crossa, S. PEREZ-ELIZALDE, J. M. Cotes, K. Viele, G. Liu and P. L. Cornelius. 2011. Bayesian estimation of the additive main effects and multiplicative interaction model. *Crop Science* 51: 1458–1469.
- Johansson, E., M. L. Prieto-Linde, G. Svensson and J. O. Jonsson. 2003. Influences of cultivar, cultivation year and fertilizer rate on amount of protein groups and amount and size distribution of mono and polymeric proteins in wheat.. *The Journal of Agricultural Science* 140: 275–284.
- JOSSE, J., F. van EEUWIJK, H.-P. PIEPHO and J.-B. DENIS. 2014. Another Look at Bayesian Analysis of AMMI Models for Genotype-Environment Data. *Journal of Agricultural, Biological, and Environmental Statistics* 19: 240–257.
- Josse, J., J. Pages and F. Husson. 2011. Multiple Imputation in PCA. *Advances in Data Analysis and Classification* 5: 231–246.
- KAYA, Y., C. PALTA and S. TANER. 2002. Additive Main Effects and Multiplicative Interactions Analysis of Yield Performances in Bread Wheat Genotypes across Environments. *Turk J Agric For* 26: 275–279.
- Kempton, R. A. 1984. The use of biplots in interpreting variety by environment interactions. *The Journal of Agricultural Science* 103: 123–135.
- Khatri, C. G. and K. V. Mardia. 1977. The von MisesFisher Matrix Distribution in Orientation Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* 39: 95–106.
- Kiers, H. A. L. 1997. Weighted Least Squares Fitting Using Ordinary Least Squares Algorithms. *Psychometrika* 62: 251–266.
- Krzanowski, W. J. 1987. Some exact percentage points of a statistic useful in analysis of variance and principal component analysis. *Technometrics* 21: 261–263.
- Letta, T. 2007. Genotype-environment interactions and correlation among some stability parameters of yield in durum wheat (*Triticum durum* Desf) genotypes grown in South East Ethiopia. *African Crop Science Conference Proceeding* 8: 693–698.
- Magari, R. and M. S. Kang. 1993. Genotype selection via a new yield-stability statistic in maize yield trials. *Euphytica* 70: 105–111.

- Mandel, J. 1971. A New Analysis of Variance Model for Non-Additive Data. *Technometrics* 13: 1–8.
- Milliken, G. A. and D. E. Johnson. 1989. *Analysis of Messy Data, Volume II: Nonreplicated Experiments*. New York: Van Nostrand-Reinhold.
- Mohamed, N. E. M., A. A. Said and K. A. Amein. 2013. Additive main effects and multiplicative interaction (AMMI) and GGE-biplot analysis of genotype \times environment interactions for grain yield in bread wheat (*Triticum aestivum* L.). *African Journal of Agricultural Research* 8: 5197–5203.
- Negash, A. W., H. Mwambi, T. Zewotir and G. Taye. 2013. Additive main effects and multiplicative interactions model (AMMI) and genotype main effect and genotype by environment interaction (GGE) biplot analysis of multi-environmental wheat variety trials. *African Journal of Agricultural Research* 8: 1033–1040.
- Nounou, M. N., B. R. Bakshi, P. K. Goel and X. Shen. 2002. Bayesian principal component analysis. *Journal of chemometrics* 16: 576–595.
- PEREZ-ELIZALDE, S., D. JARQUIN and J. CROSSA. 2011. A General Bayesian Estimation Method of Linear Bilinear Models Applied to Plant Breeding Trials With Genotype \times Environment Interaction. *Journal of Agricultural, Biological, and Environmental Statistics* 17: 15–37.
- Peterson, C. J., R. A. Graybosch, P. S. Baenziger and A. W. Grombacher. 1992. Genotype and Environment Effects on Quality Characteristics of Hard Red Winter Wheat. *Crop Science* 32: 98–103.
- Purchase, J. L. 1997. Parametric analysis to describe genotype \times environment interaction and yield stability in winter wheat. PhD thesis. Thesis, Department of Agronomy, Faculty of Agriculture of the University of the Free State, Bloemfontein, South Africa.
- R, C. T. 2015. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria.
- Rad, M. N., M. A. Kadir, M. Rafii, H. Z. Jaafar, M. R. Naghavi and F. Ahmadi. 2013. Genotype \times environment interaction by AMMI and GGE biplot analysis in three consecutive generations of wheat under normal and drought stress conditions. *Australasian Journal of Crop Science* 7: 956–961.
- Raftery, A. E. and S. M. Lewis. 1995. The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms. In: *Practical Markov Chain Monte Carlo* (W.R. Gilks, D.J. Spiegelhalter. Chapman and Hall, pp. 115–130.
- Rasul, S., M. I. Khan and M. A. Khan. 2007. COMPARISON OF SAFETY- FIRST RULE AND CLUSTER ANALYSIS FOR EVALUATING WHEAT GENOTYPES WITH SIGNIFICANT GENOTYPE \times ENVIRONMENT INTERACTION. *Pak. J. Agri. Sci.* 44: 4.
- Rodrigues, P. C., A. Monteiro and V. M. Lourenco. 2015. A Robust AMMI model for the analysis of genotype-by-environment data. *Bioinformatics*.

- Sadeghi, S. M., H. Samizadeh, E. Amiri and M. Ashouri. 2011. Additive main effects and multiplicative interactions (AMMI) analysis of dry leaf yield in tobacco hybrids across environments. *African Journal of Biotechnology* 10: 4358–4364.
- Schulthessa, A., I. Matusb and A. Schwembera. 2013. Genotypic and environmental factors and their interactions determine semolina color of elite genotypes of durum wheat (*Triticum turgidum* L. var. durum) grown in different environments of Chile. *Field Crops Research* 149: 234–244.
- Shafii, B. and W. J. Price. 1998. Analysis of Genotype-by-Environment Interaction Using the Additive Main Effects and Multiplicative Interaction Model and Stability Estimates. *Journal of Agricultural, Biological, and Environmental Statistics* 3: 335–345.
- Silva, C. P. da, L. A. de Oliveira, J. J. Nuvunga, A. K. A. Pamplona and M. Balestre. 2015. A Bayesian Shrinkage Approach for AMMI Models. *PLOS ONE* 10: 1–27.
- Simdl, V. and A. Quinnb. 2007. On Bayesian principal component analysis. *Computational Statistics & Data Analysis* 51: 4101–4123.
- Simmonds, N. W. 1962. VARIABILITY IN CROP PLANTS, ITS USE AND CONSERVATION. *Biological Reviews* 37: 422–465.
- Tolessa, T. T., G. Keneni, T. Sefera, M. Jarso and Y. Bekele. 2013. Genotype × Environment Interaction and Performer stability for Grain Yield in Field Pea Genotypes. *International Journal of plant breeding* 7: 116–123.
- Vargas, M., J. Crossa, F. A. van Eeuwijk, M. E. Ramirez and K. Sayre. 1999. Using Partial Least Squares Regression, Factorial Regression, and AMMI Models for Interpreting Genotype × Environment Interaction. *Crop Science* 39: 955–967.
- Viele, K. and C. Srinivasan. 2000. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback-Leibler Information. *Journal of Statistical Planning and Inference* 84: 201–219.
- Yan, W., L. Hunt, Q. Sheng and Z. Szlavnic. 2000. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *Crop* 40: 597–605.
- Yan, W. and N. A. Tinker. 2005. An Integrated Biplot Analysis System for Displaying, Interpreting, and Exploring Genotype × Environment Interaction. *Crop Science* 45: 1004–1016.
- Yang, R.-C., J. Crossa, P. L. Cornelius and J. Burgueo. 2009. Biplot Analysis of Genotype × Environment Interaction: Proceed with Caution. *Crop Science* 49: 1564–1576.
- Zali, H., E. Farshadfar, S. H. Sabaghpour and R. Karimizadeh. 2012. Evaluation of genotype × environment interaction in chickpea using measures of stability from AMMI model. *Annals of Biological Research* 3: 3126–3136.

Zobel, R. W., M. J. Wright and H. G. Gauch. 1988. Statistical Analysis of a Yield Trial. *Agronomy Journal* 80: 388–393.