ADDIS ABABA UNIVERSITY

GRADUATE STUDIES PROGRAM

COLLEGE OF NATURAL SCIENCE

DEPARTMENT OF STATISTICS

STATISTICAL ANALYSIS OF GENOTYPE BY ENVIRONMENT INTERACTIONS AND GRAIN YIELD STABILITY IN BREAD WHEAT USING ANOVA AND AMMI MODELS

By

Zelalem Tazu

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN STATISTICS.

June, 2011

Addis Ababa

Ethiopia

ADDIS ABABA UNIVERSITY
GRADUATE STUDIES PROGRAM
COLLEGE OF NATURAL SCIENCE
DEPARTMENT OF STATISTICS

STATISTICAL ANALYSIS OF GENOTYPE BY ENVIRONMENT INTERACTIONS AND GRAIN YIELD STABILITY IN BREAD WHEAT USING ANOVA AND AMMI MODEL

By
Zelalem Tazu

Approved by the board of examiners

……………………………………………… …………………………
Advisor Signature

………………………………… ……………. …………………………..
Examiner signature

……………………………………………….. ………………………….
Examiner Signature

Table of Contents

ACKNOWLEDGEMENT

Many thanks to the almighty God, My sincere and heartfelt gratitude to Dr. Girma Taye, thesis advisor and instructor, for giving important comments and suggestions that contributed to the successful realization of the study.

My appreciation and thanks are extended to the Department of Statistics for permission to use supporting materials that are necessary for the study. And finally but not least, I would like to thank my family and friends for their special and continuous cooperation and encouragement throughout my study period.

ABSTRACT

In agricultural experimentation involving Genotype by Environment interactions, a large number of genotypes are normally tested over a wide range of environments and the underlying statistical and genetical theories used to model this system may be rather complicated. The occurrence of the Genotype by Environment interaction further complicates the selection of superior genotypes for a target population of environments. In the absence of Genotype by Environment interaction, the superior genotype in one environment may be regarded as the superior genotype in all, where as the presence of the Genotype by Environment interaction confirms particular genotypes being superior in particular environments. Therefore, Genotype by Environment interaction plays an important role in identifying genotypes for high and stable yield. The goal of this study were to analyze Genotype by Environment interaction and stability of the Ethiopian wheat hybrids for grain yield across the target environments, and to observe the pattern of grouping of the genotypes and the environments based on grain yield response of the hybrids. This study was carried out on the yield performance of 20 bread wheat genotypes across 8 environments in Ethiopia for two growing seasons. The experimental layout was a randomized complete block design with four replications. The combined ANOVA (AMMI ANOVA) showed that environments, genotypes and Genotype by Environment interactions were highly significant ($p<0.01$) and they accounted for 80.91%, 3.37% and 4.6% of the total variation. The high percentage of the environment is an indication that the major factor that influence yield performance of bread wheat in Ethiopia is the environment. The best fit AMMI model for this multi-environment yield trial data was AMMI-4. Out of the total interactions of principal component analysis (IPCAs), the first four IPCA axes explained 82.63% of the Genotype by Environment interaction sum of squares. However, the biplot (the first two IPCAs) captures 62.32% of the interaction SS. The biplots showed G_6, G_4 and G_9 were more stable genotypes while G_{10}, G_{20}, G_{16}, G_{18} and G_3 were unstable varieties.

Key words: AMMI, Genotype by Environment interaction.

ACRONYMS

AMMI	Additive Main Effect and Multiplicative Interaction
ANOVA	Analysis of Variance
CIMMYT	International Maize and Wheat Improvement Center
CSA	Central Statistical Agency
CV	Coefficient of Variation
DF	Degrees of Freedom
EIAR	Ethiopian Institute of Agricultural Research
GEI (GXE)	Genotype by Environment Interaction
IRRI	International Rice Research Institute
IPCA	Interactions of Principal Component Analysis
LR	Linear Regression
MET	Multi-environment Trial
MS	Mean Square
NID	Normally and Independently Distributed
PCA	Principal Components Analysis
RCBD	Randomized Complete Block Design
SAS	Statistical Analysis System
SS	Sum of Squares
SV	Singular Value
SVD	Singular Value Decomposition

LIST OF TABLES

1 INTRODUCTION

1.1 BACKGROUND

Wheat constitutes one of the five major cereal crops in the world. It has been used as a main source of food since the prehistoric times. Many of the crops characteristics were probably well known 2000 years ago when it evidently was grown for food. Durum wheat (Triticum turgidum L.var.durum), sometimes called macaroni wheat, covers about 9% of the wheat area (FAO, 1994). Modern durum wheat varieties yield as the highest yielding bread wheat varieties and the kernels of durum wheat are typically larger, heavier, and harder than those of bread wheat. However, durum wheat dough is less elastic than that of bread wheat and, therefore, it is inferior for producing leavened loaves, but durum wheat is primarily used for making noodles and other pasta products such as macaroni and spaghetti. In the international trade, durum wheat of good quality generally commands a higher price than that of bread wheat.

Ethiopia is a country with a surface area of 1.2 million kilometers of which approximately 45% is arable. It is the largest wheat producer in sub-Saharan Africa with about 0.75 million ha of durum and bread wheat. Wheat is one of the major cereal crop grown in the Ethiopian highlands, which lie between 6 and 16 degree North and 35 and 42 degree East at altitudes ranging from 1500 to 3000. The most suitable areas for wheat production, however, fall between 1900 and 2700 meter. In the highlands, rainfall distribution is by modal and ranges between 600 and 2000 mm/annum. The rainy season divided in to the short rains (bulg) following from February to April and the main rains (meher) following from June to September.

Wheat of both the tetraploid (Triticum Durum Desf.) and hexaploid (Triticum aestivum L.), is the most important cereal crop in Ethiopia, ranking third in total production(17%) next to maize (zea mays L.) and teff (CSA, 2002). Wheat covers a total arable land of 110,434 ha with an average productivity of about 8.4 qt ha^{-1}, which is below the national average (14.4 qt ha^{-1}). Currently, about 60% of the wheat area covered by durum and 40% by bread wheat. Developing crop cultivars that perform well across a wide range of environmental condition has long been a major challenge to plant breeders. In practice, genotype by environment interaction complicates the identification of superior genotypes (Allard and Bradshaw, 1964). According to Hofstra (1972) growth in plants is generally the result of many complex processes each of which is influenced either directly or indirectly by environmental factors. He noted that plants depend for their growth and development on their genetic constitution and on their environment. It is widely recognize that improvement in plant types can make a very significant contribution to higher yields.

Crop breeders have been striving to develop genotypes with superior grain yield, quality and other desirable characteristics over a wide range of different environmental conditions. Genotype X Environment (GXE) interaction is one of the main complications in the selection of broad adaptation in most breeding programs. The phenotype of an organism is determined by the combined effect of the environment and the genotype which interact with one another. Numerous studies have shown that a proper understanding of the environmental and genetic factors causing the interaction as well as an assessment of their importance in the relevant GXE system could have a large impact on plant breeding (Magari and Kang, 1993; Basford and Cooper, 1998). GXE interaction occurs universally when genotypes are evaluated in several different environments (Becker and Lèon, 1988; Magari, 1989; Kang, 1990). Magari and Kang (1993) found that the contribution of different environmental factors, to the yield stability of maize in yield trials, had a significant impact on the heterogeneity of the results.

When environmental differences are large like in Ethiopia, it may be expected that the interaction of GXE will also be higher. As a result, one cultivar may have the highest yield in some environments while a second cultivar may excel in others. Hence, it is important to know the magnitude of the interactions in the selection of genotypes across several environments besides calculating the average performance of the genotypes under evaluation (Fehr, 1991; Gauch and Zobel, 1997).

Farmers and scientists want successful new wheat hybrids that show high performance for yield and other essential agronomic traits. Their superiority should be reliable over a wide range of environmental conditions but also over years. The cause of differences between genotypes in their yield stability is the occurrence of genotype by environment interactions (GEI).

Multi-locations trials play an important role in plant breeding and agronomic research. Data from such trails have three main objectives: a) to accurately estimate and predict yield based on limited experimental data; b) to determine yield stability and the pattern of response of genotypes across environments; and c) to provide reliable guidance for selecting the best genotypes or agronomic treatments for planting in future years and at new sites (Crossa, 1990).

A number of parametric statistical procedures have been developed over the years to analyze genotype by environment interaction and especially yield stability over environment. A number of different approaches have been used, for example, joint regression analysis and multivariate statistics, to describe the performance of genotypes over environments. To date considerable differences of opinion still exist between the leading protagonists of the different statistical approaches as to the best and most suitable procedure to be used for a specific data set or production region. The effects of genotype and environments are statistically non-additive, which means that differences between genotypes depend on the environment. For data sets with more than two genotypes and more than two environments, the GXE interactions are commonly calculated by analysis of variance (ANOVA), leading to an estimated variance component for GXE interactions. Performance tests over a series of environments give information on GXE interactions at population level, but from a practical point of view, it is important to measure the stability of the performance of individual genotypes (Eberhart and Russell, 1966).

1.2 STATEMENT OF THE PROBLEM

Ethiopia is known for its diverse/heterogeneous agro-ecology ranging from 100m below sea level in the Danakil depression to 4620m above sea level at mount Ras Dashan that contributes further to the problem of selecting stable wheat varieties for wider adaptation. To reduce the effect of genotype by environment interaction, crop improvement programs usually run performance trials across a wide range of environments to ensure that the selected genotypes have a high and stable performance across several environments.

Various studies have been conducted to analyze the effect of genotype by environment interaction on the Ethiopian wheat varieties. However no information is available on the genotype by environment interaction and stability in grain yield performance of these hybrids that are developed by Ethiopian seed enterprise. In addition, the changing environmental conditions of Ethiopia, the expansion of wheat to new agro-ecologies coupled with inadequate wheat varieties available for the different environments necessitate a rigorous and continuous study of genotype by environment interaction for dynamic crop improvement program.

1.3 OBJECTIVE OF THE STUDY

GENERAL OBJECTIVE

The general objectives of the study were to analyze genotype by environment interaction and stability of the Ethiopian wheat hybrids for grain yield across the target environments.

SPECIFIC OBJECTIVE

The specific objectives are:

- To evaluate the adaptability of 20 bread wheat genotypes and to identify the best performing ones for future uses.
- To utilize some statistical procedure for analyzing genotype by environment interaction and yield stability of Ethiopian bread wheat hybrids across 8 environments.
- To provide efficient statistical methods that guide breeders for releasing genotypes with adaptation to target environments , and
- To observe the pattern of grouping of the genotypes and the environments based on grain yield response of the hybrids.

2 LITERATURE REVIEW

2.1 BASIC CONCEPTS

2.1.1 GENOTYPE BY ENVIRONMENT INTERACTION

Genotype by Environment interaction is a common phenomenon in agricultural research. Differences between genotypic values may increase or decrease from one environment to another which might cause genotypes to even rank differently between environments. The Genotype by Environment interaction studies are somewhat complicated as they require integrated approaches which combine many fields including agriculture, biology, statistics, computer, and genetics.

A genotype or the genetic makeup of an organism is defined by Falconer and Mackay (1996) as the combination of alleles at a single autosomal locus in a diploid organism. The physical or visible characteristics resulting from the interaction between the genetic makeups and the environments are referred to as phenotype. Phenotype can be observed, measured, classified or counted. Organisms are determined neither by their genes nor by their environment; they are the consequence of the interaction of genes and environment (Suzuki *et al.*, 1981). Genotype describes the complete set of genes inherited by an individual that is important for the expression of a trait under investigation. Phenotype describes all aspects of the individual's morphology, physiology and ecological relationships. The genotype is essentially a fixed character of the organism; it remains constant throughout life and is unchanged by environmental effects. The phenotype changes continually and the direction of that change is a function of the sequence of environments that the individual experiences (Suzuki *et al.*, 1981).

The sum total of the effects of physical, chemical and biological factors of an individual other than its genotype is known as the environment. The individuals or populations of plants do not live in a vacuum but are surrounded and influenced by these factors. Comstock and Moll (1963) classified environments into two categories, (i) Macro-environment i.e. the environment which is associated with a given location or area at a particular period of time. (ii) Microenvironment i.e. the environment of a single organism as opposed to that of another organism growing at the same time and in almost the same place. It includes physical and chemical attributes of soil, climatic variables, solar radiation, insect pests and disease. The macro environments reflect a collection of micro-environments which are more alike within each macro-environment with the result that macro-environments substantially differ from each other. Environmental factors (non genetic factors) such as locations, growing seasons, years, rainfall, the amount of precipitation received in each season, temperature, etc. may have positive or negative impact on genotypes. Mather and Jinks (1982), Mukai (1988), and Wu and OíMalley (1998) report on two types of environmental variations: (1) micro environmental which cannot easily be identified or predicted (e.g., year-to-year variation in rainfall, drought conditions, extent of the insect damage) and (2) macro-environmental variances which can be identified or predicted (e.g., soil type, management practices, controlled temperatures). According to these investigators, the GXE interaction variance can only be estimated for the macro-environmental condition indicating that some variables that explain experiment differences are often unknown or can’t be measured.

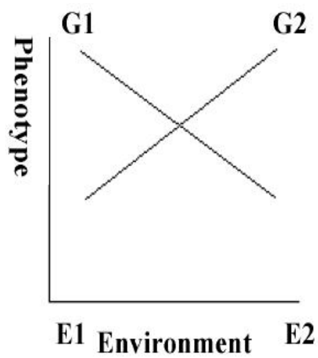
The terms ‘predictable and unpredictable environments’ were coined by Allard and Bradshaw (1964) to define and classify environments. The predictable environment includes the regular and more or less permanent features of the environment such as climate as determined by its longitude and latitude, soil type, rainfall and day length. It also includes what are called controllable variables (Perkins and Jinks, 1971) e.g. the level of fertilizer applied, sowing date and sowing density, amount of irrigation and others that can be artificially created. The unpredictable or uncontrollable environments, on the other hand, include weather fluctuations such as differences between seasons in terms of amount and distribution of rainfall and the prevailing temperature during the crop growth. The absence or low level of interaction will be useful for uncontrollable variables, whereas for the controllable variables a high level of interaction in the favorable direction is desirable to obtain maximal performance (Chahal and Gosal, 2002).

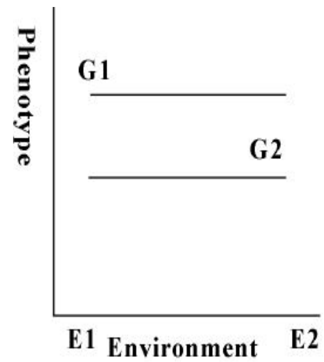
The association between the environment and the phenotypic expression of a genotype constitute the GXE interaction. The GXE interaction determines if a genotype is widely adapted for an entire range of environmental conditions or separate genotypes must be selected for different sub environments. When GXE interaction occurs, factors present in the environment (temperature, rainfall, etc.), as well as the genetic constitution of an individual (genotype), influence the phenotypic expression of a trait. The impact of an environmental factor on different genotypes may vary implying that the productivity of an animal or plant may also vary from one environment to the next. Breeding plans may focus on the GXE interaction to select the best genotype for a target population of environments.

A basic principle indicated by the GXE interaction is that even if all animals or plants were created equal (same genotypes), they will not necessarily express their genetic potential in the same way when environmental conditions (drought, temperature, disease pressure, stress, etc.) vary. This important concept may require genetic engineering of plants or animals specifically tailored to their environmental conditions.

2.1.1.1 INTERACTION ILLUSTRATION

Statistically, GXE interactions occur if the performance of genotypes varies significantly across environments or when differences between genotypes are not the same in all locations within and across years (Edmeades et al., 1989). Assuming 2 genotypes (G1 and G2) tested in 2 environments (E1 and E2), Fig 2.1 indicates the presence of GXE interaction since G1 is phenotypically superior to G2 in Environment 1 (E1) but inferior to G2 in E2. The phenotypic difference between G1 and G2 remains the same in the two environments representing no interaction between the genotype and the environment in Fig.2.2. Considering 3 environmental conditions (E1, E2, and E3) and 2 genotypes (G1 and G2), interpretation of the results could be more complicated. Fig 2.3 shows one type of GXE interactions for this situation where G1 is superior in performance to G2 in E1 and E3, but is inferior to G2 when exposed to E2. Agricultural researches have demonstrated that a genotype resulting in a good phenotype in one environment might not necessarily result in a good phenotype in another environment.





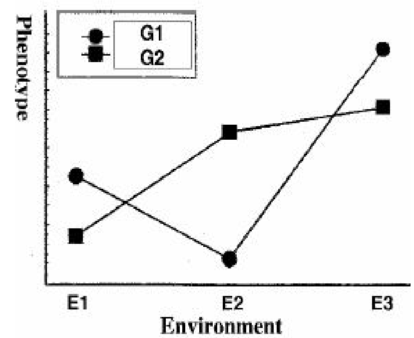


Fig 2.1:Shows GXE interaction present

Fig 2.2:Shows no GXE interaction

Fig 2.3: GXE interaction involving 2 genotypes (G_1 and G_2) and 3 Environments (E_1-E_3).

The presence of GXE interaction indicates the inconsistency of relative performance genotypes over environments (Hill et al., 1998). If two genotypes, A and B are evaluated in two environments 1 and 2, G x E interaction occurs when:

A_1-B_1 $\neq$ A_2- B_2 or A_1-B_1-(A_2-B_2) $\neq$ 0, where, A_1 is the performance of genotype A in environment 1, A_2 is the performance of genotype A in environment 2, B_1 is the performance of genotype B in environment 1, B_2 is the performance of genotype B in environment 2.

2.1.1.2 SIGNIFICANCE OF G X E INTERACTION

What breeders can do to overcome the problem of G x E interaction depends upon the relative importance of variance components. Moreover, breeding programmes aimed to develop stable genotypes also depend upon whether a breeder is dealing with predictable or unpredictable environmental variation. Whenever dealing with predictable environmental variation, the first step that should be taken is to identify the differences. There is no difficulty when differences are recognizable, for example, differences in the seasons such as varieties to be developed for the rainy season or post-rainy season. Breeders can develop varieties suitable for both these seasons because environmental variation is defined.

For variety trials, which are tested in the same locations (L) and genotypes (G) and over years (Y), G x E analysis of variance may be partitioned into components due to G x L, G x Y and G x L x Y. Significance of mean square for G x L generally suggests that the region for which genotypes are being bred comprises of a number of special environments. In such circumstances the geographic region could be subdivided into sub regions which are relatively homogeneous. Varieties should be bred which are specifically adapted to these ecotypes. Implication of G x Y interaction is very different from G x L interaction. This is so because year-to- year fluctuations cannot be predicted in advance and breeders can hardly aim their programmes to develop varieties suited to particular years (Dabholkar, 1999).

In some situations, environmental variation is predictable but can also be corrected. For example, saline soils can be corrected by certain agronomic practices or by addition of some amendments. This is easier and quicker than evolving varieties suitable for such situations. However, breeding of varieties suitable for saline or acidic soils is low cost input and also a relatively permanent solution to the problem.

It is relatively easier to develop varieties specifically adapted to predictable environmental situations than to breed for unpredictable environmental variations. The aim of the breeding programme should, therefore be to develop genotypes that can withstand unpredictable transient environmental fluctuations. In other words, breed widely adapted genotypes (Dabholkar, 1999).

According to Allard and Bradshaw (1964) "a variety which can adjust its genotypic or phenotypic state in response to transient fluctuations in environment in such a way that it gives high and stable economic returns for place and year, is termed as well buffered". Plant breeders generally agree that the new variety must show a high degree of stability in performance.

According to DeLacy *et al. (*1996), phenotypic performance of genotypes in combination with different environments can be analyzed to qualify the amount of variation attributable to the effects of the environment, genotype, and G x E interactions. DeLacy *et al*. (1996) recommended the use of restricted maximum likelihood (REML) analysis of variance and prediction of genotype performance by the use of the best linear unbiased predictors (BLUPs) to investigate patterns of adaptation of genotypes across environments.

The existence of G x E interactions complicates the identification of superior genotypes for a range of environments. G x E interactions can be an outcome of genotype rank changes from one environment to another, a difference in scale among environments, or a combination of these phenomena. According to Becker and Léon (1998), cultivar rank changes are of greater importance than scale change interactions in cultivar trials conducted over a series of environments. Hence, G x E interaction is critical only if it involves significant crossover interactions (significant reversal in genotypic rank across environments) (Becker and Léon, 1988).

The statistical analysis of G x E is important in applied statistics as well as for the analysis of experiments in plant breeding and crop production (Kang, 1996). Different statistical methods such as variance components, regression models, and multivariate analysis and cluster techniques have been proposed for the estimation and partitioning of G x E interactions (Freeman, 1973; Hill, 1975; Cox, 1984; Skroppa, 1984; Freeman, 1985, 1990; Westcott, 1986; Crossa, 1990). In many practical situations, the researcher is not interested in knowledge of the numerical amount of G x E interaction *per se*, but interested in the existence (or non-existence) of different rankings of genotypes. This concept of G x E interaction is closely related to the concept of selection in plant breeding. The breeder is mainly interested in the ranking of genotypes in different environments and in the changing of these rankings (Kang, 1996).

Breeders are interested in questions such as whether the best genotype in one environment is also the best in the other, which means that the relative characterizations and comparisons of the genotypes (orderings) are often more important than absolute characterizations and comparisons.

2.1.2 THE CONCEPT OF STABILITY

The term "stability of genotypes" is central to all types of analyses of G x E interactions especially with reference to plant breeding. Stability has been described in many different ways over the years and there have also been different concepts of stability (Lin *et al.*, 1986). Researchers use the terms adaptation, phenotypic stability and yield stability in different ways (Becker and Léon, 1988). Stability in common usage connotes consistency in performance that would mean minimum variation among environments for a particular genotype (Chahal and Gosal, 2002).

The stability with which a plant breeder is concerned implies stability in those aspects of phenotype which are important economically, such as grain yield and quality. Such stability may depend upon holding some aspects of morphology and physiology in a steady state but allowing others to vary. In this way, the desirable varieties will show low G x E interaction for agriculturally important characters, especially grain yield, but not necessarily for other characteristics. Two basic concepts of phenotypic stability are distinguished: i) the biological concept, and ii) the dynamic concept. The biological concept of stability refers to the constant performance of a genotype over a wide range of environments. This idea of stability is in agreement with the concept of homeostasis widely used in genetics. According to Becker and Léon (1988) in static stability a genotype posses unchanged performance regardless of variation of the environments, thus implying that its variance among environments is zero. This type is seldom a desired feature of crop cultivars, since no response to improved growing conditions would be expected. On the other hand dynamic stability, also termed as agronomical concept of stability, implies that a stable genotype should always give high yield expected at the level of productivity of the respective environments, i.e., a variety with G x E interaction as small as possible (Becker, 1981; Dabholkar, 1999). With quantitative traits, the majority of genotypes often react similarly to favorable or unfavorable environmental conditions. Becker and Léon (1988) stated that all stability procedures based on quantifying G x E interaction effects belong to the dynamic stability concept. This includes the procedures for partitioning the G x E interactions of Wricke's (1962) ecovalence and Shukla's (1972) stability of variance, procedures using the regression approach such as proposed by Finlay and Wilkinson (1963), Eberhart and Russell (1966) and Perkins and Jinks (1968), as well as non-parametric stability statistics.

All living things can make physiological adjustments which permit them to cope with fluctuations in their immediate environment. These adjustments themselves are known as adaptations. Adaptation is the property of a genotype which permits its survival under selection. An adapted genotype or population is simply one which performs better than the standard under comparison (Dabholkar, 1999). According to Simmonds (1962) adaptation has four separable aspects. These are:

1. Specific genotypic adaptation: it is close to adaptation of the corresponding genotypes to a limited environment.

2. General genotypic adaptation: is the capacity of a genotype to produce a range of phenotypes adapted to a variety of environments.

3. Specific population adaptation: is analogous to (1) and is the aspect of specific adaptation of heterogeneous population that is attributable to interaction between components rather than to the adaptations of components themselves.

4. General population adaptation: is analogous to general genotypic adaptation and is the capacity of a heterogeneous population to adapt to a variety of environments. The aim of a breeding programme is to identify genotypes which are widely adapted. Ramagosa and Fox (1993) concluded that if a genotype maintains high yield over a wide range of environments, it is referred to as having general or wider adaptation. On the other hand, if this is true only for a limited range of environments, that genotype has specific or narrow adaptation.

Further to the stability concept by Becker and Léon (1988), Lin *et al.* (1986) categorized stability in to three types:

I. If the among-environment variance of a genotype is small, the genotype is considered to be stable. This concept is useful for quality traits, disease resistance or for stress characters. According to this concept a genotype performs the same in different environments or under different environmental conditions. This stability is static or can be seen as a biological concept of stability (Becker and Léon, 1988). Genotype variances across environments (Si2) and the coefficient of variability (CVi) are used as parameters to describe this type of stability (Francis and Kannenburg, 1978).

II. A genotype is considered to be stable if its response to environments is parallel to the mean response of all genotypes in the trial. According to Becker and Léon (1988) this concept is called the dynamic or agronomic concept of stability. In this case, a stable genotype has no deviations from the general response to environments and creates a possible way of predicting the response of a genotype to a certain environment. Parameters used to describe this type of stability are regression coefficients (*bi*) (Finlay and Wilkinson, 1963) and Shukla's (1972) stability variance.

III. A genotype is considered to be stable if the residual mean square from the regression model on an environmental index is small. The environmental index is the mean yield of all the genotypes in each location minus the grand mean of all the genotypes in all locations. The method of Eberhart and Russell (1966) and Tai (1971) can be used for estimating type III stability.

2.2 REVIEW OF LITRETURES SPECIFIC TO THE STUDY

The information on the subject is not lacking but the inferences of various investigations are not consistent and differ greatly according to the material used and place of experimentation. However, the results of studies having some relevance to the subject are reviewed here briefly.

The variation in genotypic response from one environment to another is an intrinsic part of a genotypic behavior and without its estimation, assessment of a genotype remains incomplete (Westcott, 1987). Several researchers have studied this phenomenon and tried to specify and estimate the stability and adaptability of many wheat characters and their response to changing environments but the information regarding the subject is not consistent due to different genotypes and place of experimentation.

In the study of grain yield stability of wheat genotypes under different environment in Punjab, Pakistan by Rasul et al (2004), eighteen genotypes of wheat were planted at twenty four agro ecological zones. The combined analysis of variance revealed that there were significant differences among environments and genotypes for grain yield indicating the presence of variability in genotypes as well as diversity of growing conditions at different locations. In the analysis of variance for linear regression of hybrid mean yield on environmental mean yield, the sums of squares due to environments and GXE are partitioned into environments (linear), GXE (linear) and deviations from the regression model. They conduct stability analysis to check the response to GXE interactions and they obtained the mean squares due to GXE (linear) were non-significant depicting lack of genetic differences among genotypes for linear response to varying environments, while mean squares due to pooled deviations were highly significant, reflecting considerable differences among genotypes for non-linear response. Out of eighteen genotypes, only two wheat lines showed non-significant deviation from regression and their regression coefficient values were close to unity classified as stable varieties. However, one genotype with non-significant deviation from regression, unit regression coefficient and being superior in yield appeared as a prominent variety, to be the most stable for yield performance under varying environment.

Carvalho et al (1983) analyzed genotype-environment interactions using data on grain yield of bread wheat genotypes and observed that the highest yielding genotypes were the least stable, while genotypes with average yields were generally intermediate in stability but genotypes with above average in yield were the most stable.

Ashraf et al. (2001) analyzed genotype-environment interactions using the data on grain yield of wheat genotypes. Fifteen genotypes were studied and found that both linear and non linear components were highly significant and only two genotypes showed the regression coefficient closer to unity along with low deviation from regression. Similarly in a study involving twelve genotypes at ten locations revealed that GXE interaction was significant indicating the influence of environment on grain yield. The linear component has major contribution toward difference in stability of genotype. Highly significant differences for genotypes, environments and GXE interactions are reported earlier (Sial et al., 2000; Naazar et al., 2002).

Mosisa et al (2001) studied the interaction of 20 maize (Zea mays L.) genotypes of east African and CIMMYT origins in 9 locations, using linear regression analysis. The authors reported that none of the tested genotypes exhibited broad adaptability. Gelana et al. (2001) calculated Eberhart and Russell's stability statistics for 20 maize (Zea mays L.) genotypes tested in 18 location-year environments and recommended cultivars for drought stressed areas of Ethiopia. Regression analysis has also been used by Adugna and Elias (1994), Fekadu (1994), Asrat and Daniel (2004), among others.

Genotype x Environment interactions and correlation among some stability parameters of yield was observed by Letta (2007) in durum wheat genotypes. Twenty genotypes of durum wheat genotypes were tested over 15 environments during 2003-2005 cropping seasons in south east Ethiopia. Location within year and year variability were dominant sources of interactions. Nearly all the sources of variation in the combined analysis were highly significant except for genotype-year interaction and the analysis of AMMI showed that the four interactions of principal component analysis were highly significant. The stability analysis identified genotype 3 and 4 as more stable genotypes and recommended for commercial production in the south east Ethiopia. Highly significant rank correlation were found among Sd_i^2, W_i, Sx_i^2 and ASV implying their close similarity and effectiveness in detecting stable genotypes and they are equivalent in measuring stability.

Annicchiarico (1997) stated that AMMI analysis appears particularly useful for depicting adaptive responses of small grain cereals tested over whole Italy. At the same time, the researcher explained that joint regression and AMMI analysis are more likely to perform alike, and provide similar results, for small grain cereals over coastal and southern areas of Italy, where cold stress is limited.

Nachit et al. (1992) determined that the post-dictive AMMI models are superior to the linear regression techniques in accounting for and partitioning GEI in Mediterranean multi-location test trials of durum wheat. In addition, they expressed that predictive assessment is a useful statistical tool in estimating precise yield to make accurate and therefore successful selection in durum wheat breeding programs.

Kaya et al. (2002) suggested that the interaction of the 20 genotypes with six environments was best predicted by the first two principal components of genotypes and environments. Also, they proposed that bi-plots generated using genotypic and environmental scores of the first two AMMI components can be used by breeders and have an overall picture of the behavior of the genotype, the environment and GEIs. At the same time, Kaya et al. (2006) also evaluated bread wheat genotypes in multi-environment yield trials by using GGE bi-plot analysis and they determined that there were two proper rain fed mega environments in the Central Anatolian Plateau, also they recommended that two mega environments should be used by rain fed wheat improvement programs in order to enhance yield based selection gain in multi-environment yield trials.

In analyzing yield trial data for 36 field pea (Pisum sativum L.) genotypes and 8 environments, Girma et al. (2000) used AMMI analysis to partition the genotype x environment interaction matrix in to individual genotypic and environmental scores and came to the conclusion that AMMI-2 tends to be the best model for extracting patterns and rejecting noise from the data. Girma and colleagues further compared AMMI score-based analysis and observed values for clustering environments and field pea cultivars in to homogenous subsets. They indicated that cluster analysis following the AMMI modeling of GXE effects has the theoretical advantage of assessing similarity after separating the pattern from the noise portion of GXE variation and therefore suggested using the expected yields genotype-environment combinations according to the selected AMMI model, rather than the observed yield values. Using the results from AMMI plus cluster analysis, the authors identified few varieties presenting good yield and stability for the target areas.

3 MATERIALS AND METHODS

3.1 MATERIALS

Twenty wheat genotypes, listed in Table 3.1, were evaluated at four locations in Y1 and Y2 cropping seasons under irrigated conditions.

Table 3.1: List of wheat genotypes included in the study

No.	Genotype code
1	G_1
2	G_2
. . .	. . .
20	G_{20}

These wheat hybrids were selected based on their relative yield performance among the different experimental hybrids developed by the Ethiopian Seed Enterprise (ESC). These hybrids were a released variety adapted to the medium altitude wheat growing areas of Ethiopia. All the hybrids are categorized under the medium maturity group (between 140 and 145 days) and their broad adaptation zone is mid-altitude sub-humid which includes areas with an elevation range of 1000-2000m above sea level and an annual rainfall between 1000-1200mm.

The experimental layout was a randomized complete block design (RCBD) with four replicates. Planting method was on 30 cm apart at a seed rate of about 120 kg/ha. Plots were managed conventionally and followed the established local practices but usually the plot area ranged from 10 to 15m^2.

The trials were conducted under irrigated conditions and fertilization at each site and other management activities were done according to the practices of each farmer (co-operator) for his farm and the specific field. The whole plot was harvested to estimate grain yields and to reduce border effects, data were recorded from the two central rows of each plot. Grain yields are expressed in qt/ha at 12.5 moisture content.

The locations (Table 3.2) where the experiment was conducted were different in soil type and mean seasonal rainfall. Also the years differentiated in terms of mean seasonal rainfall. Therefore, locations in each year were considered as different environments. Hence, an environment is defined here as a location-year combination. Consequently, combinations of seasons (Y1 and Y2) and four locations were treated as eight environments (E_1-E_8).

Table 3.2: List of locations included in the study.

No.	Name	Year	
		Y1(2007)	Y2(2008)
1	Adet	E_1 (Adet-2007)	E_5 (Adet-2008)
2	Holeta	E_2 (Holeta-2007)	E_6 (Holeta-2008)
3	Kulumsa	E_3 (Kulumsa-2007)	E_7 (Kulumsa-2008)
4	Sinana	E_4 (Sinana-2007)	E_8 (Sinana-2008)

The data being considered here are obtained from trials conducted by the Ethiopian Institute of Agricultural Research (EIAR).

3.2 STATISTICAL METHODS

An important consideration in the analysis of genotype by environment data is whether environments are considered as fixed or random. If interest is in performance of genotypes in the particular testing environments, environments usually regarded as fixed. In such cases, inferences pertain to specific genotypes and specific environments rather than some population of genotypes and environments. If environments can be regarded as a random sample from a population of environments, it is appropriate to regard environmental and interaction effects as random.

A combined analysis of variance procedure is the most commonly used method to identify the existence of GEI from replicated multi-location trials. The analysis of multi-environment yield trials is usually complicated by the presence of GEI. If the GEI variance is found to be significant, one or more of the various methods for measuring the stability of genotypes can be used to identify the stable genotype(s). A wide range of methods is available for the analysis of GEI and can be broadly classified into four groups: the analysis of components of variance, stability analysis, multivariate methods and qualitative methods.

3.2.1 CONVENSIONAL ANALYSIS OF VARIANCE

Consider a trial in which the yield of **g** genotypes is measured in **e** environments each with **r** replicates. The classic model for analyzing the total yield variation contained in **ger** observations is the analysis of variance.

In the analysis of combined experiment of data from several environments, the first requirement is to assess the homogeneity of the error variance at the various environments. If the errors are homogeneous, the analysis can proceed. However, if the error variances are heterogeneous, the data will be transformed to produce homogenous variance or the locations may be separated into groups within which the variance is homogenous. Therefore, in order to combine the data, the error with each mean is measured should be tested for homogeneity, which is one of the basic assumptions of analysis of variance. This can be done first by providing a separate analysis of variance (ANOVA) for each environment. If the individual experiments are laid out as RCBD, separate ANOVA structure with sources of variation and degrees of freedom for each environment is:

Table 3.3: The general separate ANOVA structure for each environment

Source of variation	Df
Genotypes(entries)	g-1
Replicates(blocks)	r-1
Error	(g-1)(r-1)

The statistical model for this design is

$$y_{ij} = \mu + G_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, \ldots, g \\ j = 1, 2, \ldots, r \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

Where μ is an overall mean, G_i is the effect of the i^{th} genotype, β_j is the effect of the j^{th} block (replication) and ε_{ij} is NID (0,σ^2) random error term. Genotype and blocks are considered as fixed factors. Furthermore, the genotype and block effects are defined as deviations from the overall mean so that $\sum G_i = 0$ and $\sum \beta_j = 0$.

Then, quick test for homogeneity of variance is applied using either by F-test or chi-square test depending on the number variances or environments. The F-test for homogeneity of variances is applied whenever there are only two variances, with the F value computed as the ratio of the two variances-the larger variance in the numerator and the smaller variance in the denominator. This is well demonstrated through the standard F-test in the analysis of variance, which is used to test the homogeneity of two error mean squares. The chi-square test, which is commonly known as the Bartlett's test, should be used when more than two variances are tested. Since there are more than two environments in the study and hence have more than two variances (or error mean squares), the investigator will use the Bartlett's test for testing the equality (or homogeneity) of several variances.

The step-by-step procedure to apply the chi-square test (Bartlett's test) to test for homogeneity of e variances is found in appendix I.

When data are combined over locations and years for analysis, the analysis of variance structure may take different forms. These are:

- When the same location and randomization is used each year;
- When the same location is used, but different randomization is adopted each Year and
- When different locations are used each year.

The difference among the three ANOVA is mainly on the relationship between replications on one hand, and year and location on the other hand. When the same location and randomization is used each year, then replication is nested under location. But when the same location is used with different randomization, then replication is said to be nested under location by year interaction. When locations are different each year, the test also become slightly different.

The investigator will use the second forms of combined analysis, that is, when the same location is used, but different randomization is adopted each year. In this case,

$$SS_{Environment} = SS_{Year} + SS_{Location} + SS_{Location*Year},$$

$$SS_{Genotype*Environment} = SS_{Year*Genotype} + SS_{Location*Genotype} + SS_{Year*Location*Genotype}, \text{ and}$$

$$SS_{Replication\ within\ Environment} = SS_{Replication\ within\ Location*Year}$$

A combined ANOVA can be performed using either plot values or data of genotypes in individual environments that have been averaged across experiment replicates (i.e. genotype-environment cell means). To convert the result of an ANOVA performed on the basis of a cell mean in to result on a plot basis, for a constant number of experiment replicates (r), the sum of squares (SS) of effects must be multiplied by r (Cochran and Cox, 1957).

In multi-environment yield trials of g genotypes (i=1,2,…,g), e environments(j=1,2,…e) and r replicates(l=1,2,…,r) arranged in RCBD, the liner model for the conventional analysis variance(ANOVA) is

$$Y_{ijl}=\mu + G_i + E_j + GE_{ij} + B_{jl} + e_{ijl} \quad \text{.................................... (3.2)}$$

where $\mathbf{Y_{ijl}}$ is the observed yield response of the l^{th} plot of the i^{th} genotype at the j^{th} environment.

$\boldsymbol{\mu}$ is the overall mean yield of genotypes at all possible environments.

$\mathbf{G_i}$ is the effect of i^{th} genotype; and $\sum_{i=1}^{g} G_i = 0$

$\mathbf{E_j}$ is the random effect of the j^{th} environment drawn from a population with mean 0 and variance σ_E^2 and $\mathbf{E_j}$ is distributed as NID $(0, \sigma_E^2)$.

$\mathbf{GE_{ij}}$ is the interaction effect of the i^{th} genotype in the j^{th} environment. Since environments are random, this interaction is usually considered to be a random effect with mean 0 and variance σ_{GE}^2.

$\mathbf{B_{jl}}$ is the effect of the l^{th} replication in the j^{th} environment, and

$\mathbf{e_{ijl}}$ is the usual random error term with mean 0 and variance σ_e^2 and $\mathbf{e_{ijl}}$ is distributed as NID $(0, \sigma_e^2)$.

The following were determined from the ANOVA analysis, the effects of the genotypes, environments as well as their first order interactions. Genotypes were assumed to be fixed and environment effects random.

Table 3.4: The general analysis of variance (ANOVA) and mean square expectations.

Source	DF	MS	Expected mean square
Environment (E)	e-1	MS_E	$\sigma_e^2 + g\sigma_{R/E} + r\ \sigma_{GE}^2 + rg\ \sigma_E^2$
Replication within E	e(r-1)	$MS_{R/E}$	$\sigma_e^2 + g\ \sigma_{R/E}^2$
Genotype (G)	g-1	MS_G	$\sigma_e^2 + r\ \sigma_{GE}^2 + re\ \sigma_G^2$
GXE	(g-1)(e-1)	MS_{GXE}	$\sigma_e^2 + r\ \sigma_{GE}^2$
Error	e(g-1)(r-1)	MS_e	σ_e^2
Total	egr-1		

The within-environment residual mean square measures the error in estimating the genotype means due to differences in soil fertility and other factors, such as shading and competition from one plot to another.

The pooled error mean square (MS_e) to be inserted in the ANOVA can be estimated from the experimental errors of individual trials. If there is little variation in residual mean squares from one environment to another and for trials with the number of replicates, the pooled error variance (MSe) is found by averaging the residual mean squares of all environments:

$$\text{MSe} = \frac{\sum_{j=1}^{e} MS_{e(j)}}{e}$$

Where $MS_{e(j)}$ is the residual mean square (MS) for the j^{th} (j=1, 2,…, e) environment.

Analysis of variance of multi-location trials is useful for estimating variance components related to different sources of variation, including genotypes and genotype-environment interaction. In general, variance component methodology is important in multi-location trials, since errors in measuring the yield performance of a genotype arise largely from genotype-environment interaction. Therefore knowledge of the size of this interaction is required to (a) obtain efficient estimates of genotype effects and (b) determine optimum resource allocations, that is, the number of plots and locations to be included in future trials.

For balanced multi-location trials, that is, those with the same number of experimental units (genotypes or agronomic treatments) observed per site, estimation of the variance component is accomplished using the analysis of variance method. Each of the mean squares is known to estimate a linear function of the variance component defined in the model. These linear functions are called expected mean squares. By solving simultaneous equations we can obtain each of estimates of the variance components. Genetic and genetic-environment variance components can also be estimated by the restricted maximum likelihood (REML) method. This method is analogous to the analysis of variance, and both produce identical estimators for balanced data.

The ANOVA method for estimating variance components consists of equating mean squares to their expectations and solving the resulting set of simultaneous equations we obtain as follows.

Table 3.5: The general estimates of variance components and methods of determination.

Variance component	Method of determination
Environment(σ_E^2)	$\dfrac{MS_E - MS_{R/E} - MS_{GE} + MS_e}{rg}$
Replication within environment($\sigma_{R/E}$)	$\dfrac{MS_{R/E} - MS_e}{g}$
Genotype(σ_G^2)	$\dfrac{MS_G - MS_{GE}}{re}$
Genotype X Environment(σ_{GE}^2)	$\dfrac{MS_{GE} - MS_e}{r}$
Pooled error(σ_e^2)	MS_e

3.2.2 STABILITY ANALYSIS

3.2.2.1 THE ENVIRONMENTAL VARIANCE

For the specific case of evaluating phenotypic stability by using the static concept Römer(1917), as explained in Lin et al. (1986); Weber et al. (1996), proposed the use the variance of each genotype over the environments. Thus, to measure the static phenotype stability of the i^{th} genotype across a set of environments, the following could be used:

$$S_i^2 = \frac{\sum_{j=1}^{e}(y_{ij} - \bar{y}_{i.})^2}{e} \qquad i=1, 2,\ldots, g$$

Where y_{ij} is the mean yield of the i^{th} genotype in the j^{th} environment, and

$\bar{y}_{i.}$ is the marginal mean of genotype i.

A stable genotype has small variance. A problem with this method is that, in general, genotypes with high phenotypic stability measured through the environmental variance show low yield. In consequence, plant breeders do not use this method to evaluate yield stability across environments, or other related random variables. Derived quantities include the coefficient of variation ($CV_i = \sqrt{S_i^2} / \bar{y}_{i.}$) (Lin et al., 1986; Weber et al., 1996).

3.2.2.2 ECOVALENCE (W_i)

Wrick (1962) defined the concept of ecovalence as the contribution of each genotype to the GEI sum of squares. The ecovalence (W_i) or stability of the i^{th} genotype is its interaction with the environments, squared and summed across environments, and expressed as

$$W_i = \sum_{j=1}^{e} \left[y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right]^2$$

Where y_{ij} is the mean performance of genotype i in the j^{th} environment.

$\bar{y}_{i.}$ is the marginal mean of the i-th genotype.

$\bar{y}_{.j}$ is the marginal mean of the j-th environment.

$\bar{y}_{..}$ is the overall mean.

Genotypes with a low Wi value have smaller deviations from the overall mean across environments and are thus more stable. According to the meaning of ecovalence, this stable genotype possesses a high ecovalence (low values of Wi = high ecovalence).

Becker and Lèon (1988) illustrated ecovalence by using a numerical example of plot yields of genotype i in various environments against the respective mean of environments (fig 3.2).

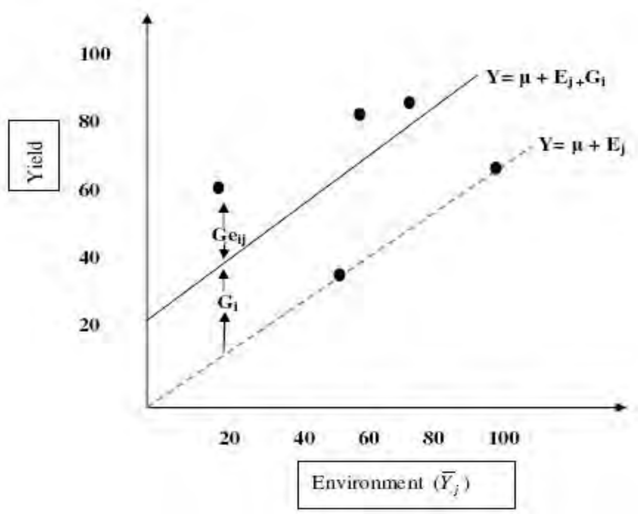


Fig 3.2: Graphical representation of GEI: The stability statistics ecovalence (W_i) is the sum of squares of deviations from the upper unbroken line.

The lower broken line estimates the average yield of all genotypes simply using information about the general mean (μ) and the environmental effects (E_j), while the upper unbroken line takes into account the genotypic effect (G_i) and therefore estimates the yield of genotype i deviations of yield from the upper unbroken line are the GEI effects of genotype i and are summed and squared across environments and constitute ecovalence (W_i).

3.2.2.3 SHUKLA'S STABILITY VARIANCE

Shukla (1972) defined the stability variance of genotype i as its variance across environments after the main effects of environmental means have been removed. Since the genotype main effect is constant, the stability variance is thus based on the residual (GE_{ij}+e_{ij}) matrix in a two-way classification. The stability statistics is termed "stability variance" (σ_i^2) and is estimated as follows:

$$\hat{\sigma}_i^2 = \frac{1}{(g-1)(g-2)(e-1)}\left[g(g-1)\sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 - \sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \right]$$

Where y_{ij} is the mean yield of the ith genotype in the jth environment, $\bar{y}_{.j}$ is the mean of all genotypes in jth environment, $\bar{y}_{i.}$ is the mean of all environments in ith genotype and $\bar{y}_{..}$ is the mean of all genotypes in all environments.

Since $W_i = \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$ and $MS_{GXE} = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2}{(g-1)(e-1)}$, $\hat{\sigma}_i^2$ can be rewritten as

$$\hat{\sigma}_i^2 = \frac{g}{(g-2)(e-1)} W_i - \frac{MS_{GXE}}{g-2}$$

Or equivalently

$$\hat{\sigma}_i^2 = \frac{g}{(g-2)(e-1)} W_i - \frac{\sum_{i=1}^{g} W_i}{(g-1)(g-2)(e-1)}$$

A genotype is called stable if its stability variance (σ_i^2) is equal to the environmental variance σ_E^2. A relatively large value of σ_i^2 will thus indicate greater instability of genotype i. As the stability variance is the difference between two sums of squares, it can be negative, but negative estimates of variances are not uncommon in variance component problem. Negative estimates of σ_i^2 may be taken as equal to zero as usual (Shukla, 1972).

The stability variance is a linear combination of the ecovalence, and therefore both W_i and σ_i^2 are equivalent for ranking purposes (Wrick and Weber, 1980).

3.2.2.4 REGRESSION COEFICIENT (b_i) AND DEVIATION MEAN SQUARE ($S_{d_i}^2$)

Joint linear regression (JLR) is a model used for analyzing and interpreting the nonadditive structure (interaction) of two-way classification data. The GEI is partitioned into a component due to linear regression (bi) of the i-th genotype on the environment mean, and a deviation (dij):

(GE_{ij}) = biEj + dij……………………………..(3.3) and thus, the model in equation (2) become

$$Y_{ijl}=\mu + G_i + E_j + (biEj + dij) + B_{jl} + e_{ijl} \quad \text{.........................(3.4)}$$

This model uses the marginal means of the environments as independent variables in the regression analysis and restricts the interaction to additive form. The method divides the (g-1) (e-1) df for interaction into g-1 df for heterogeneity among genotype regressions and the remainder (g-1) (e-2) for deviation. Hence, the general analysis of variance for the regression model relative to ANOVA model described in table 3.3 is given as follows.

Table 3.6: The general analysis of variance (ANOVA) for the regression model.

Source	DF
Environment (E)	e-1
Replication within E	E(r-1)
Genotype (G)	g-1
GXE	(g-1)(e-1)
Regression	g-1
Residual	(g-1)(e-2)
Error	E(g-1)(r-1)
Total	egr-1

Further details about interaction are obtained by regressing the performance of each genotype on the environmental means. Finlay and Wilkinson (1963) determined the regression coefficient by regressing variety mean on the environmental mean, and plotting the obtained genotype regression coefficients against the genotype mean yields. Figure 3.1 is a generalized interpretation of the genotype pattern obtained when genotype regression coefficients are plotted against genotype mean yields.

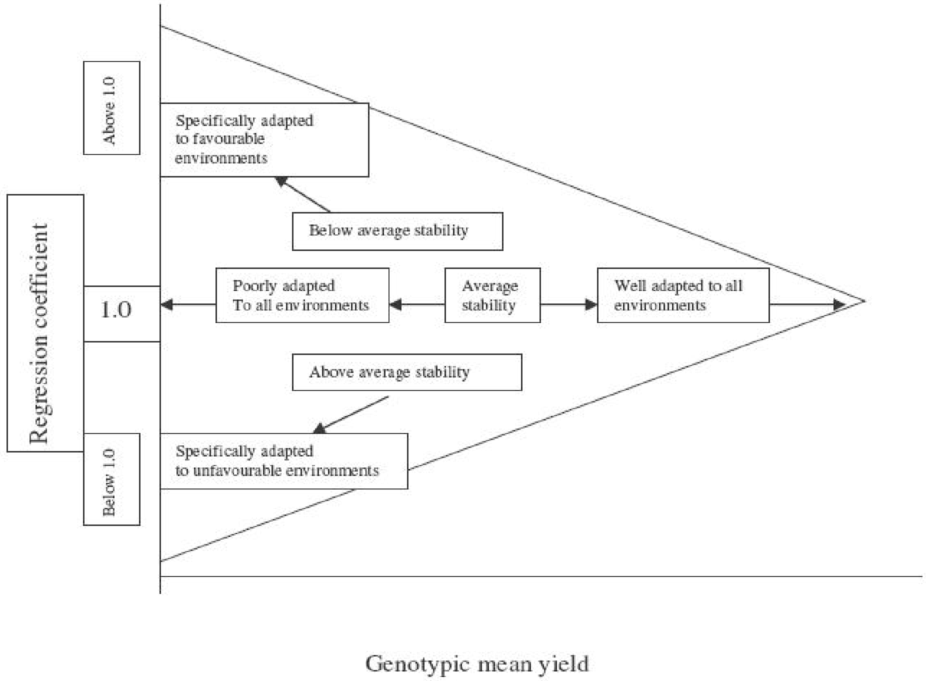


Fig 3.1 A generalized interpretation of the genotypic pattern obtained when, genotypic regression coefficients are plotted against genotypic mean, adapted from Finlay and Wilkinson (1963).

Eberhart and Russell (1966) defined a genotype with bi = 1 to be stable. Perkins and Jinks (1968) proposed an equivalent statistical analysis whereby the observed values are adjusted for environmental effects before the regression. Eberhart and Russell (1966) proposed pooling the sum of squares for environments and GEI and subdividing it into a linear effect between environments (with 1 df), a linear effect for genotype x environment (with e-2 df). In effect the residual mean squares from the regression model across environments is used as an index of stability, and a stable genotype is one in which the deviation from regression mean squares ($S^2_{d_i}$) is small.

Where $$S^2_{d_i} = \frac{1}{e-2}[\sum_{j=1}^{e} (Y_{ij} - \overline{Y}_i - \overline{Y}_j + \overline{Y}_{..})^2 - (b_i - 1)^2 \sum_{j=1}^{e} (\overline{Y}_j - \overline{Y}_{..})^2]$$

Several authors have criticized the regression approach for its statistical and biological limitations.

The statistical limitations are:

- Genotype mean is not independent of the overall mean of the environments,
- Errors associated with slopes of genotypes are not statistically independent since the sum of squares for deviation cannot be sub divided orthogonally among the g genotypes and
- It assumes a linear relationship between interaction and environmental means.

The biological limitations are:

- The relative stability of two genotypes depends not only on particular site, but also the set of other genotypes included in the analysis and
- When only few very low- or high-yielding sites are included in the analysis, fit of regression may be influenced by performance of genotype in those few extreme environments, which might mislead conclusions.

Nevertheless, the regression approach of Eberhart and Russel for determining stable genotypes is still extensively used.

3.2.3 THE ADDITIVE MAIN EFFECTS AND MULTIPLICATIVE INTERACTION (AMMI) METHOD

One of the main deficiencies of the combined analysis of variance of multi-location yield trials is that it does not explore any underlying structure within the observed non additivity (genotype-environment interaction). Analysis of variance fails to determine the pattern of response genotypes and environments. The valuable information contained in (g-1) (e-1) degrees of freedom is practically wasted if no further analysis is done.

Since the non additivity structure of data matrix has a non random (pattern) and random (noise) component, the advantages of the additive model are lost if the pattern component of the non additive structure is not further partitioned in to functions of one variable each. Williams (1952), Mandel (1961, 1969, 1971), and Gollob (1968) have delineated methods for analyzing and interpreting two-way tables with interaction. They show that the sum of squares for interaction can be further partitioned in multiplicative components related to Eigen values. Such method of analysis that links the analysis of variance with the principal component analysis is called additive main effects and multiplicative interaction (AMMI).

The additive main effects and multiplicative interaction (AMMI) method integrates analysis of variance and principal component analysis in to a unified approach (Bradu and Gabriel, 1978; Gauch, 1988). It can used to analyze multi-location trials (Gauch and Zobel, 1988; Zobel et al., 1988; 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, and linear regression models accounts for only a small portion of the interaction sum of squares. But AMMI analysis reveals a highly significant interaction component that has a clear agronomic meaning and it has no specific design requirements, except for a two-way data structure.

AMMI analysis first fits the additive main effects of genotypes and environments by the usual analysis of variance and then describes the non additive part, genotype-environment interaction, by principal component analysis.

The AMMI method is used for three main purposes:

i) 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 data set (Bradu and Gabriel, 1978; Gauch, 1985).

ii) To clarify genotype-environment interactions: AMMI summarizes patterns and relationships of genotype and environments(Kempton, 1984; Zobel et al., 1988; Crossa et al., 1990)

iii) 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; Crossa et al., 1990). Such gains may be used to reduce costs by reducing the number of replications, to include more treatments in the experiment, or to improve efficiency in selecting the best genotypes.

The AMMI model combines the analysis of variance for the genotype and environment main effects with principal component analysis of the GXE interaction. It has proven useful for understanding complex GXE interactions. The results can be graphed in a useful bi-plot that shows both main and interaction effects for both genotypes and environments. The bi-plot main effects of means vs. the first interaction principal component analysis axis (IPCA1) from AMMI analysis was used to study the patterns of response of G, E and GEI. It was also used to identify genotypes with broad and specific adaptation to target environments for grain yield.

3.2.3.1 THE AMMI MODEL

AMMI combines analysis of variance (ANOVA) in to a single model with additive and multiplicative parameters. After removing the replicate effect when combining the data, the ge observations are portioned in to two sources:

a) Additive main effects for genotypes and environments and
b) Non additive effects due to genotype-environment interaction.

The analysis of variance of the combined data expresses the observed (Y_{ij}) mean yield of the i^{th} genotype at the j^{th} environment as

$$Y_{ij} = \mu + G_i + E_j + GE_{ij} + e_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.5)$$

Where μ is the general mean; G_i, E_j and GE_{ij} represent the effect of the genotype, environment, and genotype-environment interaction, respectively; and e_{ij} is the average of the random errors associated with the l^{th} plot that receives the i^{th} genotype in the j^{th} environment.

We can write model (3.5) in matrix notation as:

$$Y = \mu 1_g 1_e^{'} + G 1_g^{'} + 1_e E^{'} + X + e \quad \ldots\ldots\ldots\ldots\ldots\ldots (3.6)$$

Where

$$Y = \begin{bmatrix} y_{11} & y_{12} \cdots y_{1e} \\ y_{21} & y_{22} \cdots y_{2e} \\ \vdots & \vdots \quad\; \vdots \\ y_{g1} & y_{g2} \quad\; y_{ge} \end{bmatrix} \quad \textit{is the data matrix of dimension gxe of grain yield of g genotypes in e environments;}$$

μ is a scalar representing the grand mean;

1_g and 1_e are gx1 and ex1 vectors with all elements equal to one, respectively;

$$G = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_g \end{bmatrix} \quad \textit{is a gx1 vector of main effects of genotypes;}$$

$$E = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_e \end{bmatrix} \quad \textit{is a ex1 vector of main effects of environments;}$$

$$X = \begin{bmatrix} GE_{11} & GE_{12} \ldots GE_{1e} \\ GE_{21} & GE_{22} \ldots GE_{2e} \\ \vdots & \vdots \;\; \ddots \;\; \vdots \\ GE_{g1} & GE_{g2} \quad GE_{ge} \end{bmatrix}$$ *is a gxe matrix where each element of a matrix specifies the interaction effect for the i^{th} genotype in the j^{th} environment ;*

And

Prime $(')$ is used to denote the transpose operation.

Now, the objective is to divide the non additive structure (GE_{ij}) from model (3.2) into functions of one variable each. This can be done by using the singular value decomposition of the matrix X.

Let X be a gxe matrix of real numbers. Then there exist a gxg orthogonal matrix U and an exe orthogonal matrix V such that

$X = U\Lambda V'$

Where the gxe matrix Λ has (i, i) entry $\lambda_i \geq 0$ for i=1, 2, …, min (g, e) and the other entries are zero. The positive constants λ_i are called the singular value of X.

Hence, the matrix expansion for the singular-value written in terms of the full dimensional matrices U, V, and Λ is

$$\underset{(gxe)}{X} = \underset{(gxg)}{U} \; \underset{(gxe)}{\Lambda} \; \underset{(exe)}{V'}$$

Where U has g orthogonal eigenvectors of XX' as its columns, V has e orthogonal eigenvectors of $X'X$ as its columns, and Λ is specified above.

The singular-value decomposition can also be expressed as a matrix expansion that depends on the rank n of X. Specifically; there exist n positive constants $\lambda_1, \lambda_2, \ldots, \lambda_n$, n orthogonal gx1 unit vectors $\alpha_1, \alpha_2, \cdots, \alpha_n$, and n orthogonal ex1 unit vectors $\gamma_1, \gamma_2, \ldots, \gamma_n$, such that

$$X = U\,\Lambda V' = \sum_{i=1}^{n} \lambda_i \alpha_i \gamma_i' = U_n \Lambda_n V_n'$$

Where $U_n = [\alpha_1, \alpha_2, \ldots, \alpha_n] = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{g1} & \alpha_{g2} & \cdots & \alpha_{gn} \end{bmatrix}$, $V_n = [\gamma_1, \gamma_2, \ldots \gamma_n] = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{e1} & \gamma_{e2} & \cdots & \gamma_{en} \end{bmatrix}$, and Λ_n is an nxn diagonal matrix with diagonal entries λ_i.

Here $XX^{'}$ has an eigenvalue-eigenvector pair (λ_i^2, α_i), so

$$XX^{'}\alpha_i = \lambda_i^2 \alpha_i$$

With $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2 > 0 = \lambda_{n+1}^2, \lambda_{n+2}^2, \ldots, \lambda_g^2$ $(for\ g > e)$. Then $\gamma_i = \lambda_i^{-1} X^{'} \alpha_i$. Alternatively, the γ_i are the eigenvectors of $X^{'}X$ with the same nonzero eigenvalues λ_i^2.

The normalization and orthogonality constraints are

$1_g^{'} U = 1_e^{'} V =$ **0 (0** is a 1xn vector of zeros) and

$U^{'}U = V^{'}V = \mathbf{I}_n = U^{'}_n U_n = V_n^{'} V_n =$

Then, the model in equation (3.6) becomes:

$$Y = \mu\, 1_g 1_e^{'} + G1_g^{'} + 1_e E^{'} + U \Lambda V^{'} + e \ldots\ldots\ldots\ldots.. (3.7)$$

Or, equivalently

$$Y_{ij} = \mu + G_i + E_j + \sum_{k=1}^{n} \lambda_k \alpha_{ik} \gamma_{jk} + e_{ij} \ldots\ldots\ldots\ldots\ldots (3.8)$$

in which GE_{ij} is represented by:

$$\sum_{k=1}^{n} \lambda_k \alpha_{ik} \gamma_{jk}$$

under the restrictions: $\sum_{i=1}^{g} \alpha_{ik} = \sum_{j=1}^{e} \gamma_{jk} = 0$

$$\sum_{i=1}^{g} \alpha_{ik}^2 = \sum_{j=1}^{e} \gamma_{jk}^2 = 1 \quad \text{and}$$

$$\sum_{i=1}^{g} \alpha_{ik} \alpha_{ik^{'}} = \sum_{j=1}^{e} \gamma_{jk} \gamma_{jk^{'}} = 0 \ \ for\ k \neq k^{'} = 1, 2, \ldots, n$$

Where, Y_{ij} is the mean yield of i[th] genotype in the j[th] environment;

μ is the grand mean;

G_i and E_j are the genotype and environment deviation from the grand mean, respectively;

α_{ik} *and* γ_{jk} are the genotype and environment principal component scores for axis k;

n is the maximum number of multiplicative terms (or n=rank(X)).

λ_k is the k[th] singular value of X (square root of the eigenvalue of XX' or $X'X$) with $\lambda_1^2 \geq \lambda_2^2 \geq, \ldots, \geq \lambda_n^2 > 0$.

e_{ij} is the error term.

The model in (3.7) is called the **full AMMI model** or AMMI-n.

The additive parameters are μ, G_i and E_j while, λ_k, α_{ik} *and* γ_{jk} are the multiplicative parameters. The α_{ik} are genotype interaction parameters that measure genotype sensitivity to hypothetical environmental factors denoted by environmental interaction parameters γ_{jk}.

A part of the interaction may well be random error or otherwise unsuitable for bilinear modeling. Consequently, the portioning of the interaction into multiplicative terms is carried out only partially i.e., only a few multiplicative terms of the $\lambda_k \alpha_{ik} \gamma_{jk}$ type (typically, one or two or three such terms) are retained; the remaining terms are pooled together and considered as random residual error, producing a reduced model.

In a more general setting, s < n non-null eigenvalues are kept producing a **reduced AMMI model** (denoted by AMMI-s) given by

$$Y_{ij} = \mu + G_i + E_j + \sum_{k=1}^{s} \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij} + e_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.9)}$$

i=1, 2,…, g; j=1, 2, …, e.

Where, s indicates the number of multiplicative terms necessary for an adequate description of the GxE

$\rho_{ij} = \sum_{k=s+1}^{n} \lambda_k \alpha_{ik} \gamma_{jk}$ is the residual not accounted for by retained multiplicative terms for the GXE.

The AMMI model with s multiplicative terms for replicated data may be written as

$$Y_{ijl} = \mu + G_i + E_j + B_{jl} + \sum_{k=1}^{s} \lambda_k \alpha_{ik} \gamma_{jk} + \rho_{ij} + e_{ijl}$$

Where Y_{ijl}, μ, G_i, E_j, B_{jl}, and e_{ijl} are as in model (2).

The simplest AMMI model (AMMI-0), $Y_{ijl} = \mu + G_i + E_j + B_{jl} + e_{ijl}$, estimates the additive main effects (i.e., genotypes and environments) without considering interaction.

AMMI-1, $Y_{ijl} = \mu + G_i + E_j + B_{jl} + \lambda_1 \alpha_{i1} \gamma_{j1} + \rho_{ij} + e_{ijl}$, combines the main effects from AMMI-0 with interaction effects estimated from the first multiplicative term. AMMI-2, $Y_{ijl} = \mu + G_i + E_j + B_{jl} + \lambda_1 \alpha_{i1} \gamma_{j1} + \lambda_2 \alpha_{i2} \gamma_{j2} + \rho_{ij} + e_{ijl}$, considers main effects plus the first two multiplicative terms. AMMI-3 to AMMI-n (the full AMMI model) includes sequentially, one more multiplicative term each.

To estimate the unknown parameters in the model: $Y_{ij} = \mu + G_i + E_j + \sum_{k=1}^{n} \lambda_k \alpha_{ik} \gamma_{jk} + e_{ij}$, one usually first uses row/column means for the main effects and then performs a singular value decomposition of the residual matrix for the interaction parameters. This classical approach corresponds essentially to a least square fit of the full model. That is, estimates of the overall mean (μ) and the main effects (G_i and E_j) are obtained in the context of a simple two-way ANOVA of the array of means $\underset{(gxe)}{Y}$. The residuals from this array then constitute the array of interactions $\underset{(gxe)}{Z} = \lfloor z_{ij} \rfloor$, where, $z_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$, and the multiplicative interaction terms are estimated from the singular value decomposition (SVD) of this array. Thus, λ_k is estimated by the k^{th} singular value of Z, α_{ik} is estimated by i^{th} element of the left singular vector $\alpha_{k(g x1)}$ and γ_{jk} is estimated by j^{th} element of the right singular vector $\gamma'_{k(1xk)}$ associated with λ_k (Mandel, 1971).

Denoting estimates by a caret ($\wedge$) placed over the symbols of the parameters, the least square estimators of μ, G_i, and E_j are $\hat{\mu} = \bar{y}_{..}$, $\hat{G}_i = \bar{y}_{i.} - \bar{y}_{..}$ and $\hat{E}_j = \bar{y}_{.j} - \bar{y}_{..}$ respectively.

And $\lambda_k \alpha_{ik} \gamma_{jk}$ is estimated from the k[th] component of the SVD of $\underset{(gxe)}{Z} = \lfloor z_{ij} \rfloor$. That is, from SVD, the matrix Z with rank n can be expressed as

$$Z = \hat{U}\,\hat{\Lambda}\,\hat{V}' = \hat{U}_n \hat{\Lambda}_n \hat{V}_n' = \sum_{i=1}^{n} \hat{\lambda}_i \hat{\alpha}_i\, \hat{\gamma}_i'$$

Where $\hat{U}_n = [\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_n]$ has n orthogonal unit eigenvectors of ZZ' as its columns, $\hat{V}_n = [\hat{\gamma}_1, \hat{\gamma}_2, \ldots \hat{\gamma}_n]$ has n orthogonal unit eigenvectors of $Z'Z$ as its columns, and $\hat{\Lambda}$ is an nxn diagonal matrix with diagonal entries $\hat{\lambda}_i > 0$.

Here ZZ' has an eigenvalue-eigenvector pair $(\hat{\lambda}_i^2, \hat{\alpha}_i)$, so

$$ZZ'\,\hat{\alpha}_i = \hat{\lambda}_i^2\,\hat{\alpha}_i$$

With $\hat{\lambda}_1^2, \hat{\lambda}_2^2, \ldots, \hat{\lambda}_n^2 > 0 = \hat{\lambda}_{n+1}^2, \hat{\lambda}_{n+2}^2, \ldots, \hat{\lambda}_g^2$ $(for\ g > e)$. Then $\hat{\gamma}_i = \hat{\lambda}_i^{-1} Z'\,\hat{\alpha}_i$. Alternatively, the $\hat{\gamma}_i$ are the eigenvectors of $Z'Z$ with the same nonzero eigenvalues $\hat{\lambda}_i^2$.

3.2.3.2 ANALYSIS OF VARIANCE (ANOVA) FOR AMMI MODEL

The computations of the SS and DF for genotype, environment and genotype x environment follow the standard procedure but the SS and DF for interaction components are described as follows.

Consider the identity from the cell-means:

$$SS_{GxE} = \sum_{i=1}^{g} \sum_{j=1}^{e} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = trace(ZZ') = trace(Z'Z)$$

Where

$$Z = [z_{ij}] = [y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}]$$

From SVD, we have

$$trace(ZZ') = trace(\hat{U}_n\,\hat{\Lambda}_n\,\hat{V}_n'\,\hat{V}_n\,\hat{\Lambda}_n'\hat{U}_n') \quad \text{Since } Z = \hat{U}\,\hat{\Lambda}\,\hat{V}'$$

$$= trace(\hat{U}_n\,\hat{\Lambda}_n\,\hat{\Lambda}_n'\hat{U}_n') \quad \text{Since } \hat{V}_n'\hat{V}_n = I_n$$

From the property of trace of matrix (i.e. trace (AB) =trace (BA), for any two square matrix A and B of the same order), we have

$$trace(ZZ') = trace(\hat{U}'_n \hat{U}_n \hat{\Lambda}_n \hat{\Lambda}'_n)$$

$$= trace(\hat{\Lambda}_n \hat{\Lambda}'_n) \qquad \text{Since } U'_n U_n = I_n$$

$$= trace(\hat{\Lambda}_n^2) = \sum_{k=1}^{n} \hat{\lambda}_k^2$$

Therefore,

$$SS_{GxE} = trace(ZZ') = \sum_{k=1}^{n} \hat{\lambda}_k^2$$

For replicated data,

$$SS_{GxE} = r\sum_{i=1}^{g}\sum_{j=1}^{e}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = r\sum_{k=1}^{n} \hat{\lambda}_k^2$$

The sum of squares of the k-th component in the AMMI model, S_k, is given by

$$S_k = \sum_{i=1}^{g}\sum_{j=1}^{e}(\hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk})^2$$

$$= \hat{\lambda}_k^2 \sum_{i=1}^{g}\sum_{j=1}^{e}(\hat{\alpha}_{ik}^2 \hat{\gamma}_{jk}^2)$$

$$= \hat{\lambda}_k^2 (\sum_{i=1}^{g} \hat{\alpha}_{ik}^2)(\sum_{j=1}^{e} \hat{\gamma}_{jk}^2)$$

$$= \hat{\lambda}_k^2$$

For replicated data,

$$S_k = r\hat{\lambda}_k^2 \qquad for \;\; k = 1, 2, \ldots, rank(Z) = n$$

Two systems for assigning DF in the AMMI model, those of Gollob (1968) and Mandel (1971), are particularly popular (Gauch and Zobel, 1996). However, the authors warn that, unfortunately, there are disagreements between these methods. Choosing one requires both theoretical and practical considerations. The approach of Gollob (1968) is very easily applied, since the number of degree of freedom for component k (v_k) of the interaction is simply defined to be

$$v_k = (g-1) + (e-1) - (2k-1)$$

Mandel (1971) defines the number of degrees of freedom for component k (v_k) to be

$$v_k = \mathrm{E}(\hat{\lambda}_k^2 / \sigma^2)$$

Where σ^2 is the error variance. However, simulations then have to be conducted to evaluate the number of degrees of freedom in particular cases. Mandel gives some tables derived from such simulations for a limited set of conditions. These tables, however, are not exhaustive and this reduces the practical utility of the method.

Gauch (1992) discusses the question of obtaining the degrees of freedom for the multiplicative components of an AMMI model. He concludes that rigorous simulation seems unnecessary or impractical, and generally recommends the use of Gollob's system when one is using an F-test approach. Therefore, the full joint analysis of variance based on Gollobs system has the structure as shown in table.

Table 3.7: The general analysis of variance for the full AMMI model.

Sources of variation	DF	Sum of square
Environment (E)	e-1	SS_E
Rep within environment (R/E)	e(r-1)	$SS_{R/E}$
Genotype (G)	g-1	SS_G
Genotype x Environment interaction (GxE)	(g-1)(e-1)	SS_{GxE}
PCA-1	v_1=g+e-1-(2x1)	$r\hat{\lambda}_1^2$
PCA-2	v_2=g+e-1-(2x2)	$r\hat{\lambda}_2^2$
.	.	.
.	.	.
.	.	.
PCA-n	v_n=g+e-1-(2xn)	$r\hat{\lambda}_n^2$
Experimental error (e)	e(g-1)(r-1)	SS_e
Total	Ger-1	

METHOD OF DETERMINING THE OPTIMAL NUMBER OF INTERACTION COMPONENTS (OR s)

Nonadditive effects are frequently observed in two-way tables, and as observed by Snee (1982), the interpretation of nonadditivity is less of a problem if replicate observations are present for each of the cells of the table. Daniel (1976) points out that the nonadditivity is often associated with just a few rows and columns of the table. Hence, good prediction of the true trait response in each cell of the two way table can be achieved by truncating the AMMI model, so criteria for determining the number of components needed to explain the patterns of interactions have been the objects of some research.

There are many tests for deciding how many components should be retained in the additive and multiplicative interaction models. Dias and Krzanowsky (2003) present two methods outlined by Krzanowsky (1987) and Gabriel (2002) based on a full ‘leave-one-out’ procedures that optimizes the cross-validation process (i.e. maximizes the number of data points left in the set at each iteration without incurring bias due to resubstitution.), and two other methods (the Gollob and Cornelius tests) based on F-tests to determine the number of components (or s) to be retained in the AMMI model. For this study, the method based on F-test suggested by Gollob (1968) will be used. The Gollob (1968) approximate F test assumes $r\hat{\lambda}_k^2 / \sigma^2$ is distributed as a chi-square variable, and he suggests using the statistic

$$F = r\hat{\lambda}_k^2 / (f_1 MS(error\ mean))$$

against an F distribution with f_1=g+e-1-(2k) and e(g-1)(r-1) degrees of freedom to test the kth multiplicative term of the model for significance. Therefore, selection of the optimal model is based on F tests for the successive terms of the interaction, the number of included terms corresponding to the number of significant components. Hence, ANOVA for the reduced (truncated) AMMI model looks like as follows.

Table 3.8: Analysis of variance (ANOVA) for reduced (truncated) AMMI model

Sources of variation	DF	Sum of square
Environment (E)	e-1	SS_E
Rep within environment (R/E)	e(r-1)	$SS_{R/E}$
Genotype (G)	g-1	SS_G
Genotype x Environment interaction (GxE)	(g-1)(e-1)	SS_{GxE}
PCA-1	v_1=g+e-1-(2x1)	$r\hat{\lambda}_1^2$
PCA-2	v_2=g+e-1-(2x2)	$r\hat{\lambda}_2^2$
.	.	.
.	.	.
.	.	.
PCA-s	v_n=g+e-1-(2xs)	$r\hat{\lambda}_n^2$
Residual	$(g-1)(e-1)-\sum_{k=1}^{s} v_k$	$SS_{GxE}-r\sum_{k=1}^{s}\hat{\lambda}_k^2$
Experimental error (e)	e(g-1)(r-1)	SS_e
Total	ger-1	

3.2.3.3 AMMI MODEL AND THE BIPLOT

The concept of Bi-plot was developed by Gabriel (1971) to graphically display a rank-two matrix. The significance of this concept is that if a two-way data set can be sufficiently approximated by a rank-two matrix, then it can be graphically displayed and investigated. Bradu and Gabriel (1978) explored the use of bi-plot as a diagnostic tool for choosing an appropriate model for the analysis of two-way data.

Genotypes and environments are depicted as points on a plane where the position of the point for genotype i is given by the estimates for the genotypic scores ($\lambda_2^{1/2}\alpha_{i2}, \lambda_1^{1/2}\alpha_{i1}$) while the point coordinates for environment j originate from the estimates for the environmental scores ($\lambda_2^{1/2}\gamma_{j2}, \lambda_1^{1/2}\gamma_{j1}$). In a vector representation, the genotype and environment point determine lines starting at the origin (0, 0). The interaction effect of genotype i in environment j is approximated by projecting the genotype point ($\lambda_2^{1/2}\alpha_{i2}, \lambda_1^{1/2}\alpha_{i1}$) onto the line determined by the environmental vector, which has a slope ($\lambda_1^{1/2}\gamma_{j1} / \lambda_2^{1/2}\gamma_{j2}$), where distance from the origin provides information about the magnitude of the interaction. The angle between the vector of genotype i and environment j tells us something about its nature: the interaction is positive for acute angles, negligible for right angles, and negative for obtuse angles. The abscissa shows the PCA-2 scores and the ordinate shows the PCA-1 scores, and thus genotype/environments that appear almost on a vertical line have similar interaction pattern for PCA-2, and those that fall almost on a horizontal line have similar interaction pattern with PCA-1. Genotypes/environments that fall in the bottom left quadrant have a negative interaction along both axes, and those that fall in the top right quadrant have positive interaction with both axes, while the others have different sign of interactions for both axes. Genotype with large PCA-1 or PCA-2 or both (either positive or negative) have high interactions, whereas genotypes with PCA-1 or PCA-2 scores near zero have small interactions for the corresponding axis. Since PCA-2 is more exposed to noise than PCA-1, genotypes that have large negative or positive values along the latter axis have stronger interaction than those along the former axis. Points representing genotypes and environments that are close to the origin contribute little to the interaction and can be well approximated by additive terms alone. A biplot of PCA axis 1 versus PCA axis 2 depicts the level of interaction inherent in the data; it does not present a contrast of interaction and main effects. To investigate main effects and interaction, a biplot of PCA axis 1 versus genotype and environment yield mean will be used.

To generate a biplot that can be used in visual analysis of MET data, the SVs have to be partitioned into the genotype and environment eigenvectors so that equation (3.9) can be written in the form

$$Y_{ij} = \mu + G_i + E_j + \sum_{k=1}^{s} \lambda_k \alpha_{ik}^* \gamma_{jk}^* + \rho_{ij} + e_{ij} \quad \text{..................................(3.10)}$$

Where α_{ik}^* and γ_{jk}^* are called interaction PCA axis k scores for genotype i and environment j, respectively. In a biplot, genotype i is displayed as a point defined by all α_{ik}^* values, environment j is displayed as a point defined by all γ^*_{jk} values (k=1 and 2 for two-dimensional biplot). Singular-value partitioning is implemented by

$$\alpha^*_{ik} = \lambda_k^{f_k}\alpha_{ik} \quad and \quad \gamma^*_{jk} = \lambda_k^{1-f_k}\gamma_{jk}$$

where f_k is the partition factor for PCA axis k. As mentioned above f_k can be any real number between 0 and 1($0 \leq f_k \leq 1$), leading to numerous ways to construct a biplot. The influence of different partitioning factors on a biplot has rarely been documented, except in Yan (2002). Three special scaling methods will be reviewed in some detail below.

a) Environment-focused scaling. It is referred to as environment focused scaling if f_k=0, i.e., $\alpha^*_{ik} = \alpha_{ik}$ *and* $\gamma^*_{jk} = \gamma_{jk}$. In this scaling, the environmental scores are in the original unit of yield (kg/ha), and the genotype scores are in normalized form (unit less). Because all the SV is partitioned into the environment scores, the range of the environment score is likely many times greater than that of the genotypes, and when directly plotted, the genotypes are likely to be crowded in the biplot. Therefore, a biplot based on environment-focused scaling is most suitable for visualizing the interrelationship among the environments but not that of genotypes. To generate a biplot in which the ranges of the genotypes and the environments are comparable, the genotypes scores for both axes can be multiplied by an arbitrary number. Multiplying both axes of the genotype scores with a positive number is equivalent to multiplying such a number to each element of the residual data matrix and will not alter the genotype x environment pattern of the data.

b) Genotype-focused scaling. It is referred to as genotype-focused scaling when f_k=1, i.e., when the SV is partitioned entirely into the genotype eigenvectors so that $\alpha^*_{ik} = \lambda_k\alpha_{ik}$ *and* $\gamma^*_{jk} = \gamma_{jk}$. In this scaling, the unit of the genotype scores (α^*_{ik}) is the original unit of yield, and the environment score (γ^*_{jk}) are unit less. Because all of the SV is partitioned into the genotype scores, the ranges of the genotype scores are likely to be many times greater than that of the environment scores. As a result, the environments in the biplot are likely to be crowded relative to the genotypes. Therefore, a biplot based on genotype-focused scaling is suitable for evaluating the genotypes but not the environments. To generate a biplot in which the ranges of the genotypes and environments are comparable, the environment scores of both axes can be multiplied by an arbitrary factor.

c) Symmetric scaling. It is called symmetrical scaling when f_k=0.5 so that $\alpha_{ik}^* = \lambda_k^{0.5}\alpha_{ik}$ *and* $\gamma_{jk}^* = \lambda_k^{0.5}\gamma_{jk}$. 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 the square root of the original [$(kg/ha)^{0.5}$]. This property makes it possible to visualize the relative magnitude of genotype variation and environment variation for both axes. This is the scaling method used in AMMI analysis (Gauch, 1988). It is intermediate between the environment-focused scaling and the genotype-focused scaling in all aspects.

The results of AMMI analysis can be presented graphically in the form of biplots (Gauch, 1988; Zobel et al., 1988; Shaffi & Price, 1998; Vargas et al., 1999; Ebdon & Gauch, 2002) in which the genotype and environment scores of the first two or three bilinear (multiplicative) terms are represented by vectors in a space, with starting points at the origin and end points determined by the scores. Usually the environmental and genotypic scores of the first and second bilinear terms are plotted. The distance between two genotype vectors (their end points) is indicative of the amount of interaction between the genotypes. The cosine of the angle between two genotype (or environments) vectors approximates the correlation between the genotypes (or environments) with respect to their interaction. Acute angles indicate positive correlation, with parallel vectors (in exactly the same directions) representing a correlation of 1. Obtuse angles represent negative correlations, with opposite directions indicating a correlation of -1. Perpendicularity of directions represents a correlation of zero. The relative amount of interaction for a particular genotype over environments can be obtained from orthogonal projection of the environmental vectors on the line determined by the direction of the corresponding genotype vector. Environmental vectors having the same direction as the genotype vectors have positive interactions where as vectors in the opposite directions have negative interactions.

4 RESULT AND DISCUSSION

4.1 INTRODUCTORY REMARKS

Most of the statistical analyses presented in this thesis were done using the SAS software. The SAS program statements for individual ANOVAs for each environment, the combined ANOVA across environments, LR analysis and AMMI analysis are found in Appendix II. The software IRRISTAT, released by the International Research Institute (IRR) of Manila, Philippines, was used for the Gollob's F-test.

The usual diagnostic plots-including a normal probability plot of residuals, a histogram of residuals, plot of residuals versus fitted values, plot of residuals versus level of regressor variable-and formal statistical procedures to assess model assumptions for the individual ANOVAs (i.e. for yield data at each environment), and the separate LR models (for each genotype yield) were performed. Examination of the results do not reveals any series violations of the assumptions that errors are normally and independently distributed with mean zero and constant variance. In the Shapiro-Wilk W test for normality, the p-value is based on the assumption that the distribution is normal. In this study, the p-value for each of the separate ANOVA model is very large (>0.05), indicating that we cannot reject the hypothesis that the residual is normally distributed (see the output of the proc univariate in Table 3 of Appendix III.). For homogeneity of residual variance, in each of the separate ANOVA, the investigator uses the Bartlett's test by considering genotype as a group. Based on this, the p value for each of the separate ANOVA is much greater than 0.05 (see Table 2 in Appendix III) indicating the hypothesis that the residual variances in each of the separate ANOVA are homogeneous cannot be rejected. Thus, the residuals have constant variance. Plots of residual versus the fitted value for each of the separate ANOVA are given in fig 1 of Appendix III. There should be no relationship between the size of the residual and the fitted values. These plots reveal nothing of unusual interest. Thus, the usual interpretation for each of the separate ANOVA is valid since all assumptions concerning the data are met.

Separate analyses of variance (ANOVA) were first done with the classical RCBD model with replications (block) at each environment. Summary results of the separate analysis are given in Table 1 of appendix III. In the analysis of combined experiment of data from several environments, the first requirement is to assess the homogeneity of the error variance at the various environments. If the error variances are homogeneous, the analysis can proceed with the original data. This can be done by importing error mean square and degree of freedom for error from the individual analysis of variance as follows.

Table 4.1: Error mean squares and their logarithms of each of the separate ANOVA model

Environment	Error mean square (MSE)	log(MSE)
1	217170.16	5.33680015
2	133636.19	5.12592409
3	255759.47	5.407831172
4	191435.24	5.28202189
5	123295.46	5.09094709
6	197121.67	5.29473437
7	281604.49	5.44963958
8	198605.44	5.29799114
Total	1598628.12	42.28589

Where log refers to logarithm base 10.

For our data set, the computed chi-square value ($\chi^2_{cal} = 15.667$) is smaller than the corresponding tabular chi-square value (χ^2_{tab}) with (8-1) =7 d.f. and at the 1% level of significance of 18.47. Thus, the hypothesis that the eight error variances are homogenous cannot be rejected. Therefore, a combined analysis of variance (ANOVA) was performed on the original (untransformed) yield data for the complete set of trials. The p-value based on the Shapiro-Wilk W test of normality for the combined ANOVA is 0.997695, which is very large, indicating the residuals in the combined ANOVA model are normally distributed. The p-value based on the Bartlett's test for the combined data is 0.0661, indicating constant error variance. The plot of residuals versus the fitted values for the combined ANOVA model is shown in fig 2 of Appendix III. There is no severe indication of dependency between the size of the residuals and the fitted values. Likewise, all assumptions of the LR model for the yield data of each genotype are met and as an illustration, SAS output of the separate LR for each genotype are presented in Table 4 of Appendix III.

4.2 ANALYSIS OF VARIANCE AND ESTIMATION OF VARIANCE COMPONENTS

The relative performance of genotypes based on the mean grain yield over environments is presented in Table 4.2. The first three hybrids with highest mean grain yield were G_{19}, G_1, and G_{16} respectively. The hybrid with the lowest mean grain yield was G_{20}. Means across environments are adequate indicators of genotypic performance only in the absence of GxE. If GxE is present, means across environments does not tell us how genotypes differ in relative performance over environments.

Table 4.2: Mean grain yield (kg/ha) of 20 bread wheat genotypes over 8 test environments.

Genotype	Mean grain yield	Rank
G_1	3670.45	2
G_2	3331.55	7
G_3	3411.16	5
G_4	3302.92	8
G_5	3139.42	15
G_6	3169.59	13
G_7	3165.52	14
G_8	3183.89	12
G_9	3384.82	6
G_{10}	3064.68	16
G_{11}	2995.02	17
G_{12}	2923.38	19
G_{13}	3290.64	9
G_{14}	3227.03	10
G_{15}	3448.73	4
G_{16}	3567.6	3
G_{17}	3193.88	11
G_{18}	2959.99	18
G_{19}	3686.3	1
G_{20}	2774.02	20

The ranking of genotypes according to yield indicated that G_{19} was at the top followed by G_1, G_{16}, and G_{15}. A significant GXE interaction may be a non-cross-over type when the ranking of genotypes remains constant across environments and the interaction is significant, because of change in magnitude of response of genotypes (Blum, 1983; Baker, 1988; & Matus et al., 1997). It may be a cross-over GXE interaction in which case a significant change in rank occurs from one environment to another. In the present investigation, the interaction is of cross-over type as the ranking of genotypes changed at every location (Matus et al., 1997). For example, the genotype G1 ranked 7th in environment E1 but it is ranked 2nd in environment E2 for mean grain yield. In general, the ranking of genotypes changes from one environment to another and hence the interaction is of cross-over type.

The combined analysis of variance (Table 4.3) revealed that there were significant differences among environments ($p<0.01$) and genotypes ($p<0.01$) for grain yield indicating the presence of variability in genotypes as well as diversity of growing conditions at different locations. The GXE interaction was highly significant ($p<0.01$) reflecting the differential response of genotypes in various environments (Zubair et al., 2001). The total variation explained (%) is 80.91% for environment, 3.37% for genotype and 4.6% for GXE. The high percentage of the environment is an indication that the major factor that influence yield performance of bread wheat genotypes in Ethiopia is the environment. The relatively large proportion of Genotype x Environment variance, when compared to that of genotypes as a main effect, is a very important consequence.

Table 4.3: Combined ANOVA for yield and the percentage sum of squares of the 20 hybrids tested at 8 environments over a period of two years.

Source	**DF**	**SS**	**%SS**	**MS**	**F-val**	**Pr>F**
Environment (E)	7	854301285.3	80.9	122043040.8	610.7	<0.0001
Replication within E (R/E)	24	26262837.9	2.49	1094284.9	5.48	<0.0001
Genotype (G)	19	35607378.1	3.37	1874072.5	9.38	<0.0001
G x E	133	48600223.2	4.6	365415.2	1.83	<0.0001
Error	456	91121803	8.63	199829.6		
Total	639	1055893527				

The restricted maximum likelihood (REML) estimates of variance components for environment, genotype and genotype x environment interaction are shown in Table 4.4. Estimated variance component due to environment ($\hat{\sigma}_E^2 = 1509789.6$) made the greater contribution to the total estimated variance for grain yield. Genotype X Environment interaction and residual components of variance were 41396.7 and 199828.6 respectively.

Table 4.4: Estimates of variance components for grain yield.

Variance component	Yield	% variance component
Environment($\hat{\sigma}_E^2$)	1509789.6	
Replication within environment($\hat{\sigma}^2{}_{R/E}$)	44722.8	
Genotype($\hat{\sigma}_G^2$)	47145.5	16.349
Genotype X Environment($\hat{\sigma}_{GE}^2$)	41396.7	14.355
Pooled error($\hat{\sigma}_e^2$)	199828.6	69.296

When individual estimates of variance for grain yield (Table 4.4) were expressed as a percent of the total variation ($\hat{\sigma}_G^2 + \hat{\sigma}_{GE}^2 + \hat{\sigma}_e^2$), the $\hat{\sigma}_G^2$ component accounted for 16.349% of the total variation. The $\hat{\sigma}_{GE}^2$ was 14.355% of the total variation, indicating that the genotypes were less consistent over environments. This means that location selection needs more effort. All the components were highly significant (p<0.01), and the importance of the $\hat{\sigma}_{GE}^2$ component indicates that factors such as rainfall, temperature, and disease incidence can result in conditions unique to each year location combination and that the genotypes respond differently to these conditions.

The GEI is highly significant (p<0.01) accounting for 4.6% of the sum of squares implying the need for investigating the nature of differential response of the genotypes to environments. Presence of the GEI indicates that the phenotypic expression of one genotype might be superior to another genotype in one environment but inferior in a different environment. In other words, when significant GXE interactions are present, the effects of genotypes and environments are statistically nonadditive (or the differences between genotypes depend on the environment). The presence of a significant GXE interaction complicates interpretation of the results. That means, it is difficult to identify superior genotypes across environments when GXE interaction is highly significant.

From the combined ANOVA in Table 4.3, GXE interaction is highly significant and hence superiority of genotypes across environments cannot be identified by considering their mean yield performance (see Table 4.2). Furthermore, the traditional analysis of variance determines the values of each variance source and the significance of the contribution of each component, but it does not partition the interaction in to several components and thus other types of analyses should be performed. Hence, such multi-location trial data along with a highly significant GXE interaction requires measures of stability analysis.

4.3 STABILITY ANALYSIS

4.3.1 THE ENVIRONMENTAL VARIANCE

The environmental variance ($S^2{}_i$) is one of the major stability measures for the static stability concept, i.e., the variance of genotype yields recorded across test environments. The smaller the $S^2{}_i$, the more stable the ith genotype.

Table 4.5: Genotype mean grain yield, environmental variance (S_i^2), and coefficient of variation (CV_i) for the 20 bread wheat varieties.

Genotype	Mean yield	Rank	Environmental variance (S_i^2)	Rank	CV_i
G_1	3670.45	2	2316730.73	20	41.4685
G_2	3331.55	7	1399332.52	5	35.5070
G_3	3411.16	5	1962213.21	17	41.0649
G_4	3302.92	8	1474673.44	8	36.7663
G_5	3139.42	15	1635147.28	12	40.7313
G_6	3169.59	13	1450959.54	7	38.0036
G_7	3165.52	14	1840385.54	16	42.8558
G_8	3183.89	12	2270522.53	19	47.3265
G_9	3384.82	6	1318120.22	3	33.9189
G_{11}	3064.68	16	1779333.80	14	43.5255
G_{10}	2995.02	17	1567800.39	10	41.8067
G_{12}	2923.38	19	1327215.83	4	39.4080
G_{13}	3290.64	9	1511515.57	9	37.3616
G_{14}	3227.03	10	1787695.85	15	41.4328
G_{15}	3448.73	4	1434902.21	6	34.7338
G_{16}	3567.6	3	2047855.23	18	40.1120
G_{17}	3193.88	11	1660490.92	13	40.3460
G_{18}	2959.99	18	1275156.29	2	38.1497
G_{19}	3686.30	1	1584762.13	11	34.1501
G_{20}	2774.02	20	601669.2	1	27.9621

Genotype's variance across environments and coefficient of variation are listed in Table 4.5. Based on these two measures, genotypes G_{20}, G_{18} and G_9 can be considered relatively more stable. Among these genotypes, G_9 have mean grain yield of 3384.82 which is slightly above the grand mean, while genotypes G_{18} and G_{20} have respectively mean yield of 2959.99 and 2774.02 which are less than the grand mean. Relatively, genotypes G_1, G_8 and G_{16} can be regarded as unstable genotypes (see Table 4.5.). A problem with this method is that, in general, genotypes with high phenotypic stability measured through the environmental variance show low yield. For instance, according to the environmental variance G_{20} and G_{18} were the most stable genotypes with low mean yield of 2774.02 and 2959.99 respectively. These hybrids were not the best ranked for mean yield, being 20^{th} and 18^{th} respectively. In consequence, plant breeders do not use this method to evaluate yield stability across environments.

4.3.2 WRICKE'S ECOVALENCE ANALYSIS (W_i)

Wricke (1962) defined the concept of ecovalence, to describe the stability of a genotype, as the contribution of genotype to the genotype x environment interaction sum of squares. The ecovalence (W_i) or the stability of the i-th genotype is its interaction with environments, squared and summed across environments. Genotypes with low ecovalence have smaller fluctuations across environments and therefore, are stable. Wricke's ecovalence was determined for each of the 20 genotypes evaluated at 8 environments (Table 4.6).

The amount of interaction (ecovalence) contributed by each genotypes given in Table 4.6. The most interactive genotype was G_{20} followed by genotypes G_{16} and G_8. According to the ecovalence method of wrick's (1962), the most stable hybrids were G_6, G_4, G_{19}, G_{15} and G_9. These hybrids were ranked for mean yield 13^{th}, 8^{th}, 1^{st}, 4^{th} and 6^{th} respectively.

The most unstable hybrids based on the ecovalence method were G_{20}, G_{16}, G_{18} and G_3. These hybrids were ranked for mean yield 20^{th}, 3^{rd}, 12^{th} and 5^{th} respectively.

Table 4.6: Wricke's ecovalence value for 20 the hybrids at 8 environments.

Genotype	W_i	Rank	$\%SS_{GXE}$	Mean yield	Rank
1	793111.69	15	6.528	3670.45	2
2	847927.04	16	6.980	3331.55	7
3	946220.52	17	7.788	3411.16	5
4	145500.01	2	1.196	3302.92	8
5	278890.54	8	2.295	3139.42	15
6	132771.02	1	1.093	3169.59	13
7	297744.14	9	2.451	3165.52	14
8	1285750.06	18	10.582	3183.89	12
9	226399.83	5	1.863	3384.82	6
10	258444.67	7	2.127	3064.68	16
11	323671.65	10	2.664	2995.02	17
12	694911.83	14	5.719	2923.38	19
13	251943.38	6	2.074	3290.64	9
14	451199.70	13	3.714	3227.03	10
15	153224.50	4	1.261	3448.73	4
16	1378648.87	19	11.347	3567.6	3
17	405511.8	12	3.337	3193.88	11
18	329026.55	11	2.708	2959.99	18
19	145799.47	3	1.196	3686.3	1
20	2803358.51	20	23.072	2774.02	20

4.3.3 SHUKLA'S STABILITY VARIANCE

Shukla (1972) developed a modified version of the ecovalence in order to give unbiased estimate of the GXE variance for every genotype using the stability variance (σ_i^2). A genotype is called stable if its stability variance is equal to the environmental variance which means that $\sigma_i^2 = 0$. A relatively large value of σ_i^2 will thus indicate greater instability of genotype i.

According to this stability parameter the most stable hybrid were G_6, G_4 and G_{19}. Similarly, the most unstable hybrids were G_{20}, G_{16} and G_{18} (Table 4.7).

Table 4.7: Genotype mean grain yield and Shukla's stability variance (σ_i^2) for the 20 Bread wheat varieties.

Genotype	Stability variance($\hat{\sigma}_i^2$)	Rank	$\%SS_{GXE}$	Mean yield	Rank
1	120815.53	15	6.528	3670.45	2
2	129516.38	16	6.980	3331.55	7
3	145118.52	17	7.788	3411.16	5
4	18020.03	2	1.196	3302.92	8
5	39193.13	8	2.295	3139.42	15
6	15999.55	1	1.093	3169.59	13
7	42185.76	9	2.451	3165.52	14
8	1999012.10	18	10.582	3183.89	12
9	30861.27	5	1.863	3384.82	6
10	35947.75	7	2.127	3064.68	16
11	46301.24	10	2.664	2995.02	17
12	105228.25	14	5.719	2923.38	19
13	34915.8	6	2.074	3290.64	9
14	66543.79	13	3.714	3227.03	10
15	19246.14	4	1.261	3448.73	4
16	213757.94	19	11.347	3567.6	3
17	59291.74	12	3.337	3193.88	11
18	47151.23	11	2.708	2959.99	18
19	18067.56	3	1.196	3686.3	1
20	439902.33	20	23.072	2774.02	20

From the results of this analysis it can be seen that the stability estimates by Wrick (1962) and Shukla (1972) are identical for ranking purpose.

4.3.4 REGRESSION ANALYSIS

The combined analysis of variance (ANOVA) in Table 4.3 tells us there is a significant GXE interaction effect and hence the interaction cannot be ignored. One way of partitioning the GXE interaction is the linear regression of the genotype performance on the environment mean. This model partitioned the GXE interaction into two components: heterogeneity of genotype regressions and the deviations from regressions.

The regression of each genotype's yields on the environmental means partitioned the GXE interaction in this bread wheat data set's with a sum of squares(SS) of 48600223.2 and 133 degrees of freedom into: heterogeneity of genotype regressions with 19 DF and a SS of

14786776.6 and deviations from regressions with 114 DF and a SS of 33813446.6 (Table 4.8). Only 30.425% of the GXE interaction sum of squares accounted for by regression SS, and the remaining 69.575 % was accounted for by the SS of the regression residual. This indicates that heterogeneity of slopes generally explains only a small proportion of the interaction.

Table 4.8: Joint regression partitioning of the GXE interaction for grain yield (kg/ha) of 20 genotypes grown in eight environments.

Source	DF	Sum of squares	MS	F-value
Environment (E)	7	854301285.3	122043040.8	610.14
Replication within E	24	26262837.9	1094284.9	5.48
Genotype(G)	19	35607378.1	1874072.5	9.38
GXE	133	48600223.2	365415.2	1.83
Regression	19	14786776.6	778251.4	2.624
Residual	114	33813446.6	296609.18	1.48
Error	456	91121803	199829.6	
Total	639	1055893527		

The genotype regressions term was tested for significance using an F-ratio by taking the deviations from regressions MS (or residual MS) as the error term. The deviations from regressions MS were tested for significance using the error term for overall GXE interaction in the ANOVA.

The analysis of variance for the regression model (Table 4.8) shows that the regression coefficients are homogenous among genotypes. Moreover, the residual MS is substantially larger than the error MS which indicates the heterogeneity of residuals among genotypes. Therefore, stability for these bread wheat genotypes was evaluated by considering deviation for regression.

According to the joint linear regression model which was developed by Finlay and Wilkinson (1963) and modified by Eberhart and Russel (1966), a stable variety is one with a high mean yield, regression coefficient equals to one (b_i=1) and deviation from regression equals to zero (S^2_{di}=0). A genotype with b_i value less than 1.0 has above average stability and is especially adaptable to low-performing environments. A genotype with b_i value greater than 1.0 has below average stability and especially adaptable to high performing environments and a genotype with b_i value equal to 1.0 has average and is well or poorly adaptable to all environments depending on high or low mean performance (Finlay and Wilkinson, 1963). A cultivar with b_i =1 and S^2_{di}=0 may be defined as stable. However, in most cases, S^2_{di} is considered as stability parameter rather than b_i which is more about responsiveness of genotypes (Eberhart and Russel, 1966; Becker and Le'on, 1988).

Table 4.9: mean yield (kg/ha) and stability parameters of bread wheat genotypes tested in eight environments.

Genotype	SS	F-Ratio	Pr.>F	Beta	Deviation	Rank	Mean yield	Rank
G_1	15951154	359.85	0.0001	1.22218	44326.83	10	3670.45	2
G_2	9017578	69.57	0.0002	0.91893	129624.96	16	3331.55	7
G_3	12893550	91.88	0.0001	1.09882	140323.78	18	3411.16	5
G_4	10183103	437.64	0.0001	0.97652	23268.48	3	3302.92	8
G_5	11172724	245.28	0.0001	1.02287	45551.13	11	3139.42	15
G_6	10033984	490.53	0.0001	0.96934	20455.44	1	3169.59	13
G_7	12670020	357.44	0.0001	1.08925	35446.53	7	3165.52	14
G_8	14969329	97.17	0.0001	1.18397	154054.72	19	3183.89	12
G_9	9066385	339.02	0.0001	0.92142	26742.75	5	3384.82	6
G_{10}	12250847	359.46	0.0001	1.07108	34081.55	6	3064.68	16
G_{11}	10650949	197.45	0.0001	0.9987	53942.25	12	2995.02	17
G_{12}	8697193	87.95	0.0001	0.90246	98886.34	15	2923.38	19
G_{13}	10331535	248.88	0.0001	0.98361	41512.32	9	3290.64	9
G_{14}	12107508	178.77	0.0001	1.0648	67727.2	14	3227.03	10
G_{15}	9905616	428.51	0.0001	0.96312	23116.6	2	3448.73	4
G_{16}	13077778	62.41	0.0002	1.10664	209534.75	20	3567.6	3
G_{17}	11224730	168.92	0.0001	1.02524	66451.07	13	3193.88	11
G_{18}	8698518	229.33	0.0001	0.90253	37929.32	8	2959.99	18
G_{19}	10949227	455.87	0.0001	1.01258	24018.06	4	3686.3	1
G_{20}	3420287	25.93	0.00022	0.56594	131899.52	17	2774.02	20

According to the S^2_{di} values, G_6, G_{15}, and G_4 can be regarded as more stable varieties. Of these, genotype G_{15} can be considered best, judging from its mean yield (3448.73) and deviations from regression (23116.6). On the contrary, G_{16}, G_8 and G_3 can be classified as unstable varieties (Table 4.9).

There is considerable disagreement in assessing yield stability among the methods presented earlier (the biological/static concept and agronomic/dynamic concept). Results from a simple variance or CV of the genotype across environments identified G_{20}, G_{18} and G_9 as a more stable varieties. Based on the contribution of genotypes to GEI, varieties G_6, G_4 and G_{19} were classified as more stable. The slopes resulting from regressing on environmental mean proved unfruitful to detect any difference in stability pattern for the varieties analyzed in this study. Finally, varieties G_6, G_{15} and G_4 were considered more stable from Eberhart and Russell's deviation from regression. In fact, environmental variance is the biological or static concept of stability measures in which stable genotype is usually associated with relatively poor yield. Accordingly, these measures are rarely useful to the breeder who is always looking for high yield stability. All the remaining methods (i.e. Wricke's ecovalence, Shukla's stability variance and the regression analysis) are agronomic/dynamic concept of stability measures. Most of the yield stability statistics in vogue are measured according to the agronomic concept and hence, the agronomic concept of stability measures would be used to determine the stable genotype over this study.

4.4 ADDITIVE MAIN EFFECTS AND MULTIPLICATIVE INTERACTION (AMMI) ANALYSIS AND BIPLOT REPRESENTATION

Multivariate techniques are widely applied in stability analysis to provide further information on real multivariate response of genotypes to environments. Among the multivariate analysis techniques, the AMMI model is the powerful method in assessing G x E interaction and stability/adaptation of genotypes from multi-environment trials. AMMI is essentially effective where the assumption of linearity of responses of genotype to a change in environment is not full filled, which is important in stability analysis. The results can be graphed in a useful biplot that shows both main and interaction effects for both genotypes and environments (Gauch and Zobel, 1996).

The combined analysis of variance (ANOVA) of the 20 hybrids over two years and 4 locations according to the AMMI-2 model are presented in Table 4.10. The ANOVA indicated highly significant differences ($p<0.01$) for environments, genotypes and genotype x environment interaction (GEI). The IPCA are ordered according to decreasing importance.

Table 4.10: Combined analysis of variance (ANOVA) according to the AMMI model and Gollob tests of interaction PCs.

Source	DF	SS	MS	Total variation explained (%)	GXE explained(%)	Cumulative (%)
Total	639	1055893527	-			
Environment(E)	7	854301285.3	122043040.8**	80.9		
Reps within E	24	26262837.9	1094284.9			
Genotype(G)	19	35607378.1	1874072.5 **	3.4		
GXE	133	48600223.2	365415.2 **	4.6		
IPCA1	25	20621032.7	824841.3 **		42.43	42.43
IPCA2	23	9664903.9	420213.2 **		19.89	62.32
IPCA3	21	5394577.3	256884.63 **		11.10	73.42
IPCA4	19	4478174.1	235693.4*		9.21	82.63
IPCA residual	45	8441535.3	187589.67			
Residual	456	91121803	199828.52			

**p<0.05, *p<0.01; IPCA=Interaction Principal Component Axis

The total variation explained (%) was 80.9% for environment, 3.4% for genotype and 4.6% for GXE. As mentioned earlier, the high percentage of the environment is an indication that the major factor that influence yield performance of bread wheat in Ethiopia is the environment. The Gollob F-test used to measure significant of the GXE interaction components at 0.01 probability level recommended inclusion of the first four interactions PCA axes in the model. Hence, the best fit AMMI model for this multi-environment yield trial data was AMMI-4. Out of the total IPCA, the first four IPCA axes explained 82.63% of the GXE interaction sum of squares.

The criterion of postdictive success of the AMMI model identified the first four IPCA axes in the model and four principal component axes of the interaction were significant for the AMMI model. However, to simplify the complexity of the analysis and to graph the results of AMMI using a biplot, two interaction principal component axes for AMMI model were sufficient for predictive model. Other interaction principal component axes captured mostly non-predictive random variation (noise) and did not fit to predict validation observations. Therefore, the interaction of the 20 bread wheat genotypes with eight environments was predicted by the first two interaction principal components of genotypes and environments. In general, the model chosen by predictive criterion consists of two interaction principal components (Kaya et al., 2002).

Furthermore, the first two IPCA axes explained 62.32% of the GXE interaction, indicating that with only 48 DF from the 133 contained in the analysis of variance for genotype x environment interaction, a considerable amount of GXE was explained. In particular, the first IPCA captured 42.43% of the total interaction sum of squares while the second IPCA explained 19.89% of the interaction sum of squares. This partitions the treatment sum of squares (the sum of sum of squares of genotype, environment and genotype x environment interaction) and DF in to a model with a sum of squares of 18314286.7 in 85 degrees of freedom. Overall, this model contains 98.05% of the treatment sum of squares.

Estimates for the genotypic and environmental scores of AMMI-2 (i.e. of interaction PCA axis 1 and axis 2) are given in table 4.11 and table 4.12, respectively. The PCA scores of a genotype from AMMI analysis indicate the stability or adaptation of a genotype across environments. The larger the PCA score, either positive or negative, as it is a relative value, the more specifically adapted a genotype is to certain environments. The closer the PCA scores near zero, the more stable or adapted a genotype is over all test environments. Environment scores from AMMI analysis relating to interaction also have meaningful interpretation. Environments with large PCA scores are more discriminating of genotypes, while environments with PCA scores near zero exhibit little interaction across genotypes and low discrimination among genotypes.

Table 4.11: IPCA1 and IPCA2 scores for the 20 bread wheat hybrids.

Genotype	mean yield	IPCA1-Score	IPCA2-Score
G_1	3670.45	-10.1905	-7.9811
G_2	3331.55	11.2861	5.5439
G_3	3411.16	4.0044	5.4259
G_4	3302.92	0.9911	-5.7534
G_5	3139.42	9.9377	1.1395
G_6	3169.59	20.1193	8.1098
G_7	3165.52	6.9843	-6.8509
G_8	3183.89	-4.3044	1.3967
G_9	3384.82	9.3305	5.0434
G_{10}	3064.68	-16.6915	-16.4580
G_{11}	2995.02	1.0621	11.0328
G_{12}	2923.38	5.6825	-6.7798
G_{13}	3290.64	18.5900	-21.2657
G_{14}	3227.03	-1.9551	5.3733
G_{15}	3448.73	22.8552	-8.4763
G_{16}	3567.6	4.1265	9.1943
G_{17}	3193.88	-0.6187	1.4154
G_{18}	2959.99	-5.8774	-0.3722
G_{19}	3686.3	10.6384	-5.8551
G_{20}	2774.02	3.9296	11.9832

Table 4.12: IPCA1 and IPCA2 scores for the eight environments

Environment	Env.Mean	IPCA1-Score	IPCA2-Score
E_1	2160.05	15.7642	5.5311
E_2	2487.9	-9.7796	-10.6584
E_3	4440.37	-4.8651	9.9231
E_4	4083.35	-24.1981	29.0094
E_5	5409.12	-6.1074	1.8694
E_6	2075.67	5.4532	-10.9825
E_7	2512.72	-4.3525	3.1026
E_8	2787.05	35.0880	18.3434

Genotype and Environment combinations with PCA scores of the same signs produce positive specific interaction effect, whereas combination of opposite signs have negative specific interactions. For example, E1 and G2 have positive specific interaction effect while E2 and G3 have negative specific interaction effect. Environment which have same signs of interaction PCA scores discriminate genotypes similarly, for instance E1 and E6 or E2 and

E3; and Environments with opposite sign of interaction scores discriminate genotypes differently, for example E1 and E2.

Note that the biplot captures 62.32% of the interaction SS. Because the GXE component of the AMMI model is based on the product of interaction PCA scores, it follows that genotypes or environments with small interactions (smaller scores) will appear close to the center of the axes. Genotypes G_{17}, $G_{8,}$ G4, G3 and G_{14} exhibit this trait, and therefore are relatively more stable. Conversely, genotypes such as G_{13}, G_{15}, G_{10} and G_6 are relatively far apart from the origin and thus show strong interaction effects (see fig 4.1).

Among the eight location-year environments E8, E4, E1 and E2 exhibited larger interactions (i.e they are relatively far apart from the origin) and were more discriminating of genotypes, whereas the environment E7 relatively exhibited negligible interaction and low discrimination. The nearly additive behavior of E7 indicates that genotypic yields in that environment were highly correlated with the overall genotypic means across environments (see fig 4.1).

The direction of the genotypes and environments from the origin (center) also contain important information on the interaction. As an example, genotype G13 and environment E4 appear opposite from each other indicating their contributions to the interaction was in opposing directions (i.e. they are negatively correlated). By contrast, genotype G6 and environment E1 both have the same relative direction, so that both contribute positively to the interaction. Indeed, the best genotypes with respect to E1 were G6 and G2. For environments E5 and E7 the best genotypes were G8 and G18. Similarly G14 was best for E3; G1 was best for E2; G7 and G were best genotypes for E6 and G6 was best for E8. Thus, the biplot can give information on relative stability, as well as suggesting trends of similar or dissimilar genotypes and environments (see fig 4.1).

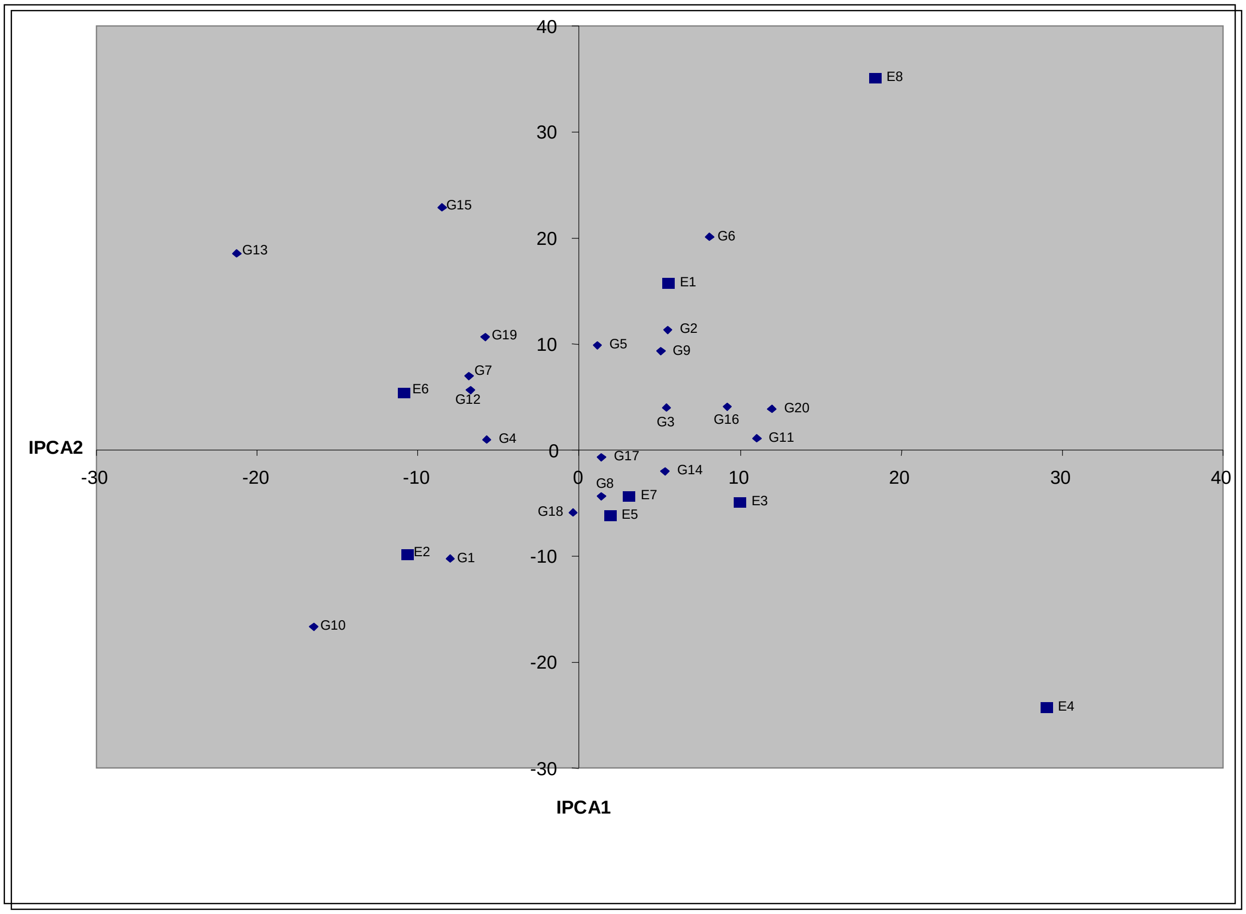


Fig 4.1: Biplot of interaction principal components analysis (PCA) axis 1 versus axis 2 for grain yield (kg/ha) for 20 bread wheat hybrids grown in 8 environment.

The second biplot (Fig 4.2) is of interaction PCA axis 1 versus mean yield of both genotypes and environments. From the biplot, environments are distributed from lower yielding environments in quadrants II(top left) and III(bottom left) to the high yielding environments in quadrants I (top right) and IV (bottom right)(fig 4.2). The high yielding environments classified according to the AMMI 1 model are E3, E4 and E5. The lower yielding environments are E1, E2, E6, E7 and E8. It is further noted that E5 was the most favorable season and E6 was the less favorable season among the eight environments, this situation is clearly indicated in fig 4.2 where the two environmental variations are plotted far apart from the mean.

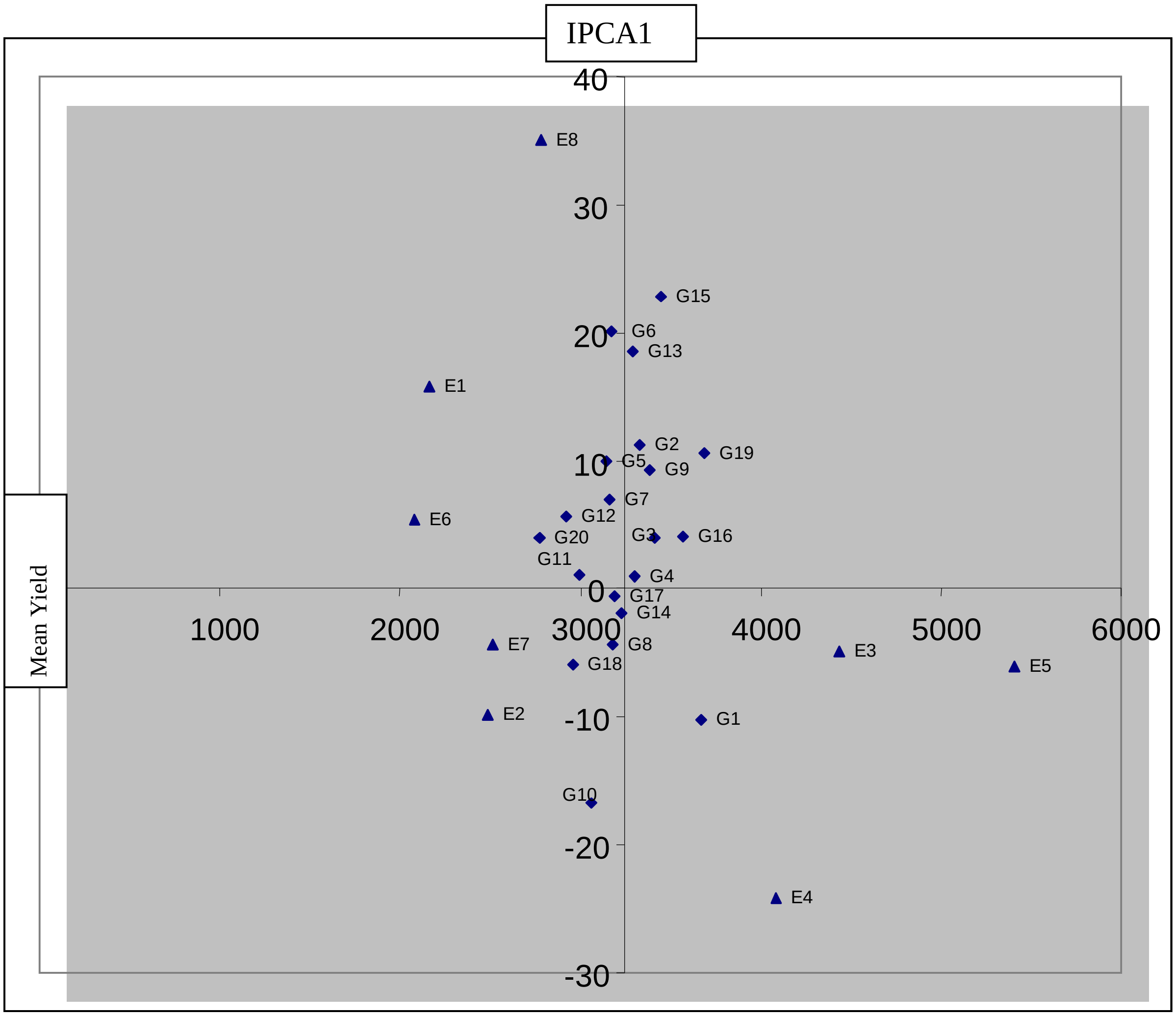


Fig 4.2: Biplot of interaction principal components analysis (PCA) axis 1 mean yield (kg/ha) for 20 bread wheat hybrids grown in 8 environments. The vertical line represents the grand mean of the experiment while the horizontal line is PCA axis 1=0.

The genotypes categorized under favorable environments with above average means are G1, G2, G4, G6, G9, G13, G15, G16 and G19 among them G4 is found to be more stable. Genotypes grouped under low yielding environments are shown at the lower left quadrant of the biplot. Generally G15 is the most unstable genotype identified by the AMMI model (fig 4.2). Genotypes that are close to each other tend to have similar performance and those that are close to environment indicates their better adaptation to that particular environment. Hence, genotypes that have similar performance with G2 were G5, G9 and G19, genotypes that have similar performance with G17 were G11, G4 and G14, and hybrids G8 and G18 showed similar performance as they are close to each other. Genotypes G1, G2 and G3 were relatively adapted to environments E3, E1 and E6 respectively. Hybrid G18 was more adapted to the environments E7 and E8. Similarly, the genotypes G19 and G20 were more adapted to E1 and E6 respectively.

In summary, interaction patterns revealed by AMMI model biplot analysis indicate that genotypes G17, G8, G4 and G18 exhibit smaller interactions with environments and are therefore more stable as observed across both interaction axes (see fig 4.1). Because of its moderately high yield and stable performance across environments, G4 was identified as the most desirable genotype among the 20 genotypes tested.

5 CONCLUSIONS AND RECOMENDATIONS

The objective of the study were: (1) to evaluate the adaptability of 20 bread wheat genotypes and to identify the best performing ones for future use; (2) to utilize some of the statistical procedures for analyzing genotype by environment interaction and yield stability of Ethiopian bread wheat hybrids across eight environments; and (3) to provide efficient statistical methods that guide breeders for releasing genotypes with adaptation to target environments in Ethiopia. Twenty bread wheat hybrids were evaluated for grain yield in a mid altitude areas of Ethiopia, for a period of two years across four locations. The genotypes were planted in a completely randomized block design.

From the combined analysis of variance, the effects environment, genotype and environment x environment were highly significant for grain yield and accounted for 80.91%, 3.37% and 4.6% of the total sum of squares. The high percentage of the environment is an indication that the major factor that influence yield performance of bread wheat in Ethiopia is the environment. In particular, the GEI is highly significant ($p<0.01$) accounting for 4.6% of the total sum of squares implying the need for investigating the nature of differential response of the genotypes to environments. The presence of the GEI indicates that the phenotypic expression of one genotype might be superior to another genotype in one environment but inferior in a different environment. In other words, presence of GEI does not permit to define an overall ranking of varieties across environments. Among the restricted maximum likelihood (REML) estimates of variance components, the estimated variance component due to environment made the greater contribution to the total estimated variance for grain yield. All of the variance components were highly significant ($p<0.01$), and the importance of the $\hat{\sigma}^2_{GE}$ component indicates that factors such as rainfall, temperature, and disease incidence can result in conditions unique to each year-location combination and that the genotypes respond differently to these conditions.

The selection process of good performing and stable genotypes is mainly complicated by the phenomenon of genotype by environment (GXE) interaction. GXE interaction is a differential genotypic expression across environments or generally the inconsistency of relative performance of genotypes over environments. The large occurrence of GXE interactions causes the relative rankings of genotypes to change from location to location and/or from year to year. Hence, it is imperative to have a proper understanding of the effects of GXE interactions on variety evaluation, which will help to apply appropriate analytical methods and wise application of resources.

The different stability measurements used in this study demonstrated association and dissociation among them in ranking of the genotypes based on stability. Eberhart and Russel's deviation from regression (S_{di}^2) procedure showed highly correlated with the parameters of Wricke (W_i) and Shukla ($\hat{\sigma}_i^2$), but it was quite poorly correlated with environmental variance. The perfect correlation between Wrick's and Shukla indicates that the two procedures are equivalent for ranking purposes.

The regression of each genotype's yield on the environmental means partitioned the GXE interaction into two components: heterogeneity of genotype regressions and the deviations from regressions. In the analysis of variance for the regression model, the residual MS is substantially larger than the error MS which indicates the heterogeneity of residuals among genotypes. Thus, stability for the bread wheat genotypes was evaluated by considering deviations for regression.

The AMMI model provides a useful technique in diagnosing genotype x environment interaction patterns. It enables clustering of genotypes based on similarity of response characteristics and identifying potential trends across environments. The number of PCA axes retained for most applications is usually $s \leq 3$, which is intended to reduce the dimension of the system and provide a more parsimonious description of the underlying interaction structure. The AMMI model provides easily interpretable information as well as the correlation between a genotype and environment.

The analysis of variance for the AMMI model indicated highly significance differences between genotypes and environments as main effects and the interaction effect of GXE was also highly significant. The first two interaction principal component axes(IPCA) of the AMMI model together accounted for 62.32% of the GXE interaction sum of squares for grain yield. The first four IPCA axes were highly significant and hence, the AMMI-4 model was used as the best fit for the bread wheat data.

The following major findings emerged for the multi-environment yield trial data set analyzed here.

- Significant variation existed for environment, genotype and genotype x environment interaction.
- Genotypes G_6, G_4 and G_9 were more stable while G10, G_{20}, G_{16}, G_{18} and G_3 were unstable varieties.
- The AMMI model is better than other statistical analyses in determining the stable and adaptability of genotypes across environments.
- The first four IPCA axes were significant and hence AMMI-4 model was the best fit for AMMI model.
- Genotypes that have similar performance with G2 were G5, G9 and G19, with that of G17 were G11, G4 and G14, and similarly hybrids G8 and G18 showed similar performance as they are close to each other. Genotypes G1, G2 and G3 were relatively adapted to environments E3, E1 and E6 respectively. Hybrid G18 was more adapted to the environments E7 and E8. Similarly, the genotypes G19 and G20 were more adapted to E1 and E6 respectively.
- Genotype G4 was the most desirable among the 20 genotypes tested.

From this study the multivariate analysis (AMMI) procedure is recommended to identify genotypes according to their adaptation environment. The categorization of genotypes according to their response to the different environments has an agronomic advantage for breeders and farmers. With AMMI it is also possible to target the different production environments by developing genotypes for specific adaptation.

Reference

- Admassu, S., Nugussie, M., Zelleke, Zelleke, H., 2008. Genotype X environment interaction and stability analysis for grain yield (Zea mays L.) in Ethiopia. Asian J. Plant Sci., 7(2): 163-169.
- Adugna Wakjira and Elias Urage ., 1994. Stability of pod yield in groundnut. Sebil, 6: 24-29.
- Allard, R.W. and Bradshaw, A.D., 1964. Implications of genotype-environment 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.
- Ashraf, M, A.S. Qureshi, A.Ghafoor and N.A.Khan, 2001. Genotype by environment interaction in wheat. Asian J.Biol.Sci., 1:356-357.
- Asrat Asfaw and Daniel Dauro ., 2004. Genotype X environment interaction and correlation among some stability parameters of yield and its attributes in blackgram [Vigna mungo (L.) Hepper]. Sebil, 11: 23-32.
- Basford, K.E. and Cooper, M., 1998. Genotype x environment interactions and some consideration of their implications for wheat breeding in Australia. Australian Journal of Agricultural Research 49:154-174.
- Becker, H.C., 1981. Correlations among some statistical measures of phenotypic stability. Euphytica 30:832-840.
- Becker, H.C and Leon, J., 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 in two-way tables. Technometrics, 20: 47-68.
- Carvalho, F.I.F., Federizzi, L.C.Nodari, R.O. and Storck, L. (1993). Comparisons among stability models in evaluating genotypes. Rev. Brasil.Genet.6:667-691.

- Chahal, G.S. and Gosal, S.S., 2002. Principles and procedures of plant breeding: Biotechnological and conventional approaches. Narosa publishing House. New Delhi, India.
- Cochran, W.G and Cox, G.M., 1957. In: Experimental designs. 2nd ed. Wiley, new York.
- Comstock, R.E and Moll, R.H., 1963. Genotype –environment interactions. In: statistical Genetics and plant Breeding. NAS-NRC.pp.164-196.
- Cox, D.R, 1984. Interaction Int.Statist. Rev.52:1-31.
- Crossa, J., 1990. Statistical analysis of multiplication trials. Advances in Agronomy 44:55-85.
- Crossa,J,, Gauch, H.G and Zobel, R.W., 1990. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. Crop science 30:493-500.
- CSA, 2002. Central statistics Authority (CSA). 2002. Statistical abstract. 96 pp.
- Dabholkar,A.,1999 Elements of biometrical genetics concept publishing company New Delhi.India .
- Daniel,C. Application of statistics to industrial experimentation. New York; John Willey .1976.
- Delacy, I.H., Cooper, M. and Basford, K.E., 1996. Relationships among analytical methods used to study genotype by environment interactions and evaluation of their impact on response to selection. In; kang, M.S. and Gauch, Jr.H.G.(eds.). Genotype –by-environment interaction, CRC press; Boca Raton, New York. Pp.51-84.
- Dias, C.T.S.; Krzanowski, W.J. Model selection and cross validation in additive main effect and multiplicative interaction (AMMI) model. Crop science, V.13.p. 865-873.2003.
- Eberhart, S.A. and Russell, W.A., 1966. Stability parameters for comparing varieties. Crop science 6:36-40.
- Edmeandes, G.O., Bolanos, J., Lafitte, H.R, Rajaram, S., Pfeiffer, W.H. and Fisher, R.A., 1989. Traditional approaches to breeding for dough resistance in cereals. In:

Baker, F.W.G.(ed.). Drought resistance in cereals. ICSU press CAB International. Mallingford. pp.27-52.

- Falconer, D.S. and Mackay, T.F.C.1996. Introduction to quantitative genetics. 4th edition, Longman, New York, P.132-133.
- Fehr, W.R., 1991. Principle of cultivar development theory and technique. IOWA state university, USA. pp. 247-260
- Fekadu Fufa., 1994. Performance of barely genotypes grown in different environments in North Western Ethiopia. Sebil, 6:20-22.
- Finlay, K.W. and Wilkinson, G.N., 1963. The analysis of adaptation in a plants breeding programme. Australian Journal of Agricultural Research 14:742-754.
- Francis, T.R. and Kannenburg, L.W., 1978. Yield stability studies in short season maize. I.A descriptive method for grouping genotypes. Canadian Journal of plant sciences 58:1029-1034.
- Freeman, G.H., 1990. Modern statistical methods for analyzing genotype x environment interactions. In: Kang, M.S.(ed.). Genotype by environment interaction and plant breeding. Louisiano State University. Agricultural center, Baton Rauge, La.pp.118-125.
- Freeman, G.H., 1973. Statistical methods for the analysis of genotype by environment interactions. Heredity 31:339-354.
- Freeman, G.H., 1985. The analysis and interpretation of interactions. Journal of Applied statistics 12:3-10.
- Gabriel K.R. Le biplot-outil d'exploration de donnees multi dimensionless. Journal de la societe Francaise de statistique. V.143.p.5-55.2002.
- Gauch, H.G and Zobel, R.W., 1988. Predictive and postdictive success of statistical analyses of yield trials. Theor. Apps. Genet. 76:1-10
- Gauch, H.G. and Zobel, R.W., 1997. Identifying mega-environments and targeting genotypes. Crop science 37(2):311-326.
- Gauch, H.G. and Zobel, R.W., 1996. AMMI analysis of yield trials. In: Genotype by environmental Interaction (Kang, M.S., Gauch, H.G., ed.). CRC press, Boca Raton, FL, pp.85-122.

- Gauch, H.G., 1988. Model selection and validation for yield trials with interaction. Biometrics 44:705-715.
- Gauch, H.G, 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam, 278 pp.
- Gollob. H.F. A statistical model which combines features of factor analytic and analysis of variance techniques. Psychometrika.V.33, p.73-115, 1968.
- Girma Taye (1997). Modeling GXE interactions: A review of procedures. Sebil, 8:147-158.
- Girma Taye, Temesgen Getachew and Geletu Bejiga (2000). AMMI adjustment for yield estimate and classification of genotypes and environments in field pea (Pisum sativum L.). Jornal of Genetics and Breeding, 54: 183-191
- Hill, J., 1975. Genotype by environment interactions a challenge for plant breeding. Journal of Agricultural science 85:477-493.
- Kang, M.S., 1990. Genotype by environment interaction and plant breeding. Louisiana state university, Baton Rouge, LA, USA.
- Kang, M.S., 1996. Using genotype by environment interaction for crop cultivar development. Advances in Agronomy 62: 199-252.
- Kaya, Y., Palta, C., Taner, S., 2002. Additive main effects and multiplicative interactions analysis of yield performance in bread wheat genotypes across environments. Turk.J.Agric. For., 26:275-279.
- Kaya, Y., Akcura, M., Taner, S., 2006. GGE –biplot analysis of multi environment yield trials in bread wheat. Turk.J.Agric.For. 30:325-337.
- Krzanowski, W.J. Cross validation in principal component analysis. Biometrics.V.43.p.575-584-1987.
- Letta, T., 2007. Genotype by 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 vol 8.pp.693-698.
- Lin, C.S., Binns, M.R. and Lefkovitch, L.P., 1986. Stability analysis: where do we stand? Crop science 26:894-899.

- Long, M.S, 1996 using genotype-by-environment interaction for crop cultivar development Advances in Agronomy 62:199-252
- Magai, R. and Kang, M.S., 1993. Genotype selection via a new yield stability statistics in maize yield trials. Euphytica 70:105-111
- Magari, R., 1989. Stability of some Albanian Maize local varieties and hybrids (in Albania). Bull Agic.Sci.4:123-129.
- Mather, K. and J.L.Jinks.1982. Biometrical Genetics /The study of continuous variation. Chapman and Hall, London, New York, 396pp.
- Mukai, T. 1988. Genotype by environment interaction in relation to the maintenance of genetic variability in populations of Drosophila melanogaste proceedings of the 2nd International conference on quantitative Genetics (Editors:B.S.Weir, E.J.Eisen, M.M.Goodman, and G.Namkoong): 21-31.
- Mandel.J., 1971. A new analysis of variance model for non additive data. Technometrics.13:1-8.
- Mosisa Worku, Habtamu Zelleke, Girma Taye, Benti Tolessa, Legesse Wolde, Wende Abera, Aschalew Guta and Hadji Tuna., 2001. Genotype x environment interaction and grain yield stability of maize (Zea mays L.) genotypes. Sebil, 10: 64-69.
- Naazar, A., F.Javid and A.A.Attary, 2002. Stability analysis of seed yield in winter type repeseed (Brasica napus) vaieties. Pakistan J.Bot., 34:151-155.
- Nachit, M.N., Sorells, M.E., Zobel, R.W., Gauch, H.G., Fischer, R.A., Coffman, W.R., 1992. Association of environmental variables with sites means grain yield and components of GXE interaction in durum wheat. J.Genet. Breed. , 46:369-372.
- Perkins, J.M. and Jinks, J.L., 1968. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. Heredity 23:339-351.
- Perkins, J.M. and Jinks, J.L., 1971. Specificity of interaction of genotypes with contrasting environments. Heredity 26:463-474.
- Ramagosa, I. and Fox, P.N. 1993. Genotype X environment interaction and adaptation. In: Hayward M.D., Bosemark N.O. and Ramagosa I.(eds.). Plant breeding: principles and prospects. Chapman and Hall, London.pp.373-390.

- Rasul, M.Zulkiffal, J.Anwar, S.B.Khan, M.Hussain and Riazuddin, 2004. Grain yield stability of wheat genotypes under different environment in Punjab in Pakistan. Pakistan J.Bot., 4:222-224.
- Shukla, G.K., 1972. Some statistical aspects of partitioning genotype environmental components of variability. Heredity 29:237-245.
- Sial, M.A., M.A. Arain and M.Ahmad, 2000. Genotype x Environment interaction on Bread wheat grown over multiple sites and year in Pakistan. Pakistan J.Bot., 32:85-91.
- Simmonds, N.W., 1962. Variability in crop plants, its use and conservation. Biol.Rev.37:422-465.
- Skroppa, T., 1984. A critical evaluation of methods available to estimate the genotype x environment interaction. Studia Forestalia Svecica 166;3-14.
- Snee.R.D. Nonadditivity in a two way classification: is it interaction or nonhomogenous variance?Journal of American Statistical Association V.77.p.515-519
- Suzuki,D., Griffiths, A. and Lewontin, R., 1981. An introduction to genetic analysis. W.H. Freeman and Company. San Francisco, USA.
- Tai, G.C.C., 1971. Genotype stability analysis and its application to potato regional trials. Crop science 11:184-190.
- Westcott, B., 1987. A method of assessing the yield stability of crop genotypes. J.Agic.Sci.108:267-274.
- Westcott, B., 1986. Some methods of analyzing genotype-environment interaction. Herdity 56:243-253.
- Williams, E.j., 1952. The interpretation of interactions in factorial experiments. Biometrika, 39: 65-81.
- Wricke, G. and Weber, W.E., 1980. Erweiterte analyse von wechselwirkungen in versuchsserien. In: Biometrie-heute and morgen. kopcke and uberla.(eds.). Springer-velag, Berlin.
- Wu, R.L. and D.M. OiMalley.1998. Non linear genotypic response to macro and micro- environments. Theoretical and applied genetics 96:669-687.
- Zobel, R.W., Wright, M.J. and Gauch , Jr.H.G., 1988. Statistical analysis of yield trial. Agronomy Journal 80:388-393.

APPENDEX

Appendix I

The step-by-step procedures to apply the chi-square test (Bartlett's test) to test for homogeneity of e variances (or error mean squares) with equal degree of freedom (d.f.).

Step1**.** For each variance estimate s^2, compute $\log s^2$, where log refers to logarithm base ten. Then, compute the totals of all e values of s^2 and of $\log s^2$.

Step2. Compute the pooled estimate of variance as:

$$S_p^2 = \frac{\sum_{i=1}^{e} S_i}{e}$$

Step3**.** Let f be the degree of freedom of each S_i^2, compute the χ^2 value as:

$$\chi^2 = \frac{(2.3026)(f)(e \log S_p^2 - \sum_{i=1}^{e} \log S_i^2}{1 + \left[(e+1) / 3ef \right]}$$

Step4**.** Compare the computed χ^2 value with the tabular χ^2 value with (e-1) d.f.; and reject hypothesis of homogeneous variance if the computed χ^2 value exceeds the corresponding tabular χ^2 value at the prescribed level of significance.

Appendix II

Almost all statistical analyses that are available in this study were done using SAS software. The most relevant SAS program statements are presented below, with classification variables for genotypes, environments and blocks labeled as gen, env, and rep, respectively.

A) SAS code for separate analysis

```
Proc sort data=zelalem;
by env;
run;
proc glm;
class rep gen;
model yield=rep gen/ss1;
by env;
output out=zele1 residual=res1 predicted=pred1;
run;
proc univariate plot normal data=zele1;
```

```
by env;
var res1;
run;
proc glm data=zele1;
by env;
class gen;
model res1=gen;
means gen/hovtest=bartlett;
run;
proc gplot data=zele1;
plot res1*pred1='*';
by env;
run;
```

B) SAS code for combined analysis

```
proc glm data=zelalem;
class env rep gen;
model yield=env rep(env) gen env*gen/ss1;
random env env*gen/test;
output out=zele2  residual=res2 predicted=pred2;
run;
proc varcomp method=REML;
class env rep gen;
model yield=gen env rep(env) env*gen;
run;
proc univariate plot normal data=zele2;
var res2;
run;
proc glm data=zele2;
class gen;
model res2=gen;
means gen/hovtest=bartlett;
run;
proc gplot data=zele2;
plot res2*pred2;
run;
quit;
```

C) SAS code for Stability analysis

```
proc sort data=zelalem;
by gen env;
run;
proc means noprint;
var yield;
output out=gxedata mean=gxe_avg;
by gen env;
run;
```

```
proc glm data=gxedata noprint;
class env gen ;
model gxe_avg=env gen;
output out=z residual=res;
run;
proc sort data=z;
by env;
run;
data b;
set z;
res_square=res**2;
run;
proc sort data=b;
by gen;
run;
proc means data=b noprint;
var res_square;
output out=wrick sum=Wi;
by gen;
run;
data wricks;
set wrick;
Ecovalence=wi;
drop wi _type_ _freq_;
run;
proc print data=wricks;
run;
proc means data=wrick noprint;
var Wi;
output out=ww  sum=sum_wi;
proc print data=ww;
run;
data shuk;
set wrick;
g=20;
e=8;
xx=12150055.79;
m=xx/((g-1)*(g-2)*(e-1));
stability_variance=(g/((g-2)*(e-1)))*Wi-m;
drop Wi g e xx m _type_ _freq_;
run;
proc print data=shuk;
run;
PROC sort data=gxedata;
by gen;
```

```
run;
proc means data=gxedata noprint;
var gxe_avg;
output out=env_var mean=environmental_mean var=environmental_variance cv=CVi;
by gen;
run;
data environmental;
set env_var;
drop _type_ _freq_;
run;
proc print data=environmental;
run;
proc gplot data=environmental;
plot CVi*environmental_mean;
run;
quit;
```

D) SAS code for Regression analysis

```
proc sort data=zelalem;
by env gen;
run;
proc means data=zelalem noprint;
var yield;
output out=gxedata mean=gxe_avg;
by env gen;
run;
proc glm data=gxedata noprint;
class env gen ;
model gxe_avg=env gen;
output out=z residual=res;/*res is an estimate of the interaction effect Gij*/
run;
proc means data=zelalem noprint;
var yield;
output out=zz mean=avg;/* avg is the mean yield of the environments*/
by env;
run;
data gg;
merge z zz;
by env;
run;
proc sort data=gg;
by gen;
run;
proc reg data=gg;
model res=avg/spec dw collin;
by gen;
```

```
output out=resres residual=res_res;
test avg=0;
run;
quit;
data regana;
set resres;
ressquare=res_res**2;
by gen;
run;
proc means data=regana noprint;
var ressquare;
output out=SSError sum=SSE;
run;
proc print data=SSError;
run;
proc reg data=gg;
model gxe_avg=avg/spec dw collin;
output out=gxeres residual=gxe_res predicted=gxe_pred;
by gen;
run;
data reganaw;
set gxeres;
ressquarew=gxe_res**2/6;
by gen;
run;
proc means data=reganaw noprint;
var ressquarew;
output out=SSError1 sum=SSE1;
by gen;
run;
proc print data=SSError1;
run;
```

E) SAS CODE FOR AMMI AANALYSIS

```
proc sort data=zelalem;
by env gen;
run;
proc means data=zelalem noprint;
var yield;
by env gen;
output out=avgyld mean=ylda;
run;
proc glm data=avgyld;
class gen env;
model ylda=env gen;
```

```
output out=residual r=res;
run;
proc IML;
use residual var{ENV GEN res};
read all var { env gen res };
m=20;
n=8;
resid=shape(res,m,n);
CALL SVD (u,q,v,resid);
axes='Axis-1':'Axis-8';
b=q#q;
sumb=sum(b);
e=(b/sumb)*100;
sqq=sqrt(q);
d=diag(sqq);
uq=u*d;
vq=v*d;
score='score 1':'score 8';
print 'eigin values', b[rowname=axes colname='ss' format=12.4];
print '%ss explained by each axes', e[rowname=axes colname='%ss' format=12.2];
print  'genotypic scores' ,uq[rowname='gen' colname= 'score' format=12.4];
print 'environmental score',vq[rowname='env' colname= 'score' format=12.4];
quit;
run;
```

Appendix III: SAS OUTPUTS

Table 1: **Individual (separate) analyses of variance (RCB design) for a trial with 20 hybrids and four replications, by environment.**

Environment	Sources of variation	Degree of freedom	Sum of squares	Mean squares
1	Replication	3	18722851.38	624283.79
	Genotype	19	10888312.31	573069.07
	Error	57	12378698.93	217170.16
2	Replication	3	147301.334	49100.445
	Genotype	19	5131004.336	270052.86
	Error	57	7617262.98	133636.19
3	Replication	3	1289011.60	429670.53
	Genotype	19	14916391.79	785073.25
	Error	57	14578289.73	255759.47
4	Replication	3	11895575.68	3965191.89
	Genotype	19	12906081.18	679267.43
	Error	57	10911808.48	191435.24
5	Replication	3	4317283.39	1439094.46
	Genotype	19	19868804.89	1045726.57
	Error	57	7027841.01	123295.46
6	Replication	3	2227983.262	742661.087
	Genotype	19	7686009.276	404526.804
	Error	57	11235935.45	197121.67
7	Replication	3	197496.721	65832.240
	Genotype	19	7644270.707	402330.037
	Error	57	16051456.2	281604.49
8	Replication	3	4315334.549	1438444.85
	Genotype	19	5166726.772	271932.988
	Error	57	11320510.03	198605.44

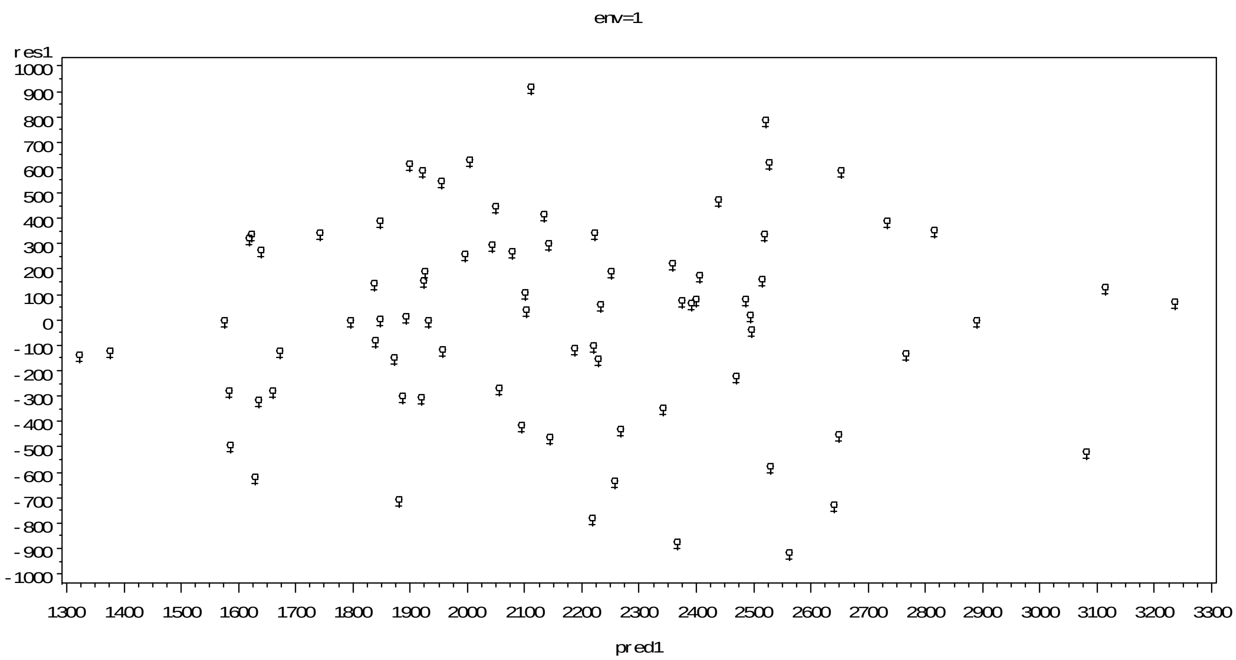
Table 2: **Bartlett's test for homogeneity of residual variance using genotype as a group in each of the separate ANOVA.**

Environment	Source	DF	Chi-Square	p-value=Pr>chi-square
1	Genotype	19	12.7759	0.8498
2	‘’	19	20.7092	0.3531
3	‘’	19	16.4989	0.6238
4	‘’	19	19.0659	0.4526
5	‘’	19	17.8538	0.5322
6	‘’	19	17.3993	0.5628
7	‘’	19	21.4293	0.3136
8	‘’	19	24.5922	0.1744

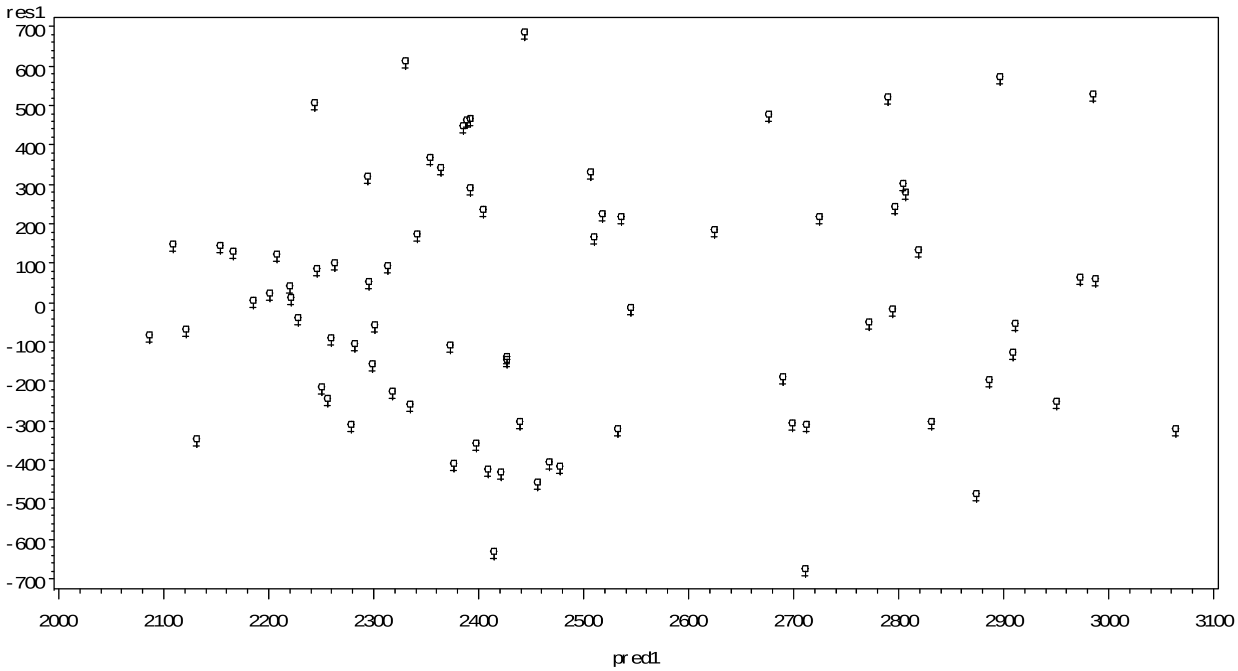
Table 3: **Test for Normality of residuals in each of the separate ANOVA model using the Shapiro-Wilk (W) statistic.**

Environment	Value of W	P-value=Pr<W
1	0.988484	0.6968
2	0.98672	0.5804
3	0.980236	0.2520
4	0.973841	0.0998
5	0.987432	0.6269
6	0.983928	0.4148
7	0.980582	0.2645
8	0.989733	0.7784

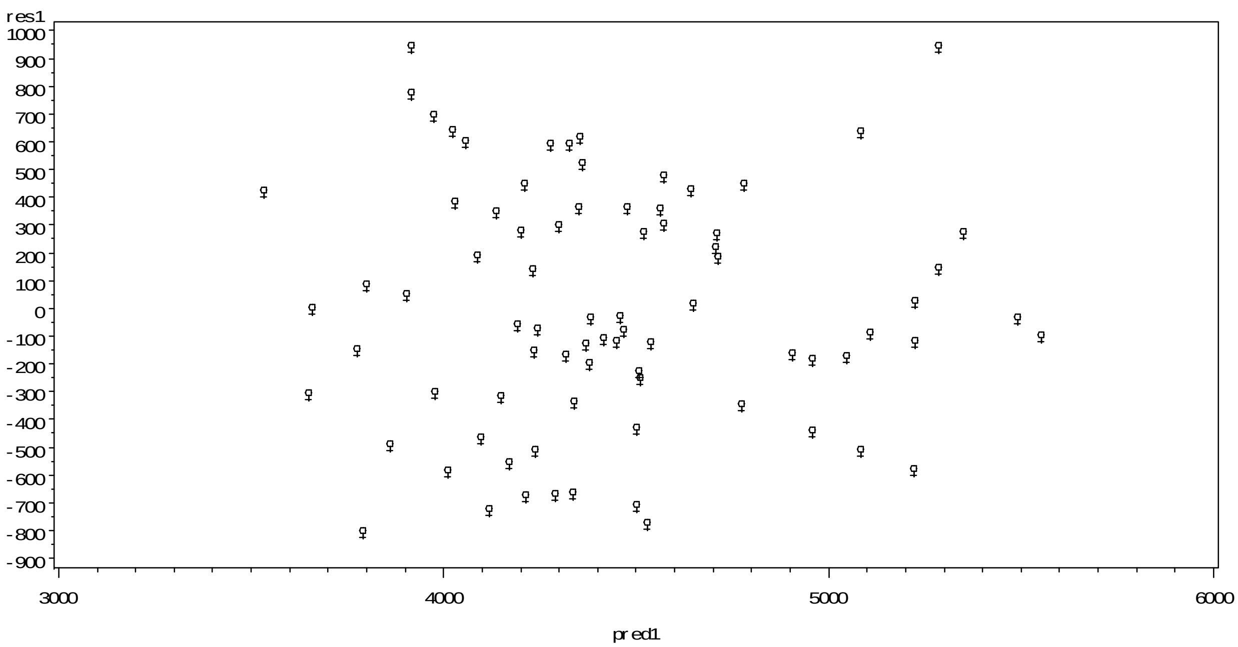
Fig 1: **plot of residuals against the predicted value in each of the separate ANOVA model**

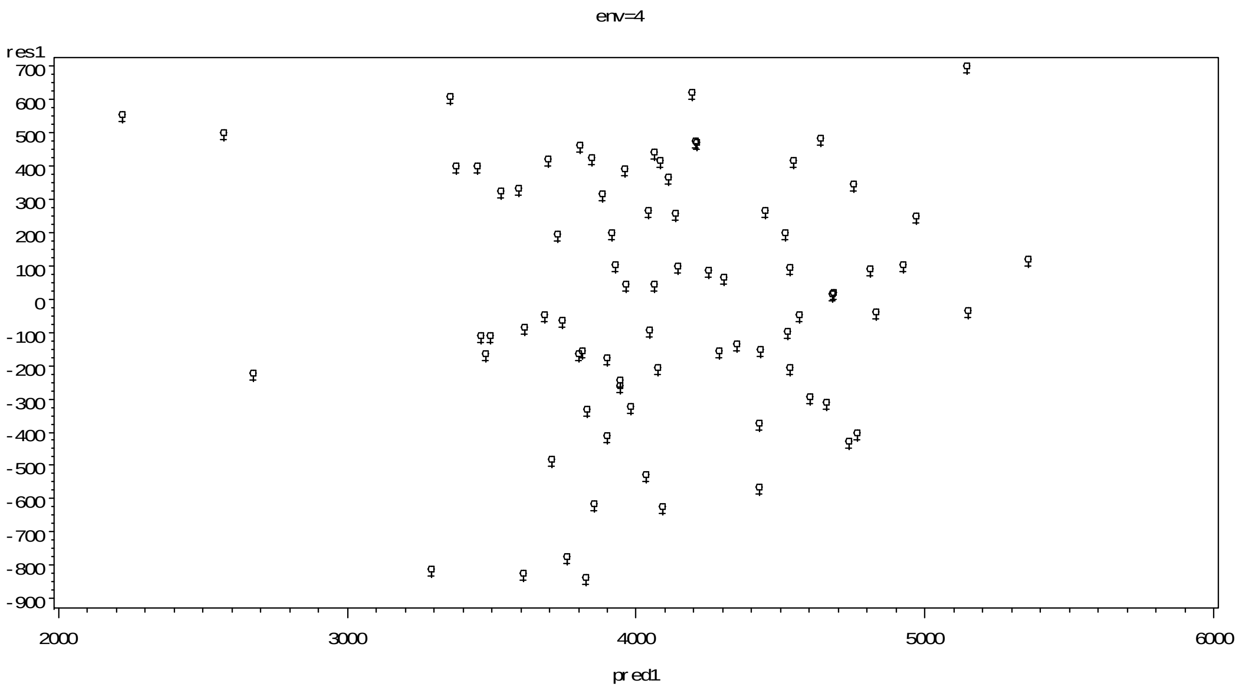


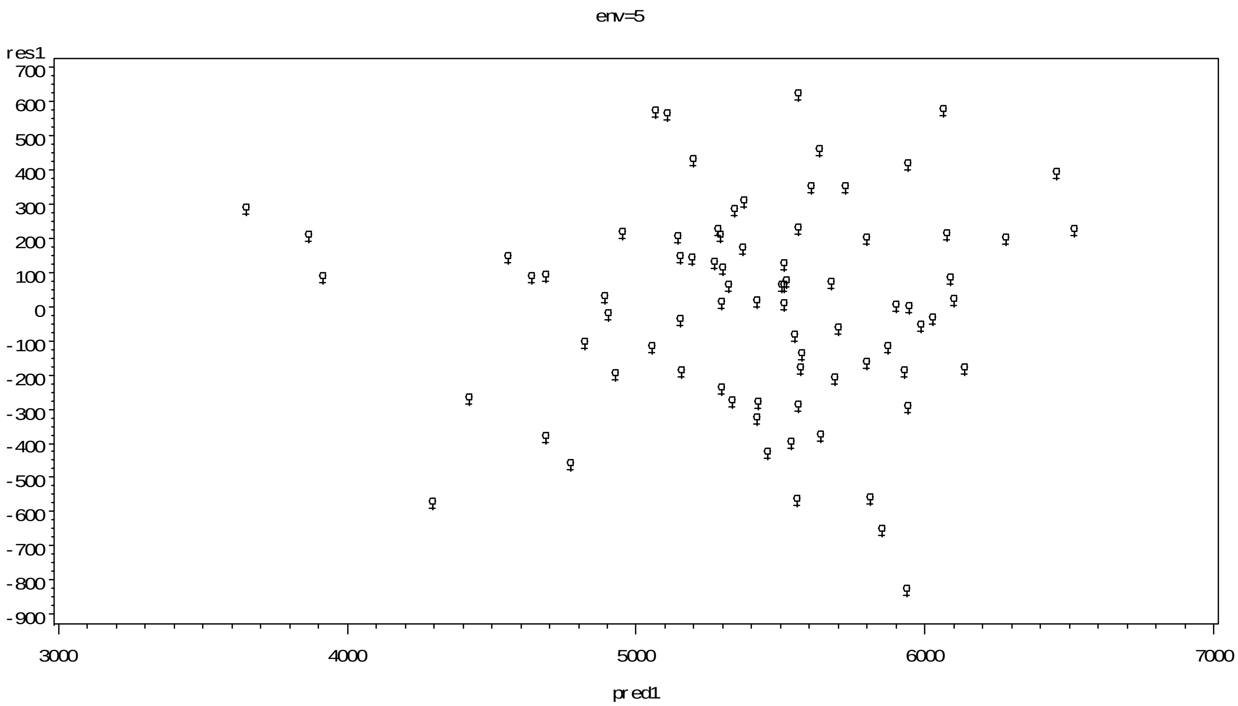
env=2


res1
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
3100
pred1

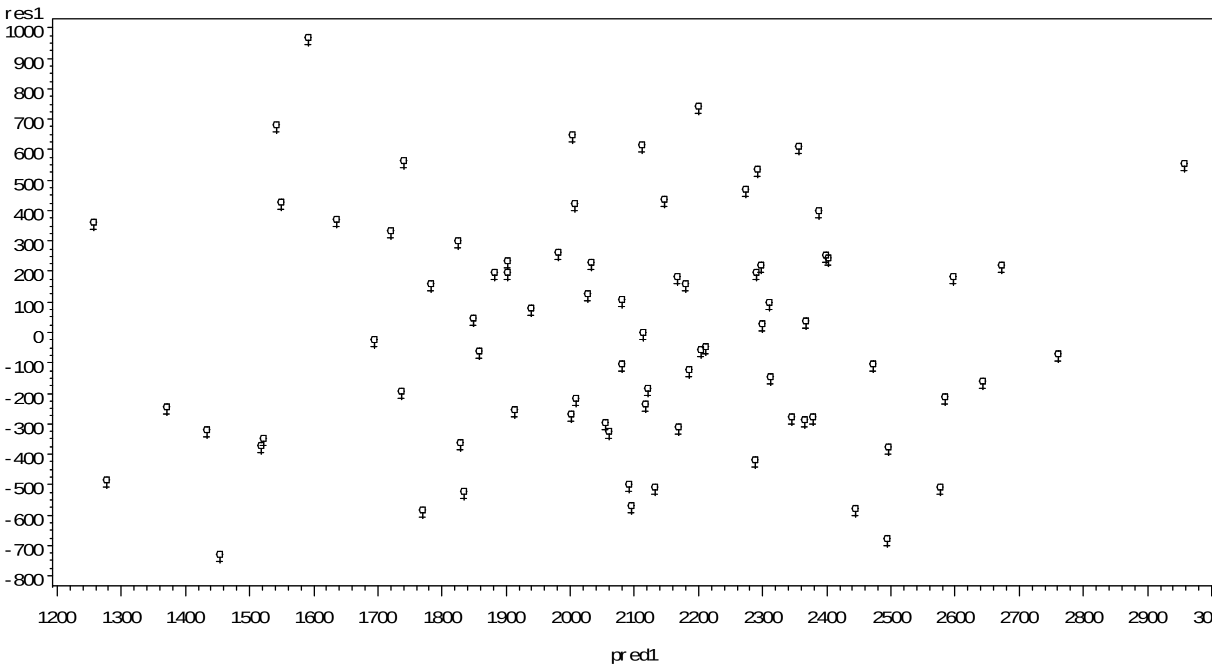
env=3


res1
1000
900
800
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
3000
4000
5000
6000
pred1

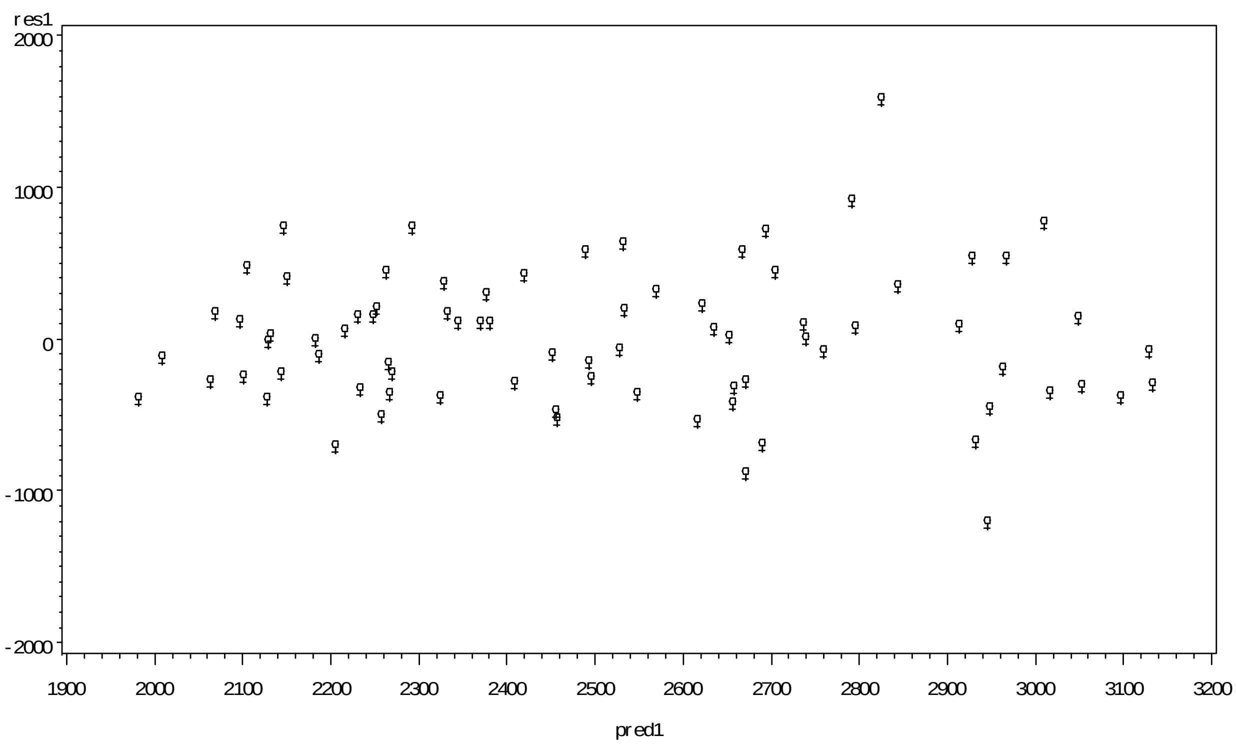

env=4
res1
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
2000
3000
4000
5000
6000
pred1

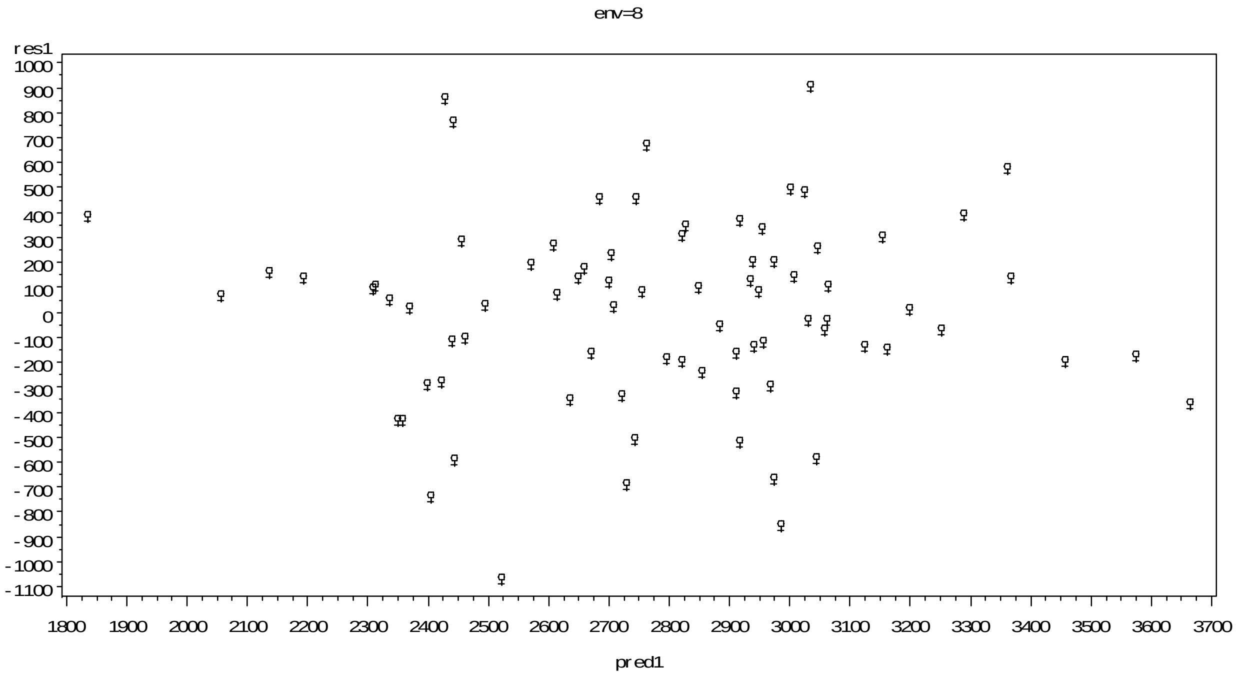

env=5
res1
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
3000
4000
5000
6000
7000
pred1

env=6


res1
1000
900
800
700
600
500
400
300
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000
pred1

env=7


res1
2000
1000
0
-1000
-2000
1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000 3100 3200
pred1



Tests for Normality for the combined ANOVA

Test	----Statistic------		-----p Value------	
Shapiro-Wilk	W	0.997695	Pr < W	0.5303
Kolmogorov-Smirnov	D	0.017266	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.019595	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.160132	Pr > A-Sq	>0.2500

Bartlett's Test for Homogeneity of residual Variance in the combined ANOVA

Source	DF	Chi-Square	Pr > ChiSq
gen	19	37.8848	0.0661

Fig 2: **Plot of residuals against the predicted value in the combined ANOVA**

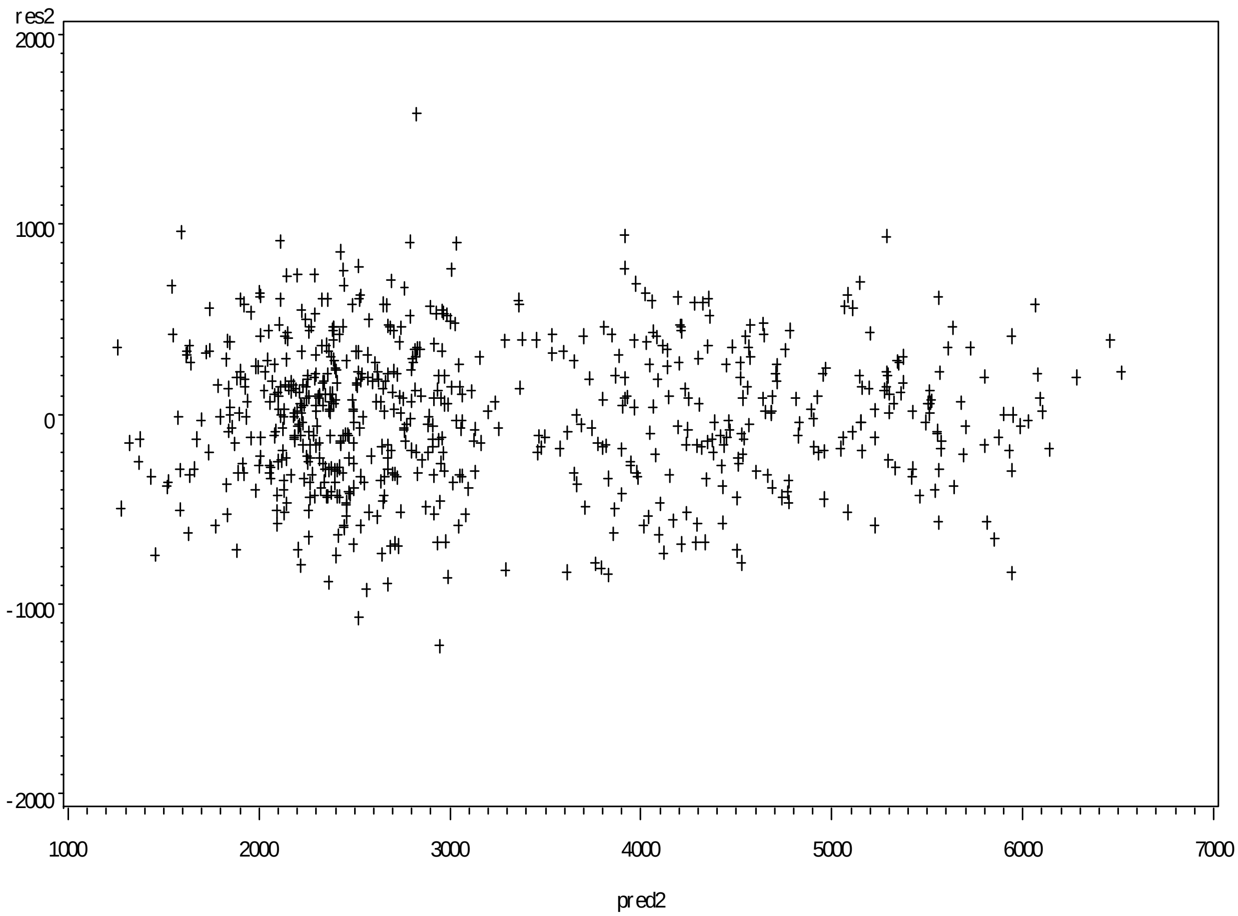


Table 4: Tests of model diagnostic in the regression of the interaction effects on the environmental mean.

Genotype	Test of first and second moment specification			Durbin Watson d	1^{st} order autocorrelation
	DF	Chi-square	Pr.>Chisq		
1	2	2.03	0.3625	1.518	0.137
2	2	1.49	0.4748	2.269	-0.151
3	2	1.45	0.4848	2.167	-0.131
4	2	1.41	0.4946	2.296	-0.245
5	2	3.26	0.1964	2.210	-0.160
6	2	3.09	0.2136	2.041	-0.264
7	2	2.58	0.2759	1.967	-0.013
8	2	1.17	0.5577	2.057	-0.198
9	2	2.58	0.2747	1.982	-0.016
10	2	1.91	0.3850	2.182	-0.113
11	2	2.66	0.2648	1.501	0.123
12	2	1.74	0.4185	2.284	-0.299
13	2	1.61	0.4463	2.186	-0.121
14	2	2.97	0.2265	2.403	-0.279
15	2	2.45	0.2939	1.909	0.031
16	2	3.09	0.2138	1.679	-0.082
17	2	1.57	0.4558	1.346	0.094
18	2	1.12	0.5715	3.249	-0.687
19	2	3.86	0.1452	2.198	-0.156
20	2	3.12	0.2097	2.674	-0.423

DECLARATION

I, the undersigned, declare that this thesis is my original work and it has never been presented in any other university. All sources of materials used for this thesis are duly acknowledged.

Declared by:

Name: ________________________________

Place: ________________________________

Date: ________________________________

Signature:______________________________

This thesis has been submitted for examination with my approval as a University advisor

……………………………………………………………

Girma Taye, Ph.D.