ANALYSIS AND INTERPRETATION OF GENOTYPE BY ENVIRONMENT INTERACTION USING ADDITIVE MAIN EFFECT AND MULTIPLICATIVE INTERACTION (AMMI) MODEL

A THESIS

presented by

FRANK OWUSU-ANSAH

in partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE

(In Mathematics)

of the

DEPARTMENT OF MATHEMATICS

FACULTY OF SCIENCE

KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

KUMASI, GHANA

JUNE, 2004.

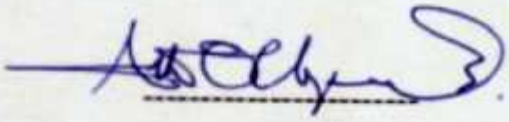
DECLARATION

I declare that I have personally undertaken the study reported herein under supervision

Date: 27/09/04

Frank Owusu-Ansah

(Student)



Date: 30/09/04

Dr. S.O. Dapaah

(Supervisor, Department of Mathematics)

DEDICATION

This piece of work is dedicated to the Almighty God, my parents and siblings.

ACKNOWLEDGEMENT

I would first of all thank the Almighty God for His abundant grace and strength provided towards a successful completion of this work. Secondly, I would like to acknowledge many individuals who have been very helpful in so many ways.

I am so much indebted to my noble supervisor, Dr. S.O.Dapaah, for suggesting the topic and whose constructive criticisms, counsel, patience and useful ideas have contributed immensely to this work. He as well provided me with the software and the data used in the analysis.

I also wish to thank all the lecturers of the mathematics department, especially the head of department, Dr. S.A Opoku, for their encouragement.

Finally my gratitude goes to my relatives and friends especially my brother David Owusu-Ansah, my only course mate and Bosom friend, David Maxwell Bessah and Eunice Agyemang for their company, encouragement and kind visits through the writing of this thesis.

May God bless you all.

Frank Owusu-Ansah.

ABSTRACT

One very important activity in plant breeding is to test a wide range of genotypes in a wide diversity of environments. Environment refers to site, year or a combination of site and year. The objective of the plant breeder is to select superior genotypes. The plant breeder however is usually confronted with the problem of genotype by environment interaction which complicates selection of superior genotypes. Genotype by environment (GE) interaction is a situation in which the performance of genotypes varies across different environments. GE interaction makes it inadequate for the plant breeder to recommend a particular genotype because its mean yield over the environments tested is high; it might have produced outstanding yield in some sites and performed poorly when grown in a particular site.

Several statistical methods have been proposed for the analysis of GE interaction. In recent years one of the most popular methodologies is the additive main effect and multiplication interaction (AMMI) model (Gauch, 1988) which was originally proposed by Gollob (1968) and Mandel (1971). This thesis is concerned with the analysis of GE data using the AMMI model.

The performance of the AMMI model is investigated by comparing genotypic correlations and their respective sums of squares using real data sets with the objective of highlighting the "optimism" associated with the fitting of the model. Results have shown that the interaction matrix exhibits high correlations between the genotype vectors which when ignored leads to optimism in the fitting sums of squares. This has prompted the development of the complement index vector as an alternative fitting procedure. The performance of the new approach is evaluated using real data sets.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

One of the objectives of the plant breeding program in agricultural research is to develop widely adapted high-yielding stable (superior) varieties for distribution to farmers to grow on their farms. The process of selecting high-yielding varieties involve the evaluation of a range of genotypes in relation to a varying environmental conditions, representing changing managements, locations, or seasons, with the aim of assessing the agronomic value of each genotype. Before the plant breeder makes a recommendation on which genotype should be selected for farmers to grow in their farms it is important to establish whether the relative performance of different genotypes varies from one environment to another.

This phenomenon often refers to as genotype by environment (GE) interaction makes it inadequate for the plant breeder to recommend a particular genotype because its mean yield over the environments tested is high; it might have produced outstanding yield in some sites and performed poorly when grown in a particular site. Thus the presence of GE interaction implies that the behaviour of genotypes in a trial depends on the particular environment in which they are grown.

In developing countries, particularly in Africa, there is a scarce use of agricultural chemicals and machinery and this limits the farmers' control over environmental stresses, including drought, heat, insects, pest and diseases. Wide variations in

climatic conditions and soil types also mean that no two growing environments are similar; hence genotype selection is crucial to avoid crop failures that subsistence farmers cannot afford.

GE interaction complicates selection of superior genotypes and can be classified into quantitative (non-crossover) and qualitative (crossover) interactions.

Figure 1a: No GE interaction.

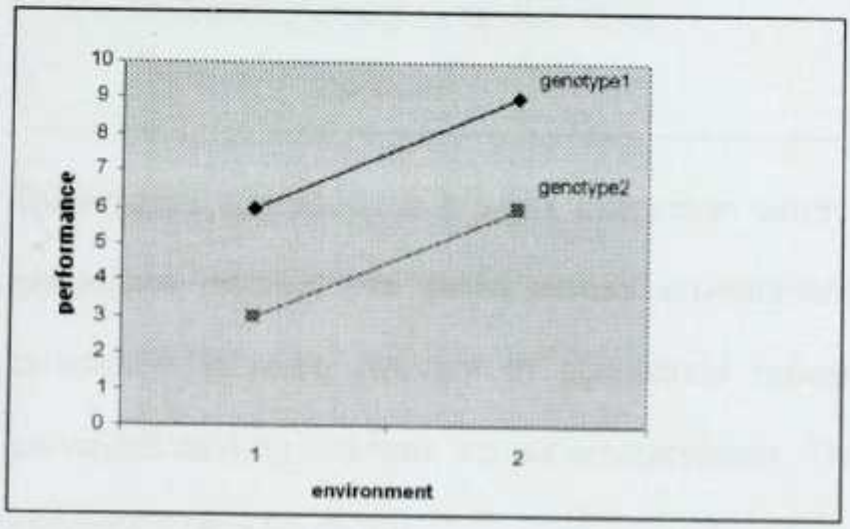


Figure 1b: Non-crossover GE interaction

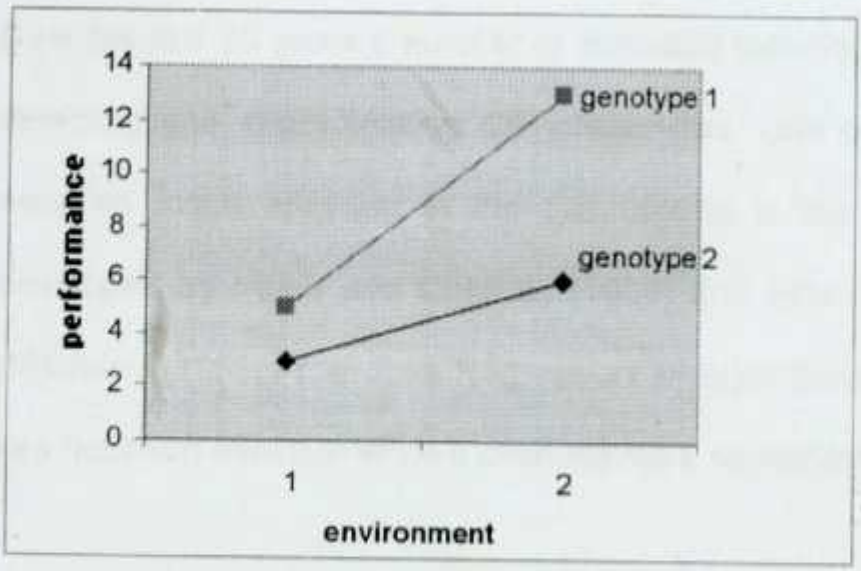
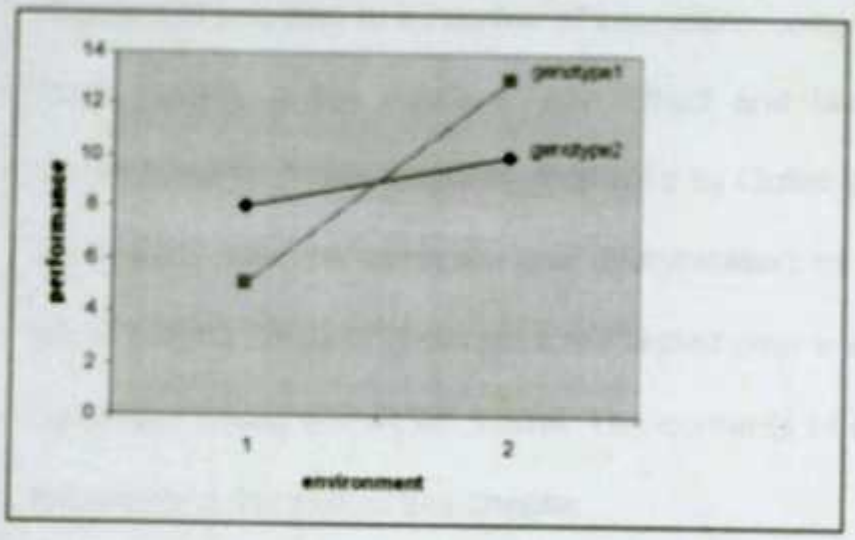


Figure 1c: Crossover GE interaction



Quantitative interaction is a weak interaction which occurs when the ranking of genotypes remains the same across environments (see figure 1b). Strong interaction is more relevant in agricultural research and is observed when genotypic ranking changes across environments. This is referred to as qualitative interaction (see figure 1c).

Over the last 30 years a number of statistical techniques have been proposed for detecting and characterising GE interactions. One of the analyses which have received much attention in the GE context is the Joint Regression analysis developed by Yates and Cochran (1938) and later rediscovered by Finlay and Wilkinson (1963). The Joint Regression analysis though partitions the interaction has not been effective since it often leaves a significant residual (Gauch, 1988).

In recent years, multiplicative models have achieved prominence over other statistical techniques simply because they breakdown the large interaction degrees of freedom to a number of interaction components (Mandel, 1971). One of such models is the Additive Main Effect and Multiplicative Interaction (AMMI) model (Gauch, 1988) originally proposed by Gollob (1968). This thesis is therefore concerned with the analysis and interpretation of GE interaction which occurs when a wide range of genotypes are tested over a wide diversity of environmental conditions using the AMMI model. The contents of each chapter in the thesis are described in the rest of this chapter.

In chapter 2 a review of the literature on some of the major methods for analysing GE interaction is presented. The AMMI model, in the context of GE interaction, involves fitting the additive model by analysis of variance and applying principal component analysis to the matrix of the interaction residuals. The problem with this approach is that the fitting sums of squares are "optimistic" due to the structure of the interaction residuals which tend to be correlated. Chapter 3 considers the theoretical and conceptual framework of the AMMI model.

In chapter 4, using a real data set we demonstrate the effectiveness of the AMMI model over analysis of variance and regression analysis. In chapter 5 the proportion of variation explained by the AMMI model is investigated using different sizes of the GE data matrix. The objective of this investigation is to highlight the "optimism" associated with the fitting sums of squares. In this thesis I propose the

use of complement index vector to overcome the problem of the optimism in the fitting sums of squares. This is developed by considering a subset of the number of genotypes to be analysed as the target set and the complement set used to estimate an environment index vector which is unrelated to the set of genotypes to be analysed. By using an externally provided index vector unrelated to the set of genotypes to be analysed we seek to eliminate the optimism in the estimation of the index vector. The performance of the proposed index vector is evaluated using real data sets.

Finally concluding remarks of this research and possibility for further study are presented in chapter 6. References of literature used are provided before the appendixes. Appendix A gives additional results which were not given in the main chapters on sums of squares and their respective correlations of randomly selected genotypic vectors of a real data set whilst appendix B gives the Genstat programs which were used for all the analyses. In Appendix C the SVT data is provided.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Since the early part of the 20th century considerable attention has been given to the nature of genotype by environment (GE) interactions and the techniques used for analysing such interactions. In chapter one the term genotype by environment (GE) interactions was explained. In this chapter we present a review of some of the most important statistical models which have been used for the analysis of GE interaction. The chapter ends with a conclusion on the methods discussed.

2.2 Analysis of Variance

The preliminary analysis of data from trials consisting of a data set of genotypes of a crop tested in different environments is usually done by analysis of variance (ANOVA) introduced by Fisher (1926). The ANOVA model partitions the total variation into three components, one measuring the differences between genotypes, one measuring the environmental differences and the last assessing the joint effects.

The ANOVA model is

$$y_{ijk} = \mu + g_i + e_j + w_{ij} + \varepsilon_{ijk} \qquad (2.1)$$

where y_{ijk} =yield of the i^{th} genotype in the k^{th} replicate of the j^{th} environment

μ=overall mean yield,

g_i =effect of the i^{th} genotype,

e_j = effect of the j^{th} environment,

w_{ij} =interaction of the i^{th} genotype and the j^{th} environment and

ε_{ijk}=the error term assumed to be normally distributed with zero mean and variance which is homogeneous over genotype and environment.

If the goal of the breeder is to select genotypes which are the "best ", average over sites is actually used; a model with fixed site term is appropriate. The results of the analysis apply to this sample of particular sites only. Mead (1988) advocates the method of selecting sites in a random way. He argues that a breeder should try to characterise the major differences within the region and select sites that span those characteristics. The other option that the plant breeder intends to develop genotypes that are superior in the region and not specifically at the sites actually used is to choose the sites at random, here creating a random sample from the population of sites. In this case inferences made after the analysis of the data apply to the whole population of sites, which represents the region.

The random selection of genotypes of maize from some region would for example represent a random factor whose mean and variance may be what is of interest to the data analyst.

2.3 Variance Components

Variance component is a technique used for analysing designs that include random effects (random or mixed models). There are several methods for estimating the variance components. Sprague and Federer (1951) showed how variance components could be used to separate out the effects of genotypes, environments and their interaction by equating the observed mean squares in the analysis of variance to their expectations on the random model.

Variance components have widely been used in genetics and plant breeding. For instance Miller, Williams and Robinson (1959) applied it on cotton data. Miller, Williams and Robbinson (1962) in considering GE interaction from several response variables measured on cotton genotypes, found that the size of genotype x site x year mean square exceeded the genotype x site mean square and genotype x year mean square and even equalled to the variance component due to genotype main effects.

Plaisted and Peterson (1959) presented a method to characterise the stability of a genotype when several genotypes were tested at a number of locations within one year. They proposed averaging the variance components σ^2_{GE} determined from each pairwise combination involving the genotype with every other in a trial. The genotype with the smallest mean square value was thus considered as the one that contributed the least to genotype x environment interaction and thus would be considered the most "stable" genotype in the tests. If a large number of

genotypes were tested, this would call for a large number of analysis, I(I-1)/2 pairs for I genotypes . Wricke (1962, 1964) proposed the sum of squares of interaction effects for each genotype as a measure of stability of its behaviour over environments. This was criticised by Freeman and Perkins (1971) on the grounds that the interaction sum of squares with (I-1)(J-1) df (where J is the number of environments) cannot be partitioned with I orthogonal components.

Shukla (1972) proposed an unbiased estimate of $\sigma^2_{(i)}$, the interaction variance over environments for the i^{th} genotype, and devised a criterion for testing whether $\sigma^2_{(i)}$ was large enough for it to be considered unstable. Hanson (1970) devised a composite measure of stability which combined the contribution of variance component approach and joint regression analysis (section 2.4).

Another measure of genotype stability based on the assumption that each genotype is exposed to a random sample of environments was proposed by Francis and Kannenberg (1978). They plotted for each genotype the coefficient of variation against the mean. They suggested that useful genotypes were those with a high mean but low coefficient of variation.

Often a breeder's assessment of the value of a genotype is based on comparison with one or more checks. Lin and Binns (1988) suggested using pairwise analysis of variance between known check genotype and all tested genotypes to detect which genotypes show the same adaptation as checks in variety trials. Genotypes

that do not elicit significant GE interaction with a check were identified and their yields classified as significantly superior to, inferior to or no different from the mean of a check. Genotypes with similar adaptation patterns to standard checks and higher yields across sites were identified for recommendation.

Lin and Binns (1988) proposed another measure of genotypic stability, P_i, the squared difference between yield of genotype i and the highest yielding genotype averaged over all environments. With this method, different checks can be used from site to site. A small value for the superiority index (P_i) implies general adaptation of a genotype. The index integrates mean yield and relative stability into one parameter and this may be influenced by the scale of observations, important when ranges of site mean yields are large. In such cases, the determination of P_i may be dominated by higher yielding sites. Standardisation of data by site mean and variance would overcome this, but could reduce the influence of genotype means.

The literature discussed so far dealt with complete tables but a characteristic feature of the data collected for GE studies is their lack of balance. In most breeding programs, the genotype list changes over years and not all genotypes are tested at all locations within each year. This situation creates unbalance two way or three way tables which in the past were tedious to analyse without substantial computing power. By the mid-1970's this situation had changed which motivated Patterson and Thompson (1971) to develop the general technique of

residual maximum likelihood (REML) analysis of unbalanced data. REML estimates the components of variances by maximizing the likelihood of all contrasts with zero expectation. Genotype means are estimated by generalised least squares with weights depending on the estimated variance components. Indeed, for balanced data, REML and ANOVA estimators are identical. For unbalanced data, using REML will allow appropriate terms of the ANOVA model to be taken as random and estimate parameters of fixed and random terms including the corresponding variance components.

All the stability measures discussed in this section seek to characterise genotype response by a mean and a variance measure without providing any explanation for the behaviour of each genotype at a given site in a given season. The paucity of information provided on interaction by ANOVA has led to the need to break down the interaction in order to understand and utilise this information.

Another disadvantage is that environments are assumed to form a randomly selected sample from a well defined population, but the choice of sites for genotype trials is far from random and as genotypes are tested only for a few years they may well not experience a random sample of weather conditions. In the next section I discuss Regression Analysis, a model which does not assume that environments are a random sample from the population and also offer various interaction partitioning methods for further analysis.

2.4 Linear Regression Analysis

One form of the analysis which has received much attention in the GE context is the joint regression analysis developed by Yates and Cochran (1938). Yates and Cochran applied regression on their barley yield data. The method was not really taken up until Finlay and Wilkinson (1963) rediscovered the same method and used it for an analysis of adaptation in a trial with 277 varieties of barely in seven environments. Further use of regression followed. Bresse (1969) used regression to analyse herbage grass.

Regression analysis of GE data involves two stage fitting procedures.

(i) Fit the additive model

$$y_{ij} = \mu + g_i + e_j + w_{ij} \qquad (2.2)$$

where y_{ij} =yield of the i^{th} genotype in the j^{th} environment

μ=overall mean yield,

g_i =effect of the i^{th} genotype,

e_j = effect of the j^{th} environment,

w_{ij} =interaction of the i^{th} genotype and the j^{th} environment

(ii) using $\hat{e}_j$ as the environment index calculate the regression dependence of the GE interaction terms on the index by regressing $\hat{w}_{ij}$ on $\hat{e}_j$ for the set of genotypes , obtaining regression coefficient $\hat{\beta}_i$ for the environments. Effectively we are fitting the model

$$\hat{w}_{ij} = \beta_i \hat{e}_j + d_{ij} \qquad (2.3)$$

where d_{ij} represent the residual interaction variation. We can therefore write the full model as

$$y_{ij}=\mu+g_i+e_j+\beta_i\hat{e}_j+d_{ij}+\varepsilon_{ij} \quad (2.4)$$

Essentially, the method partitions the (I – 1) (J – 1) df for interaction into (I – 1) df for heterogeneity among genotype regressions and the remainder (I – 1) (J – 2) df for deviation.

Following the development of Finlay and Wilkinson's stability parameters a number of definitions and measures of stability were proposed, all based on the use of environment mean as a measure of the environment.

Eberhart and Russell (1966) proposed in addition to the regression coefficient another statistic based on the deviation $S^2_{d(i)}$ of each genotype's response from the fitted line.

$$S^2_{d(i)}=\left[\frac{\sum_{j=1}^{t}\hat{d}^2_{ij}}{t-2}\right]-\frac{S^2_e}{r} \quad (2.5)$$

where $d^2_{ij}=\left(y_{ij}-\hat{\mu}-\hat{g}_i-\hat{\beta}_i\hat{e}_j\right)^2$

S^2_e = Estimates of the pooled error mean squares and

r = Number of replicates.

A variety's stability was now assessed by two indices; $\hat{\beta}_i$ and $S^2_{d(i)}$. Both indices defined a recommendable genotype as one with high mean yield , a regression coefficient equal to 1.0 and a standard deviation from regression as small as possible ($S^2_{d(i)}$=0).

2.4.1 Limitations of Linear Regression

While joint regression has been widely used for analysing GE interactions, the method has been criticised on statistical grounds by Freeman and Perkins (1971).

The main criticisms are that the estimation process of regression parameters violates basic assumptions. This is due to the fact that the environment mean is correlated with the response of the genotype that has contributed with it. This problem is however reduced for at least 15 or 20 genotypes (Crossa, 1990; Freeman and Perkins, 1971; Freeman, 1973; Hill 1975).

Following the criticisms of Freeman and Perkins (1971), a number of studies were carried out to compare the results using dependent and independent measures of the environment. Fripp and Caten (1971) used dikaryous of Schizophyllium commune and measured the growth rates of those fungus genotypes in a malt growth medium at fixed, different temperatures and with differing concentrations of the malt extract. They used general biological measures of the environment, ranging from setting aside a replicate of the test genotypes, to using the response

of unrelated genotypes not included in the set of test genotypes as well as the usual dependent environment measures.

Perkins and Jinks (1973) regressed values for a large number of genotypes on their environmental means and on values derived from another closely related sets of genotypes; they found that all the analyses generally gave similar values for significance, but that regressions on means derived from only a few independent genotypes were sometimes so insensitive as to give rise to problems of interpretation.

2.5 Principal Component Analysis

Principal Component Analysis (PCA) involves a mathematical procedure that transforms a set of correlated response variables into a smaller set (reduced dimensionality) of uncorrelated variables called principal components.

The technique which is used primary as an exploration procedure is one of the most frequently used multivariate methods (Pearson, 1901; Gower, 1966). Using the PCA for the analysis of GE interaction the model first removes the grand mean μ after which an eigenanalysis is performed to partition the residual into;

i. A multiplicative model with PCA axis 1 to N, each comprising the product of singular values λ_n and their corresponding genotype and environment eigenvector (indices), ξ_{ni} and η_{nj} respectively.

ii. A residual ρ_{ij} if not all IPCA axes are used

iii. The error, ε_{ij}

Thus essentially, we are fitting the model

$$Y_{ij}=\mu+\sum_{n=1}^{n}\lambda_n\xi_{in}\eta_{nj}+\rho_{ij}+\varepsilon_{ij} \qquad (2.6)$$

Under certain conditions PCA is a generalisation of the linear regression analysis (Mandel, 1969; Johnson, 1977; Digby, 1979). Herosaki et al. (1975) found that PCA was more efficient than linear regression method in describing genotypic performance. Freeman and Dowker (1973) used PCA to interpret the causes of GE interaction in carrot trials. Perkins and Jink (1968), Hardwick and Wood (1972), and Hill and Goodchild (1981) have also used this approach to find groupings in genotypes.

An additional facility that helps to understand GE interaction after performing principal component analysis is the method of biplot. Biplot was proposed by Gabriel (1971) for the graphical representation of any matrix whose rank is 2 by means of the rows and column vectors. This method summarises the most important pattern of a GE data.

Kempton (1984) illustrated the versatility of biplot for GE data and the method has since been used. Recently extension to the use of the biplot has involved the study of the underlying statistical analysis rather than its associate set of plotted points. Important questions like which genotype performed well in which environment,

mega environment investigation, mean performance and stability of genotypes etc. are graphically addressed by the biplot analysis (Gauch, 1992).

2.5.1 Limitations of Principal Component Analysis

Though PCA has proven to be valuable in GE studies, a number of criticisms have been noted (J. Crossa, 1990) which includes the following;

i. In reducing the dimensionality of multivariate data, distortions may sometimes occur.

ii. A lack of correlation between variables prevents few dimensions from accounting for most of the variation (Williams, 1976).

iii. Sometimes the PCA components may not have any obvious relationship to environmental factors.

iv. Contrary to ANOVA which assumes a complete additive model and treats the interaction as a residual, PCA assumes a complete multiplicative model without any description of the main effects of genotypes and environments (Zobel et al, 1988).

v. Non-linear association in the data prevents PCA from efficiently describing the real relationships between entities (Williams, 1976)

2.6 Additive Main Effect and Multiplication Interaction (AMMI) Analysis

The additive main effect and multiplication interaction (AMMI) model involves an initial fitting of the additive component for genotypes and environments by analysis of variance followed by principal component analysis to fit the multiplicative (interaction) component. The AMMI model is

$$y_{ijk} = \mu + g_i + e_j + \sum_{l=1}^{L} \Gamma_{il} \Psi_{jl} + \varepsilon_{ijk} \qquad (2.7)$$

y_{ijk}= yield of the kth replicate of the ith genotype in the jth environment,

μ= the grand mean,

g_i = the genotype deviation

e_j=the environment deviation.

L = the number of multiplicative components (IPCA axes)

e_{ijk}= the random factor

Γ_{il} and Ψ_{jl} are the genotype and environment scores (IPCA scores) respectively obtained from the eigenanalysis of the interaction residual matrix (See chapter three).

This model (Gauch, 1988) originally proposed by Gollob (1968) and Mandel (1969, 1971) has proved valuable in GE interaction studies (Gauch, 1988; Gauch and Zobel, 1988; Zobel et al, 1988; Crossa et al, 1990).

Zobel et al. (1988) used AMMI model to explain 72% of the GE sums of squares and thus provided a more adequate biological explanation of GE interaction than regression, which accounted for 19% of interaction. Boonseng et al (1997)

reported AMMI analysis of dry root yield of 5 cassava genotypes grown in early raining season. Chatwachirawong (1993) in his studies in sugarcane experiment indicated that only one multiplicative component (IPCA 1) was enough to explain 90% of the GE interaction variation. Ebdon and Gauch (2002) used the AMMI model to analyse the GE interaction of National Turfgrass.

There is debate about appropriate test for retaining multiplicative components in the AMMI model (Cornelius, 1993). Cornelius found that a conservative test used by several researchers resulted in far too many significant components and warned AMMI users about the problem.

Gauch and Zobel (1988) distinguished two processes of determining the number of IPCA axes to be considered using the terms Predictive and Postdictive accuracy (see chapter three). They recommended predictive modelling. Annicchiarico (1997b), however, reporting alternative tests of the AMMI model recommended the F ratio test.

It is possible to visualise the first few multiplicative components in a biplot diagram. One problem about AMMI analysis is that when the best AMMI model includes more than two IPCA axes, assessment and presentation of genotypic stability are not as simple as the AMMI model with 1 or 2 IPCA axes (called AMMI1 and AMMI2 respectively) case (P.N Fox). Despite the outstanding performance of the

AMMI model (Gauch, 1992) in terms of predictive assessment there has been some instances where the AMMI model has shown controversial results.

Crossa(1991) from biological standpoint indicated the possible danger of over-riding a message that is asynchronous with pattern when a single site is with a particular stress among the majority of sites without this stress. Additionally he noted an alarming result of 9 genotypes grown in 38 subtropical environments, where GE interaction was highly significant, yet AMMI model with no multiplicative component (AMMI0) showed the best predictive value indicating that the significant interaction is noise (Crossa et al, 1991; Crossa et al, 1990).

2.7 Other Methods of Analysis

There are other methods which have been used to analyse GE data. One of the popular methods is the Cluster Analysis. The objective of cluster analysis is to group observations into clusters such that each cluster is as homogeneous as possible with respect to the clustering variables.

The clustering procedure begins with the selection of a measure of dissimilarity (similarity), followed by a decision on the type of clustering technique to be used (hierarchical or non-hierarchical). Next is the clustering method (for example, centroid method in clustering technique) followed by a discussion regarding the number of clusters required. Finally the cluster solution is interpreted. Several techniques for evaluating the cluster solution as to whether clusters are quite

homogeneous and well separated at any given step are available, with the more common ones being Root-Mean-Square Standard Deviation (RMSSTD) of the new clusters , Semi-partial R-Squared (SPR), R-Squared (RS) and Distance between two clusters.

A number of researchers have used cluster analysis to analyse GE data. Abou-El-Fittouh, Rawlings, and Miller (1969) applied cluster analysis to classify locations used in cotton variety trials in the USA. Mungomery et al. (1974) employed an unstandardised Euclidean distance as a dissimilarity measure, using an unw eighted group average link clustering strategy (Sokal and Michener, 1958). Others who have applied cluster analysis to GE data include Byth et al (1976), Lin and Thompson (1975), Lin (1982), Fox and Rosielle (1982), Ramey and Rosielle (1983) and Ghaderi et al (1982).

A limitation of clustering is that grouping will be achieved even when there may not be any natural groupings or clusters in the data. Gordon (1981) and Cormack (1971) criticised cluster analysis on the grounds that different clustering methodologies can lead to different results producing misleading conclusions.

Pattern analysis refers to the successive use of classification and ordination techniques to present the maximum variation in a multidimensional data on a reduced dimension. Whilst classification techniques such as clustering assumes discontinuities within the data, ordination techniques which seeks to represent a

higher dimensional space on a lower dimension assumes continuous distribution. Principal component analysis is one of the popular ordination techniques. An example of pattern analysis therefore constitutes the successive application of cluster analysis and Principal component analysis.

Principal component analysis combined with cluster analysis was effective in forming subgroups among 29 populations of faba bean (Vicia faba L.), which differed in mean performance and response across environments (Polignano et al., 1989). Jose Crossa et al. (1990) used AMMI1 estimates as input to cluster analysis resulting in a more cohesive and clearer explanation of response patterns of 18 CIMMYT bread wheat genotypes in 25 locations.

There are other methods which have also been used. Principal Coordinate Analysis (Gower, 1966) developed by Schoenberg (1935), Young and Householder (1938) and Torgerson (1952), sometimes referred to as "classical scaling" has been used to analyse GE data. It has been used to study the adaptation of soybean lines evaluated across environments in Australia (Mungomery et al, 1974; Shorter et al, 1977).

Factor analysis has been used to understand relationship among yield components and Morphological characteristics of crops (Walton, 1972; Seiler and Stafford, 1985). Peterson and Pfeiffer (1989) applied principal factor analysis to study the underlying structures and relationships of tests sites, based on winter

wheat performance. Correspondence analysis (Hill 1973, 1974) also termed "reciprocal averaging", redundancy analysis (Rao, 1964) and stochastic dominance (Menz, 1980) have also been used. Fox et al (1990) used stratified ranking method to evaluate the proportion of sites where any genotype ranks in the top, middle or bottom third of the entries.

2.8 Conclusion

This chapter has reviewed the literature of a large range of statistical methods applicable to the analysis of GE data. The big question is that which model or method is preferable and for which data and which objective. More often than not it is necessary to perform several analyses and compare the results.

A large number of these statistical methods including AMMI can be useful for model diagnosis. Sometimes a preliminary AMMI analysis may be extremely effective, motivating no further analyses, whilst in other cases it may serve to diagnose a different model that is likely to be the best (Gauch, 1992). Currently AMMI is popular for two-way factorial data with the objective of accounting for the main effect and interaction effect separately. We proceed to the next chapter where the conceptual and theoretical framework of the AMMI model is been discussed.

CHAPTER THREE

THEORY AND CONCEPT OF THE AMMI MODEL

3.0 Introduction

In the advent of modern powerful computers, one can easily analyse data without knowing the mathematical concepts behind it. However, to be able to effectively analyse data by a statistical tool (Model) it is very important to understand the theoretical and conceptual framework of the model.

This chapter is therefore designed to highlight the theoretical and conceptual framework of AMMI model. Section 3.1 discusses the data organization after which section 3.2 presents the AMMI model in a very systematic fashion coupled with the mathematical theory behind the model.

3.1 Data Organisation

An experiment of assessing the yield performance of I genotypes in J environments with K replication of each genotype results in a two-way table with an IJK number of observations. The observations (yield) are organized in a matrix form called a data matrix such that y_{ijk} is the yield of the i^{th} genotype in the j^{th} environment for the k^{th} replication. Fig. 3.1 shows the yield distribution of 3 genotypes in 4 environments with 3 replications.

Fig 3.1 Yield of 3 genotypes in 4 environments with 3 replications

Genotype	Environment			
	1	2	3	4
1	y_{111} y_{112} y_{113}	y_{121} y_{122} y_{123}	y_{131} y_{132} y_{133}	y_{141} y_{142} y_{241}
2	y_{211} y_{212} y_{213}	y_{221} y_{222} y_{223}	y_{231} y_{232} y_{233}	y_{241} y_{242} y_{243}
3	y_{311} y_{312} y_{313}	y_{321} y_{322} y_{323}	y_{331} y_{332} y_{333}	y_{341} y_{342} y_{343}

For example, from Fig 3.1, y_{321} is the 1st replicate of the 3rd genotype in the 2nd environment. An experimental effect is removed from the data matrix by averaging the data across the effect.

1. The replicate effect can be removed from the matrix by averaging the replicate of a particular genotype in a given environment denoted by y_{ij} Thus,

$$Y_{ij.} = \frac{\sum_{k=1}^{K} y_{ijk}}{K} \qquad (3.1)$$

For instance, from the above figure (fig 3.1),

$$y_{23\bullet} = \frac{y_{231} + y_{232} + y_{233}}{3}$$

The removal of the replicate effect reduces the data matrix to comprise IJ points (cells). For instance Fig 3.1 reduces to the matrix

$$Y_{ij} = \begin{pmatrix} y_{11\bullet} & y_{12\bullet} & y_{13\bullet} & y_{14\bullet} \\ y_{21\bullet} & y_{22\bullet} & y_{23\bullet} & y_{24\bullet} \\ y_{31\bullet} & y_{32\bullet} & y_{33\bullet} & y_{34\bullet} \end{pmatrix}$$

2. The environmental effect is further removed to obtain the genotypic effect by averaging each genotype across the environments denoted by $y_{i..}$

$$Y_{i..} = \frac{\sum_{j=1}^{J} y_{ij.}}{J} = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}}{JK} \qquad (3.2)$$

For instance

$$y_{1\bullet\bullet} = \frac{y_{11\bullet} + y_{12\bullet} + y_{13\bullet} + y_{14\bullet}}{4}$$

The genotypic effect in a 3 x 4 matrix is given as

$$Y_{i..} = \begin{pmatrix} y_{1\bullet\bullet} & y_{1\bullet\bullet} & y_{1\bullet\bullet} & y_{1\bullet\bullet} \\ y_{2\bullet\bullet} & y_{2\bullet\bullet} & y_{2\bullet\bullet} & y_{2\bullet\bullet} \\ y_{3\bullet\bullet} & y_{3\bullet\bullet} & y_{3\bullet\bullet} & y_{3\bullet\bullet} \end{pmatrix}$$

3. In a similar way as above, the environment effect is obtained by

$$y_{j.} = \frac{\sum_{i=1}^{I} y_{ij}}{I} = \frac{\sum_{i=1}^{I}\sum_{k=1}^{K} y_{ijk}}{IK} \qquad (3.3)$$

The environment effect in a 3 x 4 matrix is given as

$$Y_{\cdot j \cdot} = \begin{pmatrix} y_{\cdot 1 \cdot} & y_{\cdot 2 \cdot} & y_{\cdot 3 \cdot} & y_{\cdot 4 \cdot} \\ y_{\cdot 1 \cdot} & y_{\cdot 2 \cdot} & y_{\cdot 3 \cdot} & y_{\cdot 4 \cdot} \\ y_{\cdot 1 \cdot} & y_{\cdot 2 \cdot} & y_{\cdot 3 \cdot} & y_{\cdot 4 \cdot} \end{pmatrix}$$

Finally the grand mean y... is evaluated as

$$y_{\cdots} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}}{IJK} \qquad (3.4)$$

The grand mean matrix is given as

$$Y_{\cdots} = \begin{pmatrix} y_{\cdots} & y_{\cdots} & y_{\cdots} & y_{\cdots} \\ y_{\cdots} & y_{\cdots} & y_{\cdots} & y_{\cdots} \\ y_{\cdots} & y_{\cdots} & y_{\cdots} & y_{\cdots} \end{pmatrix}$$

The two-way yield data is anlaysed by using the additive main effect and multiplication interaction (AMMI) model (Gauch, 1988)

3.2 AMMI Model

The AMMI model of a two-way data is given by

$$y_{ijk} = \mu + g_i + e_j + \sum_{l=1}^{L} \Gamma_{li} \Psi_{lj} + \varepsilon_{ijk} \qquad (3.5)$$

The components of equation (3.5) carry the meaning defined as

y_{ijk}= yield of the kth replicate of the ith genotype in the jth environment,

μ= the grand mean,

g_i = the genotype deviation

e_j=the environment deviation.

L = the number of multiplicative components (IPCA axes)

e_{ijk}= the random factor

Γ_{ii} and Ψ_{ij} are the genotype and environment scores (IPCA scores) respectively obtained from the eigenanalysis of the interaction residual matrix

This model (3.5) can be restructured as

$$Y_{ijk} = A + W \tag{3.6}$$

where A is the additive part and W is the multiplicative part such that

$$A = \mu + g_i + e_j \tag{3.7}$$

and

$$W = \sum_{l=1}^{L} \Gamma_{il} \Psi_{lj} + \varepsilon_{ijk} \tag{3.8}$$

3.2.1 Estimation of Model Parameters

The estimation of the parameters of the AMMI model is classified under two phases. The first phase estimates the additive part using analysis of variance and the second phase estimates the multiplicative part (which results as the residual of ANOVA together with the error term) using principal component analysis.

3.2.2 Phase I of AMMI Analysis

Analysis of variance is the primary submodel of AMMI analysis and computes the additive main effects for genotypes and environments. The standard analysis of variance model for decomposing a two-way GE data is

$$y_{ij} = \mu + g_i + e_j + W_{ij}, \tag{3.8}$$

where the least square estimates of μ, g_i, and e_j are

$$\hat{\mu} = y_{..} \tag{3.9}$$

$$\hat{g}_i = y_{i.} - y_{..} \tag{3.10}$$

$$\hat{e}_j = y_{.j} - y_{..} \tag{3.11}$$

$$\hat{W}_{ij} = y_{ij.} - y_{i..} - y_{.j.} + y_{..} \tag{3.12}$$

Since μ, g_i, e_j and W_{ij} are all matrices, it means that g_i is a column centred matrix, e_j is a row centred matrix and W_{ij} is a doubly centred matrix.

3.2.3 Constraints of the ANOVA Model (Properties of Centred Matrices)

1. Row centred matrix has the sum of rows to be zero $\sum_{j=1}^{J} e_j = 0$
2. Column centred matrix has the sum of columns to be zero $\sum_{i=1}^{I} g_i = 0$
3. Doubly centred matrix has sum of both rows and columns to be zero

$$\sum_{i=1}^{I} W_{ij} = \sum_{j=1}^{J} W_{ij} = 0$$

3.2.4 Test of Significance

The significance of the components of the ANOVA model is based on F- test which is calculated according to Table 3.1

Table 3.1 ANOVA table for a two way data

Source of variation	Degrees of freedom (df)	Sum of squares (SS)	Mean SS	F values
Total	IJK- 1	SSTO = $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - y_{...})^2$		
Treatment	IJ – 1	SST = $K\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij.} - y_{...})^2$	$MST = \frac{SST}{IJ-1}$	$F_T = \frac{MST}{MSE}$
Genotype	I – 1	SSG = $JK\sum_{i=1}^{I}(y_{i..} - y_{...})^2$	$MSG = \frac{SSG}{I-1}$	$F_G = \frac{MSG}{MSE}$
Environment	J – 1	SSNV= $IK\sum_{j=1}^{J}(y_{.j.} - y_{...})^2$	$MSNV = \frac{SSNV}{J-1}$	$F_{NV} = \frac{MSNV}{MSE}$
Interaction	(I -1) (J -1)	SSN= $K\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2$	$MSN = \frac{SSN}{(I-1)(J-1)}$	$F_N = \frac{MSN}{MSE}$
Error	IJ(K -1)	SSE = $\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ijk} - y_{ij.})^2$	$MSE = \frac{SSE}{IJ(K-1)}$	

The F ratio of each source of variation (Treatment, Genotype, Environment and Interaction) is obtained by dividing the source's mean sums of squares by the MSE. The source is said to be significant at α-level when its calculated F– ratio is greater than the tabulated F value given the degrees of freedom, df, of the source of variation, $F_{(df,\alpha)}$

Though the residual may not be significant by ANOVA partitioning it may reveal significant sources within it (Gauch, 1988). Hence the residual matrix W becomes a very important factor after ANOVA which should be decomposed in order to help explain the interaction (multiplication) effect. We therefore proceed to estimate the parameters of the multiplicative components from the residual matrix W in the next phase.

3.2.5 Phase II of AMMI Analysis

The residual matrix W obtained after the analysis of variance is subjected to Singular Value Decomposition (SVD) in a similar way as Principal Component Analysis (PCA).

3.2.6 Singular Value Decomposition

Any mxn (m by n) matrix posses a basic structure which involves a triple product. If A is an m x n matrix, then there exists an integer r, $\delta_1 \geq \delta_2 \ldots . \delta_r > 0$, an m x m unitary matrix U, an n x n unitary matrix V, and an m x n matrix S, all of whose entries are zero except $S_{ii} = \delta_i$, i = 1, 2, ...r such that

$$A = USV^T \qquad (3.13)$$

S is represented as $S = \begin{bmatrix} \delta & 0 \\ 0 & 0 \end{bmatrix}_{m,n}$ where $\delta = \text{diag}\ (\delta_1, \delta_2, \ldots\ldots \delta_r)$.

The equation (3.13) is called singular value decomposition (SVD) of A and δ_i is called singular values of A.

$V = \{V_1, V_2 \ldots Vn\}$ is an orthomormal eigenbasis ($VV^T = I$) of A^TA and δ_i^2, i = 1 ....r are non zero eigenvalues of A^TA and $V_1 \ldots . V_r$ are corresponding eigenvectors, $V_{r+1} \ldots . V_n$ are eigenvectors corresponding to the eigenvalue 0.

Similarly $U = \{U_1, U_2 \ldots U_m\}$ is an orthonormal eigenbasis for AA^T, δ_i^2, i = 1, 2, . . . r are the non zero eigenvalue of AA^T and $U_1 \ldots U_r$ are corresponding eigenvectors, $U_{r+1} \ldots U_m$ are eigenvectors corresponding to eigenvalue 0.

The subscript r is called the rank of A and r = min (m,n). The rank of a matrix is the largest number of linearly independent rows or columns of a given matrix. The least square approximation of rank r to A can be found as $A_r = U_r S_r V_r^T$ where U_r and V_r^T are the first r columns of U and V^T respectively, and S_r contains the first r singular values of A (Eckart and Young 1936). For a doubly centred m x n matrix such that k = min (m, n), the maximum rank is (k – 1) (Gollob, 1968). With the consideration of the singular value decomposition and the least square approximation of a double centred matrix, we assume a least square approximation of the residual matrix W as

$$W = U_r S_r V_r^T \qquad (3.14)$$

The components of (3.14) have the same meaning as defined above. Geometrically, the least square approximation of the residual matrix (3.14) is meant to reduce its bloated dimensionality to its true dimensionality (which is equal to the rank of W) in which the data lie.

In other words, PCA which is the second sub model seeks to generate fewer new uncorrelated variables $\Upsilon_1\ \Upsilon_2 \ldots \Upsilon_n$ out of the old correlated variables of the doubly centred data matrix where

$$\Upsilon_i = \sum_{j=1}^{n} U_{ij} W_j \tag{3.15}$$

Where i = 1 . . . n and the coefficients are being chosen so that Υ_i accounts for the maximum variance for all i as well as satisfying the uncorrelated condition stated above. Thus for Υ_i to be uncorrelated with Υ_j for all $i \neq j$

$$E(\Upsilon_i \Upsilon_j) = E\left(\sum_{k=1}^{n} U_{ik} W_k \sum_{p=1}^{n} U_{jp} W_p\right) = 0$$

$$= \sum_{k,p=1}^{n} U_{jk} U_{ij} E(W_k W_p)$$

$$= \sum_{k,p=1}^{n} U_{jk} U_{ij} \Sigma_{kp} = 0$$

where Σ_{kp} is the covariance of W_k and W_p.

3.2.7 How the Coefficients U_{ij} Are Calculated

Firstly without some restriction, the coefficient cannot be calculated, simply because each coefficient could be multiplied by some constants and the variance of each Υ variable increased without limit. The usual constraint imposed to overcome this problem is that the sums of squares of the coefficient defining each component equal one. Thus

$$\sum_{k,p=1}^{n} U_{jk} U_{jp} = \begin{cases} 0, j \neq k \\ 1, j = k \end{cases}$$

This orthogonality condition geometrically indicates a rotation of the coordinate axis of the data set about the origin. Equation (3.15) is expressed in a matrix notation as below;

$$\mathcal{Y} = UW \tag{3.16}$$

where $\mathcal{Y}$ = column vector of new variable, W= column vector of original variable, U = the transformation (orthogonal) matrix. For covariance of $\mathcal{Y}$, we have

$$E(\mathcal{Y}\mathcal{Y}^T) = E\{UW(UW)^T\} = (UWW^TU^T) = U\Sigma U^T = \Lambda \tag{3.17}$$

where $\Sigma = WW^T$, Λ is a diagonal matrix with diagonal elements $\lambda_1 \geq \lambda_2 \geq \ldots\ldots \geq \lambda_n$ which are eigenvalues of Σ and the columns of U are their corresponding eigenvectors (For instance the environment index for the GE case). This result (3.17) allows us to write

$$\Sigma = U^T \Lambda U \tag{3.18}$$

the equation (3.18) above is called Spectral Decomposition of $\Sigma = WW^T$. Now Σ is a symmetric matrix and as such, the sum of its diagonal elements is equal to the sum of its eigenvalues.

Thus Trace $(\Sigma) = \sum_{i=1}^{n} \lambda_i$ (3.19)

This follows that the total variance of the original variables is equal to the sum of the eigenvalues. The eigenvalues are therefore considered as the variance of the derived variables. The presence of interaction corresponds to $\hat{\lambda}_i$ being non-zero.

The new variables are called interaction principal component analysis (IPCA) axes. The full model will therefore comprise n (maximum rank) IPCA axes. From (3.19) the ith IPCA axis accounts for a proportion

$$\frac{\lambda_i}{tr(\Sigma)} \text{ of the total variance of the interaction}$$

and the first k IPCA axes accounts for a proportion of

$$\frac{\sum_{i=1}^{k} \lambda_i}{tr(\Sigma)} \text{ where } \lambda_1 \geq \lambda_2 \geq \ldots\ldots \geq \lambda_n > 0$$

The complete AMMI model which integrates the ANOVA and PCA submodels (3.8) and (3.14) therefore is

$$y_{ijk} = \mu + g_i + e_j + \sum_{l=1}^{L} \delta_l U_{il} V_{jl} + \rho_{ij} + \varepsilon_{ijk} \quad (3.20)$$

It is subject to the constraints

$$\sum_{i=1}^{I} g_i = 0, \sum_{j=1}^{J} e_j = 0, \sum_{i=1}^{I} \rho_{ij} = 0, \sum_{j=1}^{J} \rho_{ij} = 0, \sum_{l=1}^{L} U_{il}^2 = 1, \sum_{j=1}^{J} V_{jl}^2 = 1 \quad (3.12)$$

where ρ_{ij} result when some of the IPCA axes are ignored. This occurs when the higher axes are believed to be non patterned (non-systematic variation in the data). ρ_{ij} is called noise which is made up of the non patterned axes together with the error term. It is this noise that complicates the explanation of the interaction which AMMI model is believed to be able to separate it (ρ_{ij}) from pattern, $\mu + g_i + e_j + \sum_{l=1}^{L} \delta_l U_{il} V_{lj}$.

U_{il} and V_{jl} are the l^{th} genotype and environment eigenvectors respectively due to the l^{th} singular value δ_l. From (3.14) $W = USV^T$ can be restructured as

$$W = US^{1/2}S^{1/2}V^T = (US^{1/2})(S^{1/2}V^T) = \Gamma\Psi \qquad (3.22)$$

Where $\Gamma = US^{1/2}$ and $\Psi = S^{1/2}V^T$.

Γ is called Genotype IPCA scores and Ψ is called Environment IPCA scores. AMMI generates a family of models and the simplest model, AMMI0 estimates only the additive main effects of genotypes and environments to explain the data matrix ignoring GE interaction. AMMI 1 considers main effects as well as one principal component axis (IPCA 1) to interpret the residual matrix. Similarly, AMMI 2 has two IPCA axes. In effect AMMI submodels derive their names from the number of the multiplicative components (IPCA axes) considered. The full model denoted by AMMI N considers all the N IPCA axes.

Practically the full model though fits the data perfectly, is believed to be excessive because of noise. The question as to how many axes to include is treated under AMMI model selection in the next section.

3.2.8 AMMI Model Selection

AMMI model selection deals with the number of IPCA axis that are considered significant – to regard as mostly pattern rather than noise. Several proposals on model selection have been tabled among which we have Gauch's (1988, 1992) proposal on predictive assessment in contrast to postdictive assessment.

1. Postdictive Criterion

A statistical model is constructed for a data set and success (in separating pattern from noise) is measured in terms of the model's ability to fit this data set with consideration of parsimony (reduced model with minimal degrees of freedom). The IPCA scores break the interaction degrees of freedom (I-1)(J -1) by assigning the n^{th} axis I +J + 1 – 2n (Gollob, 1968) degrees of freedom except for the last IPCA axis which has 1 fewer because of the constraints that the residual sum to zero. Postdictive evaluation of significance is by using F – test.

There is also another testing criteria proposed by Mandel (1969). Mandel's test differs from Gollob's by the way in which the interaction degrees of freedom are partitioned. Mandel suggests that degrees of freedom be assigned in such a way that under the hypothesis of no interaction the expected values of the mean

squares for each component should be equal to the variance of y_{ij} and he derives the appropriate degrees of freedom by simulation (Williams et al., 1993). The two approaches by Gollob and Mandel could lead to different conclusions from the same data. In a comparative study by Williams et al. (1993), their results support Mandel's conclusions.

2. Predictive Criterion

The data to be analysed are partitioned into;

i. Data used to construct a model (Modelling data)

ii. Data used to validate the model (Validation data).

Success is measured in terms of model's ability to fit the validation observation. Predictive evaluation of significance is by using sums of squared difference between predicted values and observed validation value (SSD)

3.2.9 AMMI Biplot

Biplots give a visual (two dimensional) representation of the SVD based on (3.22) of the AMMI model. Using subscripts on (3.22) we have.

$W_{ij} = \sum_{l=1}^{n} \gamma_{il}\tau_{jl}$, where γ_{il} and τ_{jk} are the genotype and environment scores such that γ_{il} and τ_{jk} are respectively plotted against $\gamma_{il'}$ and $\tau_{jk'}$ ($I^1 = I - 1$) to obtain the I^{th} ($I \neq 1$) AMMI biplot except for AMMI1 biplot where IPCA 1 scores, $\gamma_{il'}$ and τ_{jl}

are plotted against their respective mean (deviations). For instance AMMI 2 biplot is IPCA 2 scores against IPCA 1 scores.

The prefix bi therefore refers to the fact that genotypes and environments are both presented in the same graph.

3.2.10 Interpretation Rules of Biplot

One big problem is that interpretation of biplots of higher axes is quite difficult. AMMI 1 and AMMI 2 are however interpreted as follows;

AMMI 1 biplot

1. Points of same kind
    i. Displacement along abscissa indicates difference in main (additive) effects
    ii. Displacement along ordinate indicates difference in interaction effects
    iii. Displacement between abscissa and ordinate indicates both main and interaction effects
2. Clustered points of the same kind are said to be similar with respect to both main and interaction effects with dispersed points being dissimilar

AMMI 2 biplot

This graph is a useful supplement of AMMI 1 biplot when IPCA 2 is sizable and significant.

1. Points near origin have little or no interaction. These are well fit by the additive submodel, whilst points far from origin show high interaction pattern.
2. Clustered points have similar patterns whilst distant points have different interaction patterns.

CHAPTER FOUR
DATA ANALYSIS

4.0 Introduction

In the attempt to streamline seed production and sales, breeders have given more attention to the production of cultivars that are as stable and widely adaptable as possible. The identification of a mega-environment becomes paramount when adaptation is the goal. Nevertheless it is difficult to achieve this goal due to the fact that there is scarcely a genotype that 'wins' (is superior) every time in every place, basically because of genotype-environment (GE) interaction.

Plant breeders and agronomist therefore undertake series of trails of several genotypes in several environments (mulitilocation trial) to generate data, which eventually is analysed to facilitate guidance for selecting the best genotype or agronomic treatments for planting in future years and at new sites.

This chapter is devoted to the demonstration of analysis of a two-way GE yield data both in the classical and the modern perspectives.

Section 4.1 discusses the material used while sections 4.2 and 4.3 discuss some widely used methods for analysis – Analysis of variance, regression analysis and AMMI analysis respectively.

4.1 Materials and Methods

4.1.1 The Data

Crop Research Institute (CRI), Kumasi through Dr. S. Ohemeng Dapaah, provided the data, which will be used for the analysis in this chapter. They comprise a yield response of a trial of 10 genotypes in 9 environments over a one year (1999) period with 4 of the environments being dry areas whilst the rest are from relatively rainy areas. The trial is code-named station variety trial (SVT) and the data is provided at Appendix C. Table 4.1 shows the sites and site codes as well as genotypes and genotype codes.

Table 4.1: site and site code and genotype and genotype code of the SVT data

	SITE	CODE	GENOTYPE	CODE
1	KPEVE	KPE	DORKE SR	DR
2	DAMANGO	DMG	NAES	NS
3	NYANKPALA	NKP	SAFITA – 2	S2
4	WA	WAA	90 DYF	9F
5	MANGA	MNG	DODZI	D1
6	FUMESUA	FMS	90 DWD	9D
7	EJURA	EJR	DMR – ESR	MR
8	KWADASO	KDS	EV EJ91	E1
9	POKUASE	PKS	FU91 90DWPP	FP
10			LOCAL	LL

The genotypes were evaluated in a randomized complete block experiment with 4 replications in each experiment. Missing cultivar entries were omitted from the data set.

4.1.2 Software

Genstat software was used for all analysis with programs for analysis of variance, joint regression analysis and additive main effect and multiplicative interaction given at Appendix B.

4.2 Analysis Of Variance

The Analysis of variance (ANOVA) model for genotype by environment data (two way data) is usually represented by

$$y_{ijk} = \mu + g_i + e_j + w_{ij} + \varepsilon_{ijk} \qquad (4.1)$$

where y_{ijk} is the yield of the ith genotype in the jth environment and the kth replication, μ is the grand mean, g_i are genotype mean deviations (mean minus the grand mean), e_j are the environment mean deviations, w_{ij} are the ANOVA residual also the interaction of genotype and environment, and ε_{ijk} is the error term. Details of the ANOVA model and table are presented in Chapter Three (Section 3.3.1). Using model 4.1 the SVT data were analysed and the results are shown in Table 4.2.

4.2.1 Results of ANOVA

Summary of the analysis is presented in Table 4.2 below.

Table 4.2 Analysis of variance for the SVT data.

Source	df	SS	MS	F	Prob
Total	359	675500000	1881615.60		
Treatment	89	550367735	6183907.13	13.34	<0.001*
Genotype	9	55210000	6134444.44	13.24	<0.001*
Environment	8	442400000	55300000	119.32	<0.001*
Interaction	72	52757735	732746.32	1.58	0.005*
Error	270	125132265	463452.83	1.62	0.006*

4.2.2 Discussion (ANOVA)

From the analysis (Table 4.2) the main effects are significant together with the interaction at 1% level. The interaction being significant means the performance of a genotype differs across the environments. Meanwhile the sums of squares accounted by the various components of the ANOVA model is highest for Environment (80%), followed by genotype (10.3%) and Interaction (9.6%) which are almost equal. The trend follows the usual pattern whereby GE SS is relatively similar to that of the genotype

It therefore follows that though some genotypes showed relatively high mean yields, they cannot be selected as superior genotypes since the environment is found to exhibit a significant effect on the performance of the genotype (significant

interaction). Thus a serious deficiency of the ANOVA additive model is the limited information provided on the interaction (Gauch, 1988) since it does not partition the interaction for a complete analysis of the data.

Table 4.3. Means of genotypes across environments

Genotype	DR	NS	S2	9F	D1
Means	4805.8	4572.6	4371.6	4486.1	3596.5

Genotype	9D	MR	E1	FP	LL
Means	4454.3	4522.8	4884.9	4504.5	3750.8

In other words genotypes E1 followed by DR might perform extremely well in few environments, over-shadowing the poor performance in the other environments. Hence selection of superior genotypes based on mean performance is not recommendable.

We therefore have to proceed to partition the interaction to explore significant sources in it. Among the practical ways of looking at this is what is proposed by Finlay and Wilkinson (1963), linear regression analysis, for partitioning and analysing interaction.

Partitioning Interaction

4.3 Linear Regression Analysis

The linear regress model falls conveniently into two parts;

(i) Conventional analysis of variance

(ii) Partitioning of the interaction

The partitioning of the interaction results into 3 different regression sub models.

1. The genotype regression model

$$y_{ijk} = \mu + g_i + e_j + \beta_i \hat{e}_j + d_{ij} + \varepsilon_{ijk} \qquad (4.2)$$

2. The environment regression model

$$y_{ijk} = \mu + g_i + e_j + \alpha_j g_i + d_{ij} + \varepsilon_{ijk} \qquad (4.3)$$

3. Joint regression model

$$y_{ijk} = \mu + g_i + e_j + \kappa g_i e_j + d_{ij} + \varepsilon_{ijk} \qquad (4.4)$$

Equations (1), (2) and (3) together forms the complete regression model given as

$$y_{ijk} = \mu + g_i + e_j + \beta_i e_j + \alpha_j g_i + \kappa g_i e_j + d_{ij} + \varepsilon_{ijk} \qquad (4.5)$$

where the old variables retain their meaning from the ANOVA model whilst

β_i is genotype slope estimated by

$$\beta_i = \frac{\sum_{j=1}^{J} W_{ij} e_j}{\sum_{j=1}^{J} e_j^2} \qquad (4.6)$$

α_j is environmental slope estimated by

$$\alpha_j = \frac{\sum_{i=1}^{I} W_{ij} g_i}{\sum_{i=1}^{I} g_i^2} \qquad (4.7)$$

κ = joint regression coefficient estimated by

$$\kappa = \frac{\sum_{i=1}^{I} \beta_i g_i}{\sum_{i=1}^{I} g_i^2} \qquad (4.8)$$

and d_{ij} is residual of regression analysis.

The regression analysis is useful only when the initial analysis of variance clearly establishes the significance of the GE interaction (J. Hill, 1975).

4.3.1 Results of Linear Regression Analysis

Summary of the linear Regression Analysis of variance for SVT trial is presented below.

Table 4.4. Linear regression analysis of variance for SVT Trial

SOURCE	DF	SS	MS	F	PROB.
Total	359	675500000	1881615.60		
Treatment	89	550367735	6183907.13	13.34	<0.001*
Genotypes	9	55210000	6134444.44	13.24	<0.001*
Environments	8	442400000	55300000	119.32	<0.001*
Interactions	72	52757735	722746.32	1.58	0.005 *

Joint Regr.	1	300870	300870	0.65	0.421
Gen. Regrs.	8	6173951	771743.86	1.67	0.106
Env. Regrs.	7	4355274	622182	1.34	0.231
Residual	56	41927640	748707.86	1.62	0.006 *
Error	270	125132265	433452.83		

4.3.2 Discussion

The analysis of variance in table 4.4 indicates that regression is not effective for this particular data set. The genotype, environment and joint regression combined accounted for only 20.53% of the interaction SS, leaving a relatively large residual of 79.49% of the interaction. Furthermore, the average regression model MS is smaller compared to the interaction MS. Meanwhile none of the regressions (joint, genotype and environment regressions) as shown in Table 4.4 was significant at 0.05 level while the residual was highly significant. This suggests that the residual do better than the model and hence the model failure. Table 4.5 gives the linear regression model parameters for the SVT data. It lists the deviations, slopes and squared correlations r^2 for each genotype and environment.

Table 4.5. Linear regression model for the SVT data.

Summary of the linear regression model for the SVT data; Grand mean = 4395, Joint regression coefficient = -0.00006659

ENTITY		DEVIATION	SLOPE	R^2
GENOTYPES	DR	410.8	0.00882	0.00212
	NS	177.6	-0.1653	0.12831
	S2	-23.4	0.01481	0.00407
	9F	91.1	-0.05468	0.05794
	D1	-798.5	-0.1303	0.13025
	9D	59.3	-0.05434	0.0688
	MR	127.8	0.1213	0.18516
	E1	489.9	-0.0046	0.00087
	FP	109.5	-0.01978	0.00194
	LL	-644.2	0.2841	0.22468
ENVIRONMENTS	FMS	187.3	0.3012	0.09511
	EJR	1669.8	-0.5162	0.1539
	KPV	1111.1	0.3688	0.14486
	DMG	-1735.2	0.1278	0.02582
	PKS	1012.3	-0.083	0.00978
	KDS	52.3	0.1889	0.04198
	NKP	-53.1	-0.2095	0.07947
	WAA	-652.1	-0.I3599	0.1483
	MNG	-1592.3	0.1822	0.0244

Finally, considering the R^2 values in Table 4.5, for each genotype and environment, no entity is found to fit well as all R^2 values are less than 0.3. The complete model is therefore ineffective due to its failure to explain significant part of the interaction.

The Additive Main effect and Multiplicative Interaction, or AMMI, model was discussed by Bradu and Gabriel (1978) and has proved valuable in genotype by environment interaction studies (Gauch, 1988; Gauch and Zobel, 1988; Zobel et al, 1988). This model is considered in the next section.

4.4 Additive Main Effect and Multiplicative Interaction (AMMI) Analysis

For detailed description of appropriate statistical fixed-effect models for AMMI and sub cases, refer to chapter 3.

The model is

$$y_{ijk} = \mu + g_i + e_j + \sum_{l=1}^{L} \delta_l U_{il} V_{lj} + \rho_{ij} + \varepsilon_{ijk} \qquad (4.9)$$

where y_{ijk}, μ g_i, e_j and ε_{ijk} retain their meaning as in the ANOVA model.

ρ_{ge} are the AMMI residuals, L is the number of SVD (singular value decomposition) axes retained in the model, δ_l is the singular value for SVD axis l, U_{il} are the genotype singular vector value for SVD axis l and V_{lj} are the environment singular values for SVD axis l.

F-test was used to determine the significance of a given SVD axis (Guach, 1992). The goal of the analysis is to summarize the interaction SS with a few SVD axes (typically, I = 1 to 2), leaving a reduced model with residuals containing mostly noise. The genotype interaction scores ($\delta^{0.5}U_i$) and the environment interaction scores $\delta^{0.5}V_j$ are in units which are the square root of the unit for y.

4.4.1 Results of AMMI Analysis

Table 4.6. Summary of AMMI 3 Analysis of variance for SVT trial

SOURCE	DF	SS	MS	F	PROB.
Total	359	675500000	1881615.60		
Treatment	89	550367735	6183907.13	13.34	<0.001*
Genotypes	9	55210000	6134444.44	13.24	<0.001*
Environments	8	442400000	55300000	119.32	<0.001*
Interactions	72	52757735	722746.32	1.58	0.005*
IPCA 1	16	23312060	1457003.81	3.14	<0.001*
IPCA 2	14	13028605	930614.64	2.01	0.017**
IPAC 3	12	7247064	603922	1.30	0.218
Residuals	30	9172270	305742.33	0.66	0.914
Error	270	125132265	463452.83		

4.4.2 Discussion

From table 4.6, IPCA 1 has a SS of 23312061 which is 44.2% of the interaction. In addition it has a root mean square (RMS) residual of 6.5% of the grand mean of 4395 kg/ha. This is significant at 1% level. The IPCA 2 is significant at 1.7% level. The IPCA 2 together with the IPCA 1 accounted for 68.9% of the interaction SS with RMS residual of 4.9% of grand mean. This indicates a fairly good performance of the model as it is able to explain a significant part of the interaction. The full model AMMIF though is AMMI 8 and fits the data perfectly is excessive since the IPCA3 is found to be insignificant at 5% level.

4.4.3 Biplot and Interpretation

Biplot was introduced by Gabriel (1971). It involves plots which show both genotypes and environments simultaneously in the same graph. Biplots are valuable for displaying patterns is multilocational trial data. The abscissa of AMMI1 biplot shows means while the ordinate show the IPCA1 scores that capture interaction effects. Refer to Chapter Three for additional information on biplots. Table 4.8 gives AMMI2 model for the SVT trial showing the genotype and environment deviations with their corresponding IPCA1 and IPCA2 values.

Table 4.8 AMMI2 model for the SVT data

ENTITY		DEVIATION	IPCA 1	IPCA 2
GENOTYPES	DR	410.8	-4.58	-4.93
	NS	177.6	26.38	10.55
	S2	-23.4	-4.37	-7.13
	9F	91.1	-1.35	-6.32
	D1	-798.5	19.47	11.41
	9D	59.3	7.93	5.95
	MR	127.8	-4.58	-3.86
	E1	489.9	1.01	0.76
	FP	109.5	-5.54	-29.45
	LL	-644.2	-34.38	23.02
ENVIRONMENTS	FMS	187.3	-14.04	-20.46
	EJR	1669.8	-26.29	16.29
	KPV	1111.1	19.35	-11.63
	DMG	-1735.2	0.14	-10.69
	PKS	1012.3	-6.23	10.45
	KDS	52.3	-16.01	-12.05
	NKP	-53.1	8.84	1.77
	WAA	-652.1	7.34	24.73
	MNG	-1592.3	26.91	1.59

Fig 4.1 AMMI 1 biplots for genotypes and environments in the SVT trial

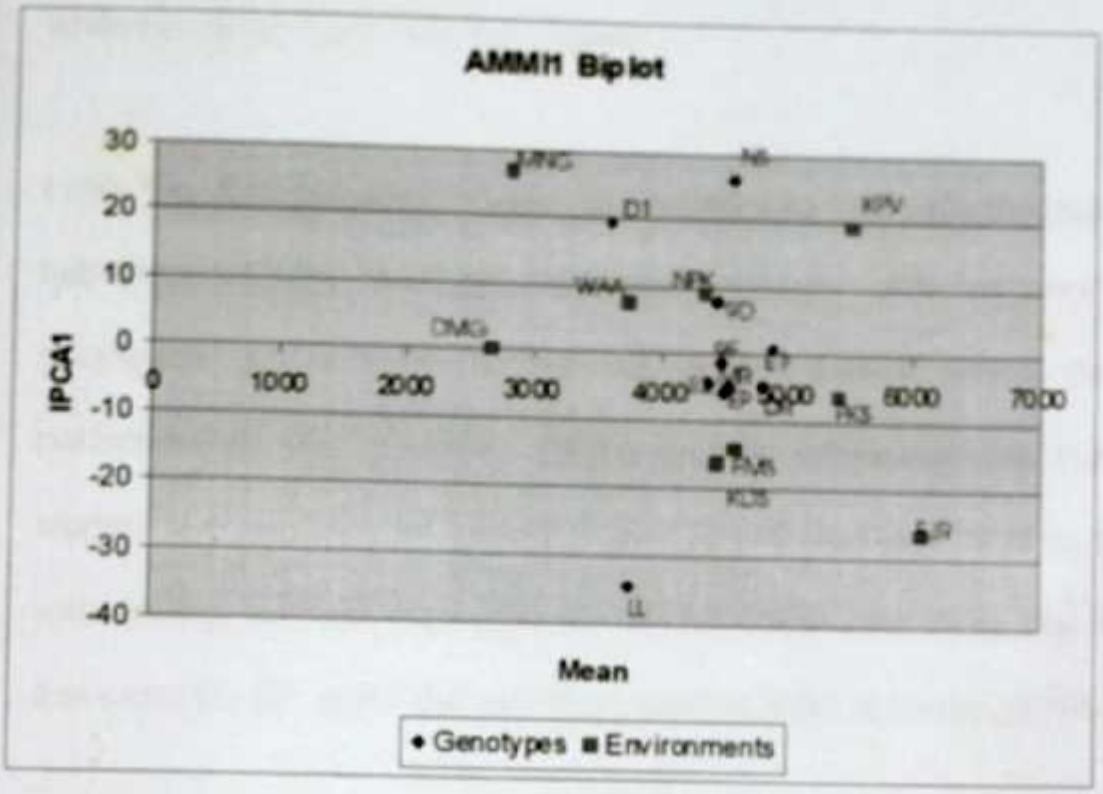
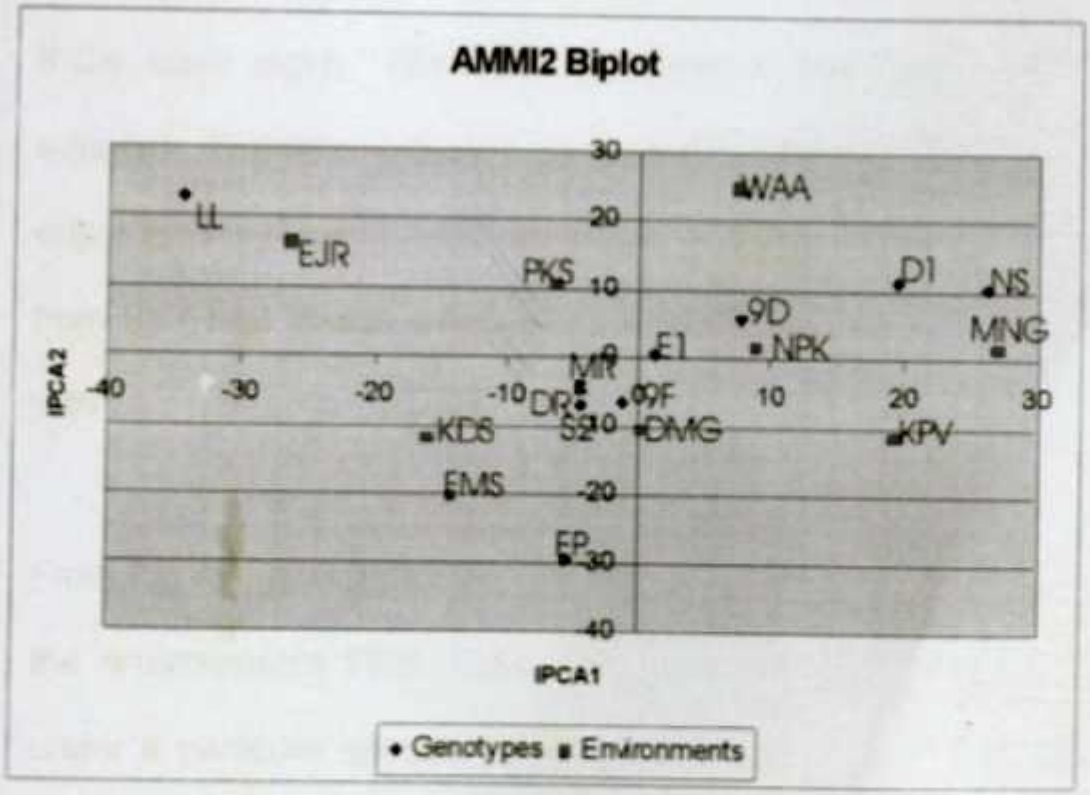


Fig 4.2 AMMI 2 biplots for genotypes and environments in the SVT trial



To draw attention to location and genotypes, all three-letter codes are locations while two letter codes are genotypes.

From Fig 4.1, the entire space is divided into two with the top and the bottom subspaces having opposite interaction scores. For instance considering the genotypes, E1 is a top ranked cultivar and overall winner (in terms of mean performance) and LL differs clearly in their interaction (as they have opposite signs). The ordinates of AMMI2 (Fig. 4.2) indicates lack of fit to the AMMI1 model with entities farthest from zero on IPCA2 being poor fit by the AMMI1 model. In this case LL, EP and WAA are described as least fit by the AMMI1 model.

Adaptation

Genotypes are said to be better adapted to environments if both have the same IPCA score signs. Genotypes classified in this sense are said to be narrowly adapted. Therefore breeding genotypes having superior performance through the entire growing region (as indicated by ANOVA) is inconsistent with these results from SVT trial. Broad adaptation (universal win in all environments) is associated with zero interaction scores.

From Fig 4.1, genotypes 9F, S2, MR, DR, EP and LL are said to be adaptable to the environments PKS, FMS, KDS and EJR. Actually these environments fall under a particular geographical zone whilst genotypes E1, 9D, D1, and NS are adaptable to another geographical zone (MNG, DMG, WAA, NPK, KPV).

Similarities

Genotypes 9F and 9D are said to be similar with respect to main effect while MR and DR are similar in terms of their interaction effects from Fig. 4.1. Additionally genotypes S2, MR, EP and DR are considered similar with respect to both main and interaction effects. On the other hand the environments FMS and KDS also appear similar in terms of both main and interaction effect.

In fact from Fig.4.2, the AMMI2 biplot, points near origin have little interaction and for that matter E1 together with MR and DR are said to be well fit by the additive submodel. Near points on the AMMI2 biplot are also as having similar interaction pattern. For instance MR, DR, S2, and 9F, have similar interaction pattern which conforms to the description by the AMMI1 biplot.

Stability

Stability refers to consistency in response of a genotype across environments. Genotypes with interaction scores near zero (Fig 4.1), for instance E1 and 9F do not interact much with environments, and therefore their rank orders are relatively stable whilst LL and NS have a poor (relatively) stability index.

Interaction

The percentage of the treatment SS captured by an AMMI biplot is a useful statistic for assessing the overall goodness of fit. From table (ANOVA), the AMMI

1 model captures an SS of 520922060 which is 94.6% of the treatment SS. Hence the graph captures the yield trial's structure quite well.

Maize quality performance is better for genotypes growing in environments with large scores of the some sign whereas poor quality performance is expected with large scores of opposite signs. Therefore a genotype such as E1 performs better in KPV compared to NKP. NS and DR perform well in KPV and EJR respectively. LL is also better performed in EJR.

Yield Reliability

Practically, high yield stability may be associated with low mean yield, (or low stability with high mean yield) which complicates selection or recommendation. The practical objective of combing high levels of mean yield and yield stability has led to the concept of yield reliability.

A reliable genotype is characterised by consistently high yield across environments. The use of a yield reliability index facilitates genotype selection or recommendation.

According to this work E1 will be declared reliable (relatively) essentially for the top environments and generally for all environments since it is observed to have high mean yield with a relatively good stability index. D1, NS and LL on the other hand are considered as relatively less reliable.

CHAPTER FIVE

EXTENSION OF AMMI MODEL

5.0 Introduction

In chapter 2, we discussed a method for the analysis of GE data which was proposed originally by Mandel (1971); additive main effect and multiplicative interaction model. Currently, it has received much attention in the GE context. The model was promoted by Gauch (1988) who gave a general descriptive name to such models with additive and multiplicative components, "AMMI".

In section 5.1 of this chapter a review of the AMMI model is presented and the fitting of the model extended in section 5.2 by proposing an alternative method for estimating an environment index vector to overcome the limitations associated with AMMI environment index. The old and the proposed fitting procedures are compared in section 5.3 whilst the best ratio of the subgroups for the complementary process is considered in section 5.5. An index vector for analysing the entire data follows in section 5.6 which ends with a summary of the complementary procedure. Finally an overall summary of the chapter is presented in section 5.7.

5.1 Review Of The AMMI Model

AMMI is a straight forward extension of the PCA and one of a range of methods which have been found useful for extracting information on genotype by environment interaction.

Gauch (1988) has shown through validation studies that AMMI gives more precise estimates of genotype yield within locations than means across replicates. However, a number of authors are beginning to report situations where AMMI models may not be appropriate for GE studies. For instance, from a biological standpoint Crossa et al (1991) in discussing the benefits of the AMMI model also indicated that relying solely on AMMI model to describe GE interaction may lead to important interactions from a single environment with a unique stress being overlooked. Romagosa et al (1993) reported an analysis of a homogeneous subset of data indicating more complex pattern than the complete data set.

In a similar study, Piepho (1994) conducted a research where the predictive accuracy of the AMMI model was compared to that of The Best Linear Unbiased Predictions (BLUP) using Faba bean data sets from 1985 to1989. His results suggested that for all the five data sets the best AMMI model had many axes. In the 1987 and 1988 data sets, the full model was the best of all the AMMI models. These results are in contrast to many other results, where AMMI0, AMMI1 or AMMI2 were identified as the most accurate models (see e.g. Gauch 1988, 1992;

Crossa et al. 1990). Essentially, the model fitting process for the AMMI model requires more investigation.

A problem that is often overlooked by users of AMMI models is that the procedure for fitting the multiplication terms is strongly "optimistic".

The optimism is due to the structure of the matrix of interaction residuals $\underline{W}(y_{ij} - y_{i..} - y_{.j.} + y_{...})$ after fitting the additive terms which exhibits high correlations between genotypic vectors over environments. Consequently, the first eigenvector from $\underline{W}\underline{W}^1$ which is used as an environment index will tend to be correlated with the highly correlated terms of the GE matrix. This will result in sums of squares for fitting the multiplicative term (s) which may be sent directly from the latent root proportions of explained variation being larger than they would be if the terms of the matrix $\underline{W}$ were uncorrelated.

Simulation result of R^2 values from completely null cases (Mandel, 1971) have been produced on which to base the formal tests from the first multiplicative terms.

To illustrate the presence and effect of correlations between terms of GE matrix, I analysed a data set consisting of seven genotypes of soybean in ten environments with four replications and a "fictitious" data set comprising five genotypes in four environments with three replications (both from Gauch, 1992).

Table 5.1 shows sums of squares explained by regressing each genotype's vector of the residuals from the main effects model on the environment index (eigenvector).

Table 5.1 Sums of squares explained by genotype vectors of residual interaction matrix from the analysis of yield data consisting of 10 environments and 7 genotypes with 4 replications (Gauch, 1992)

Genotype	1	2	3	4
SS	7,180,752	12,077,914	19,103	1,255

Genotype	5	6	7	Total SS
SS	5,753,419	3,959,715	3,764,101	32,756,558

From Table 5.1, genotype 2 has the highest sums of squares. Four other genotypes showing relatively high sums of squares are genotypes 1, 5, 6, and 7. The high sums of squares explained by the five genotypes are less surprising if the correlations (Table 5.2) between the terms of the matrix and the first eigenvector are compared.

Table 5.2 Correlations among the seven genotypic vectors (G1-G7) and with the environmental eigenvector (ef) of the soybeans data (Gauch, 1992)

	1	2	3	4	5	6	7	
G1	1.00							
G2	0.92	1.00						
G3	-0.03	0.00	1.00					
G4	0.05	0.01	-0.31	1.00				
G5	-0.81	-0.81	-0.45	0.09	1.00			
G6	-0.94	-0.93	0.17	-0.05	0.77	1.00		
G7	-0.88	-0.92	0.19	-0.31	0.58	0.86	1.00	
ef	-0.98	-0.99	-0.09	-0.04	0.86	0.95	0.89	1.00
	G1	G2	G3	G4	G5	G6	G7	ef

Apparently, the first eigenvector bears closer resemblances to the terms of the residual interaction matrix which are highly intercorrelated than the terms which are uncorrelated resulting in relatively large sums of squares for genotypes 2, 1, 5, 6 and 7. The overall proportion of the interaction variation explained by the first IPCA axis of the soybean data is 82.4%. In this instance, the GE data will be considered as best described by AMMI model with one multiplicative term while predictions from the analysis may be less reliable due to the terms of the GE matrix being correlated.

Table 5.3 Sums of squares explained by genotype vectors of residual interaction matrix from the analysis of yield data consisting of 5 environments and 4 genotypes with 3 replications (Gauch, 1992)

Genotype	1	2	3	4	5	Total SS
SS	17,280	0	270	2,430	4,320	24,300

Table 5.4 Correlations among the five genotypes (G1-G5) vectors and with the environmental eigenvector (ef) of the fictitious toy data (Gauch, 1992)

	1	2	3	4	5	ef
G1	1.00					
G2	0.00	1.00				
G3	-1.00	0.00	1.00			
G4	-0.81	-0.58	0.81	1.00		
G5	-0.93	0.37	0.93	0.54	1.00	
ef	1.00	0.00	-1.00	-0.81	-0.93	1.00
	G1	G2	G3	G4	G5	ef

It is also clear from Table 5.3 and 5.4 that genotype 1 together with genotypes 5, 4 and 3 which pulled high SS were relatively highly intercorrelated. On the other hand it is not surprising that genotype 2 approximately pulled no SS considering its relatively weak correlations (Table 5.4) with the other genotypic vectors including

the environment eigenvector (r = 0). Invariable the GE data will be considered as best described by AMMI model since the first IPCA axis was able to capture as large as 92.3% of the total interaction SS.

The characteristic feature of the first eigenvector being dependent on the size of the correlations between terms of GE matrix being analysed may explain the causes of the increased sums of squares in AMMI models.

Additional Tables on Genotypic SS with their respective correlations which follow the same trend of high correlated genotypic vectors resulting in high SS is provided at Appendix A.

5.2 Extending AMMI Analysis By Using Alternative Index Vectors

In the last section we demonstrated that there are almost always strong indications that the nature of the residuals after fitting the additive main effect inevitably produces a matrix which tends to be correlated. Hence sums of squares obtained from regression on the environment index estimated from the correlated matrix are "optimistic". In this section we propose an alternative index vector which because they are uncorrelated to the genotype vectors of the residual interaction matrix to be analysed have no mechanism for introducing optimism in the sums of squares. Using analysis of variance notation, it is customary to write a non-additive model

$$y_{ijk} = \mu + g_i + e_j + \sum_{l=1}^{L} \delta_l U_{il} V_{lj} + \rho_{ij} + \varepsilon_{ijk} \qquad (5.0)$$

as $y_{ijk} = \mu + g_i + e_j + W_{ij} + \varepsilon_{ijk}$ (5.1)

where W_{ij}, the genotype by environment interaction term is a functions of two factors. Given that we have an environment index (θ_j), we can calculate the regression dependence of the genotype – environment interaction terms by regression W_{ij} on θ_j for each genotype, obtaining regression coefficient β_i.

Effectively, we are fitting the model

$$W_{ij} = \beta_i\theta_j + \delta_{ij} \quad (5.2)$$

where δ_{ij} represents the random component. By combing equations (5.1) and 5.2), the full model can be rewritten as an additive multiplicative mixture.

$$y_{ij} = \mu + g_i + e_j + \beta_i\theta_j + \delta_{ij} \quad (5.3)$$

Many other models are special cases of this general model and the model is also equivalent to the AMMI models of Gauch (1988) discussed under chapter 3. Essentially we have substituted U_{ik} and δ_k in (5.0) by β_i and V_{jk} by θ_j in (5.3) which leave us with few parameters than (5.0) with effects of β_i being orthogonal but not normalized. The full constraints for equation (5.3) are therefore shown below.

$$\sum_{j=1}^{j} \theta_j^2 = 1$$

$$\sum_{j=1}^{j} \theta_j = 0$$

$$\sum_{i=1}^{b} \beta_i = 0$$

Since we do not know the appropriate form of the environment index for (5.3) we can consider alternative choice for θ_j. An alternative form of environmental index called complement index vector is proposed. The index will be used as an environmental index vector which will be compared with the environment index vector (self), which is the first eigenvector of the residual interaction matrix.

5.2.1 Self Index Vector

The self index vector θ_s is the first eigenvector of the Matrix $\underline{WW}^1$. The regression of the terms of $\underline{W}$ on $\underline{\theta}$s is the AMMI analysis.

5.2.2 Complement Index Vector θ_c

The Complement Index Vector θ_c is defined as a vector of values estimated by splitting the total number of genotypes into two (usually unequal) subgroups (eg. In proportion 1:3 etc). One subgroup constitutes the actual genotype by environment interaction and the complement index vector θc, is estimated as the first eigenvector from a matrix of the other group.

For example, consider a trial consisting of 40 genotypes by 12 environments by 3 replications. We can arbitrarily split the data into say 2 sets A and B, where A consists of 10 genotypes by 12 environments by 3 replications and B is 30 genotypes by 12 environments by 3 replications.

We can analyse the first set of genotypes (set A) using the first eigenvector obtained from the second set (Set B). In this case the set A is called the target set, the set to be analysed. By this approach we estimate an environment index completely unrelated to the genotypes whose GE interactions are analysed.

Effectively, this is a form of cross validation technique whereby one set of the data is used for the model while the remaining set is reserved for estimating the index vector. We supposed that if an environment index θ_j has a good predictive value we will expect it to be an effective index even when applied to a different set of genotypes in the same environment. By transferring indices, we seek to eliminate the optimism in the estimation of the index.

The complement index technique is similar to the technique proposed by Malter and Caligari (1974) to overcome the problem of non-independence in the joint regression analysis. They regressed the yield of each genotype on the rest. However, this does not supply only one estimate of θ_j, and therefore the procedure was rather less attractive.

5.3 Comparing Self and the Complementary Analyses

The performance of each of the two index vectors was assessed using a model with a single multiplicative interaction term. The data used for the assessment was yield data from a set of wheat trials obtained from International Maize and Wheat Improvement Centre, Mexico (CIMMYT). The full set consisted of 40 genotypes by 45 environments by 3 replications but only 40 genotypes by 9 environments by 3 replications will be used for this assessment.

The set of 40 genotypes by 9 environments by 3 replications was randomly split into 10 target sets of 10 genotypes by 9 environments by 3 replications, and the two index vectors were estimated for each of the ten target sets as explained in the previous section. Thus for the analysis based on the complement index vector, a target set was analysed using an index vector estimated from 30 genotypes by 9 environment by 3 replications, while in the case of the self index vector, a target set was analysed using an index vector estimated from the target set.

For each target set, the proportion of the interaction variation explained from regressing each genotype's vector of the residual interaction matrix on each of the corresponding environment index vector was calculated and average R^2 value was computed for the set of ten target sets. The analysis was repeated for different target sets of sizes 20 and 30 genotypes and the performance of the 2 index vectors in terms of R^2 values are shown in Table 5.5.

Table 5.5: Average proportion of the interaction variation explained by two environment index vectors from the analysis based on subsets of a dataset consisting of 9 environments by 40 genotypes with 3 replications (subsets of the CIMMYT data for 1986-1987).

Target set	Mean R^2 values			Standard deviations		Coefficient of variation	
	Self	Complement	Mandel	Self	Complement	Self	Complement
10	43.4	16.0	39	6.6	7.9	15.2	49.4
20	37.0	19.8	30	4.0	2.4	10.8	12.1
30	34.4	19.9	19	1.6	5.3	4.7	26.6
Mean	38.3	18.6	29.3	4.1	5.2	10.2	29.4

The fourth column of Table 5.5 list expected R^2 values for corresponding target sets sizes for Mandel's simulation (Mandel, 1971) which may be read directly from Mandel's Tables.

If we compare average R^2 values based on self index vector with corresponding values from Mandel's simulated values for the three different target set sizes, then there is invariably good evidence that the interaction is non zero. This evidence points to the conclusion that there are almost always strong indications of genotype by environment interactions.

On average regressing on complement index vectors accounted for 18.6% respectively. This indicates a dramatic reduction in the sums of squares compared to values obtained from using the self index vector (average of 38.3%). The apparent differences between the self index vector and the alternative index vector in terms of R^2 values illustrate the effect of correlations between the terms of the matrix being analysed and the corresponding environment index vector.

If we compare the R^2 values based on the complement index vector with "expected "R^2 values expressed as ratio of the degrees of freedom of the fitting sum of squares with (I-1) df to the degrees of freedom of interaction sum of squares ((I-1) (J-1)), where I and J are number of genotypes and environments respectively, then the alternative index vector can be predictively useful.

5.4 Correlations of Eigenvectors

The complementary analysis allows for a random selection of target set whose complement is used to estimate the environment eigenvector for an eventual analysis of the target set. An important question however, is that can the selected genotypes have a significant influence on the estimate of the environment eigenvector for a particular target set in a given set of environments?

This is investigated by using the CIMMYT data to compute the correlations among 10 environment eigenvectors estimated from ten random samples of 30 genotypes to analyse their respective target sets of 10 genotypes (Table 5.6). The analysis is repeated for target set of sizes 20 and 30 genotypes (Tables 5.7 and 5.8).

Table 5.6 Correlation among ten eigenvectors estimated from ten random samples with each eigenvector estimated from 30 randomly selected genotypes in 9 environments.

Eigenvectors	Correlations (r)									
	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.99	1.00								
3	-0.89	-0.83	1.00							
4	-0.97	-0.94	0.97	1.00						
5	0.95	0.92	-0.97	-0.99	1.00					
6	0.99	0.97	-0.94	-0.99	0.98	1.00				
7	-0.97	-0.95	0.88	0.95	-0.92	-0.96	1.00			
8	-0.87	-0.81	0.97	0.96	-0.96	-0.92	0.83	1.00		
9	0.95	0.97	-0.72	-0.85	0.85	0.91	-0.88	-0.69	1.00	
10	-0.90	-0.94	0.67	0.81	-0.81	-0.87	0.79	0.64	-0.98	1.00

Table 5.7 Correlation among ten eigenvectors estimated from ten random samples with each eigenvector estimated from 20 randomly selected genotypes in 9 environments.

Eigenvectors	Correlations (r)									
	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.97	1.00								
3	0.74	0.79	1.00							
4	0.94	0.99	0.85	1.00						
5	0.84	0.89	0.95	0.92	1.00					
6	-0.92	-0.82	-0.72	-0.80	-0.79	1.00				
7	0.87	0.82	0.84	0.81	0.87	-0.95	1.00			
8	0.84	0.84	0.92	0.85	0.94	-0.88	0.97	1.00		
9	0.87	0.82	0.84	0.81	0.87	-0.95	1.00	0.97	1.00	
10	0.74	0.68	0.87	0.72	0.80	-0.87	0.90	0.90	0.90	1.00

Table 5.8 Correlation among ten eigenvectors estimated from ten random samples with each eigenvector estimated from 10 randomly selected genotypes in 9 environments.

Eigenvectors	Correlations (r)									
	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	-0.80	1.00								
3	-0.66	0.80	1.00							
4	-0.90	0.72	0.49	1.00						
5	0.86	-0.88	-0.65	-0.93	1.00					
6	-0.81	0.89	0.93	0.75	-0.87	1.00				
7	-0.81	0.95	0.92	0.73	-0.88	0.99	1.00			
8	-0.82	0.85	0.92	0.76	-0.85	0.99	0.97	1.00		
9	-0.93	0.65	0.62	0.93	-0.84	0.80	0.75	0.82	1.00	
10	-0.93	0.95	0.75	0.88	-0.96	0.90	0.94	0.89	0.83	1.00

From the above results (Tables 5.6 to 5.8), we note that the correlations are generally very large ($>|0.8|$). These results suggest that the set of genotypic yield values used to estimate the environment eigenvector has insignificant effect on the estimate. Thus for any given target set a particular complementary estimate of the environment index can be used to analyse different randomly selected target sets

of the same complementary size . There is the need however, to identify the target set whose complementary environment eigenvector gives the best predictions. The next section treats this issue.

5.5 The Optimum Target Set

Since the complement index vector is an externally provided index, it has a useful predictive value which can be investigated further. We can therefore as a next stage, identify a target set for the best prediction from the complement set. This involves varying the sizes of the target sets in order to generate different data matrix sizes for estimating complement index vectors.

The CIMMYT data (1986-1987) was once again used. The entire data was randomly split up with respect to the environment into the following subgroups;

i. 40 genotypes by 9 environments by 3 replications

ii. 40 genotypes by 18 environments by 3 replications

iii. 40 genotypes by 27environments by 3 replications

iv. 40 genotypes by 36 environments by 3 replications

Using each subgroup (case), target set sizes were chosen so that complement set sizes were 2, 4, and 6, up to 38 genotypes and analysed in the complementary sense. For example a target set of size 2 genotypes involved the use of two randomly selected genotypes to calculate the residual interaction matrix and the complement index vector estimated from the remaining 38 genotypes. For each

target set, 10 random samples of different combinations of genotypes were generated. The mean and variance of R^2 values for each target set are presented for the different sets of environments in Tables 5.9 -5.12.

Table 5.9. The mean and variance (s_d^2) of the proportion of variation explained (R^2 values) by 19 target sets from the analysis of a data set with 40 genotypes by 9 environments (subsets of the CIMMYT ISWYN data for 1986-1987).

Target set size	2	4	6	8	10	12	14	16	18
Mean R^2 values	20.9	12.6	19.9	24.5	14.4	21.6	17.9	19.1	18.8
Variance s_d^2	921.7	36.7	159.8	40.4	40.7	37.6	25.2	32.1	20.7

Target set size	20	22	24	26	28	30	32	34	36	38
Mean R^2 values	16.8	13.0	14.7	14.2	14.7	16.3	16.6	15.1	16.6	11.7
Variance s_d^2	29.3	15.4	12.3	25.5	20.1	10.0	16.6	15.9	17.6	12.1

Table 5.10. The mean and variance (s_d^2) of the proportion of variation explained (R^2 values) by 19 target sets from the analysis of a data set with 40 genotypes by 18 environments (subsets of the CIMMYT ISWYN data for 1986-1987).

Target set size	2	4	6	8	10	12	14	16	18
Mean R^2 values	9.7	21.0	20.2	19.8	16.5	18.5	14.7	16.8	17.1
Variance S_d^2	187.4	207.4	165.6	43.7	31.4	35.4	16.6	9.9	9.1

Target set size	20	22	24	26	28	30	32	34	36	38
Mean R^2 values	14.8	13.9	13.5	12.1	11.6	11.7	12.5	11.6	11.0	9.6
Variance s_d^2	9.2	17.9	2.9	6.1	7.6	14.4	6.0	20.1	18.3	16.3

Table 5.11. The mean and variance (s_d^2) of the proportion of variation explained (R^2 values) by 19 target sets from the analysis of a data set with 40 genotypes by 27 environments (subsets of the CIMMYT ISWYN data for 1986-1987).

Target set size	2	4	6	8	10	12	14	16	18
Mean R^2 values	11.7	19.9	21.3	22.6	19.0	21.6	14.7	19.0	20.8
Variance s_d^2	274.4	198.8	102.0	22.0	38.2	11.3	36.6	15.5	5.8

Target set size	20	22	24	26	28	30	32	34	36	38
Mean R^2 values	18.4	17.0	15.2	14.8	14.4	14.9	9.6	11.7	11.3	8.6
Variance s_d^2	4.6	34.5	16.2	9.6	18.5	9.2	24.0	23.3	22.4	23.0

Table 5.12. The mean and variance (s_d^2) of the proportion of variation explained (R^2 values) by 19 target sets from the analysis of a data set with 40 genotypes by 36 environments (subsets of the CIMMYT ISWYN data for 1986-1987).

Target set size	2	4	6	8	10	12	14	16	18
Mean R^2 values	12.1	18.8	19.2	18.5	17.3	19.0	14.8	15.3	16.9
Variance s_d^2	218.7	110.3	42.0	10.6	33.4	15.6	12.2	6.2	3.5

Target set size	20	22	24	26	28	30	32	34	36	38
Mean R^2 values	15.9	15.0	14.0	13.2	11.5	12.1	10.9	9.9	9.8	7.0
Variance s_d^2	4.2	17.3	23.5	10.6	8.9	10.8	10.9	18.1	13.2	16.6

The results show that the proportion of the variation explained is very variable for target sets of sizes less than eight genotypes as it is indicated by high s_d^2 values. For target set sizes of such few genotypes the regression fits are unduly influenced by influential or dominant genotypes within the target sets. There were few genotypes we classified dominant whose presence in the target set improved the proportion of the variation explained in every case.

The R^2 values can therefore be very high or low depending mainly on the number of dominant genotypes in the target set. These results suggest that although parameter estimates from complement index vectors are unbiased estimates, unreliable estimates can be obtained when the target set is based on too few

genotypes. The behaviour of the dominant genotypes was also observed when target sets included more of the genotypes (target set >> complement set) than the number used for estimating the complement index vectors. This is indicated by lower R^2 values for target sets of greater than twenty genotypes. As far as the optimum is concerned, the distribution of the mean R^2 values over really optimal target set sizes is generally flat with slight decline after 16-20 genotypes. It is reasonable to use half of the number of genotypes as the target set with half for estimating the complement index vector.

5.6 Average Complement Index Vector

The complement index vector was proposed for analysing subset of the GE data matrix due to the high correlations between the terms of the residual interaction matrix. If the full data set is to be analysed it is not possible to estimate its complement index vector in the same framework as described in the previous sections because it has no complement set. The objective of this section is to extend the method for estimating CI vectors in order to estimate an index vector for analysing the full data set. The problem with the use of CIV is that we end up with a set of vectors (two) rather than a single vector which gives overall measure of the index.

However, we observe from the analysis of different target sets presented in tables 5.6-5.8 that, correlations between the different complement index vectors from the same set of environments were relatively high ($\geq |0.8|$) for a uniformly distributed

set. This suggest that we can obtain a single index vector by averaging sets of environment index vectors from the same sets of environments which can be used for analysing complete datasets. We refer to this vector as the average complement index vector. The average of the sets of the complement index vectors is based on the assumption that the vectors have the same direction.

The procedure for obtaining an average complement index vector is as follows;

1. Split the data into two equal subgroups (target sets)
2. Calculate the CIV (θ_c) for each subset
3. Ensure that the CIVs have the same directions.
4. Calculate the average (arithmetic) CIV (θ_c) from the CIVs.

5.6.1 Evaluation of the Average Complement Index Vector

The performance of the average complement index vector is assessed using a model with a single multiplicative interaction term. The data used for the assessment was the CIMMYT wheat data consisting of 40 genotypes by 45 environments by 3 replications. The data is randomly divided into two subgroups with respect to genotypes to obtain a subgroup A of 20 genotypes by 45 environments by 3 replications ($A_{20\times 45\times 3}$) and its complement $\bar{A}_{20\times 45\times 3}$.

The average of the two complement index vectors from the two subgroups was used to analyse each of the subgroups A and $\bar{A}$ as well as the entire data set. Additional self analyses for all the three groups A, $\bar{A}$ and $A \cup \bar{A}$ are carried out for comparison. This analysis is repeated ten times using different randomly selected

subgroups (target sets). The proportions of the interaction variations explained by the self, complement and average complement index vectors (R^2 values) for the various subgroups as well as their respective variances are provided in Table 5.13.

Table 5.13.The mean and variance between the proportion of variation explained (R^2 values) by the self, complement and average complement index vectors of 20 genotypes of the CIMMYT ISWYN data for 1986-1987.

Target set	20 Genotypes		
Analysis	Self	complement	Average complement
Mean R^2 value	24.0	14.1	21.0
variance	13.4	2.0	7.9

From the analysis (Table 5.13) the self R^2 value is larger than the respective complement and average complement R^2 values. This conforms to expectation based on the above discussion. However, the analysis of the average complement index for the 20 genotypes with mean R^2 value of 21.0 gets closer to the self mean R^2 value of 24.0 than that of the complement which is 14.1. It therefore suggests that the procedure of averaging the complement indices has rendered the result to be biased towards the self index vector. This indicates that though the average complement index vector works well, there is the needed to evolve a scale factor that will bring it closer to the complement index than the self index vector.

5.8 Chapter Summary

1. This chapter reviewed the AMMI model essentially on the problems associated with the model as pointed out by Crossa et al (1991), Romagosa et al (1993) and Piepho (1994).

2. The estimation of the multiplicative parameters of AMMI model is strongly "optimistic". An investigation on the proportion of the interaction sums of squares explained by AMMI1 was carried out based on comparative analysis of correlations and sums of squares. This established a close link with the correlations of the residual matrix, the environment index and the sums of squares explained.

3. An alternative method of cross validation procedure which extracts the environment index from a section of the data set to analyse the remaining part was proposed; the complementary procedure.

4. The complementary procedure was justified with a series of comparative analyses using the complement and the self index vectors. The analyses were performed using random samples (subgroups) of a data from the CIMMYT wheat data (1986-1987).

5. It was finally recommended that splitting the data into two equal subgroups may give better complementary predictions and that the average of the two eigenvectors from the two subgroups may be used to analyse the entire data set.

CHAPTER SIX

CONCLUDING REMARKS AND FURTHER STUDIES

The aim of this thesis has been to develop a method for the analysis and interpretation of genotype by environment interaction. We have examined one of the popular methods (AMMI) currently being used for analysing genotype by environment interaction. Several methods for analysing GE data have been reviewed in Chapter two. AMMI analysis was noted to give precise estimates for GE data. In Chapter three the conceptual and theoretical framework of the AMMI model has been presented.

In chapter four, a real data set (SVT trial) from Crop Research Institute, Fumesua, was analysed to investigate the performance of the AMMI model, analysis of variance and linear regression analysis in selecting the best genotype. Whilst analysis of variance failed to break down the interaction component, regression analysis explained only 20.5% leaving substantial amount of the interaction (79.5%) as residual. AMMI analysis with 2 IPCA axes on the other hand accounted for a sizable amount of the interaction (69%) suggesting quite a good model performance.

In Chapter five we have first of all explored several areas of the additive main effects and multiplicative interaction (AMMI) models. The sums of squares arising from AMMI models were considered in detail. We have shown how the structure of the residual interaction matrix after fitting the additive terms exhibit high

correlations between the genotype vectors over the environments. Ignoring the correlations between the terms of the residual interaction matrix has been shown to lead to optimism in the fitting sums of squares. The optimism is dealt with by using the average complement index vector proposed in this thesis.

The recommendation given in this thesis is that half of the number of genotypes in a trial should be used to estimate the complement index vector, and the average complement index vector obtained as an average of the complement index vectors.

There are several directions in which this work could be extended. Firstly, an analysis of the average complement analyses shows that some level of biasness is introduced into the index vector as a result of the averaging of the indices. This requires a very close consideration. We therefore suggest that a future study should target at a scale factor that will bring the performance of the average complement index vector to be at par with the complement vector which is established as independent estimate of the environment index for a complement subgroup of a data.

Secondly, the method of analysis which has been proposed in this thesis is suitable for continuous data, but a major new dimension can be added to this work by extending the idea of complement index vector to handle data in the form of proportions or counts. In principle the frame work of Genstat could be used where

the error is modelled by the binomial distribution. It is important to extend the idea of complement index vector to handle proportions because yield is not the only response variable that is of interest. In many cases Plant breeders want to assess the resistance of genotypes to disease or to pests and the response variable then tend to be in the form of a proportion.

We have not particularly looked at the connection between yield data and environmental variables like day length, temperature and others due to data limitation. Methods of interpreting GE interaction in the presence of explicitly measured environmental variables are divided into two groups. The first method extracts environmental characterisation from the GE data, which are eventually related to measured environmental variables (Eeuwijk, 1992). The second method which incorporates measured environmental variables also referred to as Redundancy analysis imposes a restriction on the environmental scores to ensure a linear combination of the environmental variables. In other words redundancy analysis is an AMMI analysis with a restriction on the environmental scores (Rao, 1964; Hardwick and Wood, 1972; Izenman, 1975; Van der Wollenberg, 1977; Davies and Tso, 1982).

The need to research on which environmental variables are to be included in the main analysis cannot be over emphasised. It is preferable for the environmental variables to be selected according to the breeder's ideas about what affects the yield of the crop rather than, for example, monthly rainfall totals because these

always happen to be available. Variable selection methods like stepwise procedure (Snedecor and Cochran, 1980) and backward elimination (Joliffe, 1986) have been applied to select influential environmental variables in the analysis. Stepwise procedure for instance has been noted to have serious pitfalls (Berk, 1978).

Finally another important method that can be considered for future studies is Cluster analysis. This method is applied when the researcher's objective is to identify genotypes or environments which show similar patterns of response. Basically genotypes or environments are declared similar when they are within the same clusters and different when they are from different clusters. Cluster analysis combined with an ordination technique (e.g. Principal component analysis) can be useful for identifying patterns in a data set.

REFERENCES

1. Abou-El-Fittouh H.A, Rawlings J.O, Miller P.A. (1969). Genotype by environment interactions in cotton – Their nature and related environmental variables. *Crop Sci. 9:337 – 381.*
2. Annicchiarico P. (1997b). Additive main effect and multiplicative interaction (AMMI) of genotype-location interaction in variety trials repeated over years. *Theor. Appl.Genet., 94,1072-1077.*
3. Berk KN. (1978). Comparing subset regression procedures. *Technometrics, 20: 1 6*
4. Boonseng, O., C. Rajanaridpiched E. Sarobol, P. Duangbpatra, and P Chatwadiirawong (1997). Stability of Thai cassava varieties grown in early rainy season. *Kasetsart J. (Nat Sci) 31:281 – 290. (in Thai)*
5. Bradu D. and Gabriel K.R. (1978). *Technometrics 20,47-68.*
6. Brain Snoad and Arthur A. E. (1974). Genotype- Environment interaction in peas. *heoretical and applied Genetics 44,222-231.*
7. Brain Wescott. (1986). Some methods of analysing genotype-environment interactions. *Heredity 56, 243-253.*
8. Breese, E. L. (1969). The measurement and significance of genotype-environment interaction in grasses. *Heredity, Lond. 24: 27 – 44*
9. Chatwachirawong, P. (1993). Yield Stability of major sugarcane cultivars in Thailand *M. S. thesis, Kasetsart University, Bangkok*
10. Cormack, RM (1971). A review of classification (with discussion). *Journal of the Royal Statistical Society. A 134: 321 – 367*
11. Cornelius P.L (1993). Statistical tests and retention of terms in Additive main effect and multiplicative interaction models for cultivar trials. *Crop Sci. 33, 1186-1193.*
12. Crossa J. (1990). Statistical analysis of multilocation trials. *Advances in Agronomy, Vol. 4, 55 – 85*
13. Crossa J., Fox P. N, Pfeiffer H.W, Rajaram S and Gauch H. G.(1991). AMMI adjustment for statistical analysis of an international wheat yield trial *Theor Appl. Genet 81: 27-37*

14. Davies P.T, Tso M.K.S (1982). Procedures for reduced-rank regression. *Appl Stat 31:244 – 255*

15. Digby, P.G. N (1979). Modified Join Regression analysis for incomplete environmental data. *Journal of Agric Sci., Cambridge 93, 81 – 86*

16. Ebdon J. S and Gauch H. G. (2002). Additive Main Effect and Multiplicative Interaction Analysis of National Turfgrass Performance Trials; I. Interpretation of Genotype x Environment Interaction. *Crop Sci. 42; 489-496.*

17. Ebdon J. S and Gauch H. G (2002). Additive Main Effect and Multiplicative Interaction Analysis of National Turfgrass Performance Trials; II. Cultivar Recommendations. *Crop Sci. 42; 497-506*

18. Eberhart, S.A and Russel W.A (1966). Stability parameters for comparing varieties. *Crop Science 6, 36 – 40.*

19. Eckart C and Young G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1,211-218.*

20. Finlay K. W. and Wilkinson G. N (1963). The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric Res., 14, 742-754.*

21. Fishier R. A. (1926). Arrangement of field experiments. *Journal of the Ministry of Agriculture 33, 503 – 13.*

22. Fox P.N and Rosielle A.A (1982). Reducing the influence of environmental main effects on pattern analysis of plant breeding environments. *Euphytica, 31: 645 – 656.*

23. Fox P.N, Skovmand B., Thompson B.K, Braun H.J and Cormier R. (1990). Yield and adaptation of hexaploid spring triticale. *Euphytica, 47, 57-64.*

24. Francis TR and Kannenberg LW (1978). Yield stability studies in short-season maize. I. A descriptive method for grouping genotypes. *Can J. Plant Sci. 58: 1029 – 1034*

25. Freeman G. H, (1973). Statistical methods for the analysis of Genotype-Environment interactions, *Heredity 31 (3), 339-354*

26. Freeman G.H and Donker, B.D (1973). The analysis of variation between and within genotypes and environments. *Heredity 30, 97 – 109.*

27. Freeman G.H and Perkins J.M. (1971). Environmental and genotype – environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. *Heredity, 27, 15-23.*

28. Fripp Yvonne J. And Caten C.E (1971). Genotype –environment interactions in schizophyllum Commune. I. Analysis and Character. *Heredity 27,393-407*

29. Gabrial K.R. (1971). Biplot display of Multi-variety matrices and application to principal component analysis. *Biometrika 58:453 – 467*

30. Gauch H.G Jr. (1982). Noise reduction by eigenvector ordinations. *Ecology 63:1643-1649*

31. Gauch H.G Jr. (1988). Model selection and validation for yield trials with interaction. *Biometrics 44:705 – 715*

32. Gauch H.G and Zobel R.W. (1988). Predictive and postdictive success of statistical analysis of yield trials. *Theor. Appl. Genet., 76,1-10.*

33. Ghaderi A, Adams M.W and Seattler A.W (1982). Environmental response patterns in commercial classes of Common bean (Phaseolus vulgaris L.) *Theoretical and Applied Genetics, 63, 17 – 22*

34. Gollob H.F (1968); A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika- Vol. 33, No. 1.*

35. Gordon A.D (1981). Classification. Chapman and Hall, London.

36. Gower J.C (1975). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika, 53: 325 – 338*

37. Hardwick R. C. and Wood J. T. (1972). Regression methods for studying genotype – environment interactions. *Heredity 28,209-222.*

38. Hanson W.D.(1970). Genotype stability. *Theoretical and Applied Genetics, 40,226-231.*

39. Hill M.O. (1974). Correspondence analysis; a neglected multivariate method. *Applied Statistics, 23, 340-354.*

40. Hill J. (1973). Methods of analysing completion with special reference to herbage plants. II. Effects of associate plants. *Journal of Agricultural Science, Cambridge 81, 91-98.*

41. Hill J. (1975). Genotype-environment interactions– a challenge for plant breeders. *J. agric. Sci., camb. , 85,477-493*

42. Hill J and Goodchild, N.A. (1981). Analysing environments for plant breeding purposes as exemplified by multivariate analysis of long term wheat yields. *Theoretical and Applied Genetics, 59:317 – 325.*

43. Izenman AJ (1975). Reduced – rank regression models for the multivariate linear Model *J. Mult Anal 5:248 – 264*

44. Jolliffe I.T (1986). Principal Component Analysis Springer – *Verlag, New York*

45. Johnson G.R. (1977). Analysis of genotypic similarity in terms of mean yield and stability of environmental response in a set of maize hybrids. *Crop Science, 17, 837-842.*

46. Kempton, R.A (1984). The use of biplots in interpreting variety by environment interactions *J. Agric Sci. 103, 123 – 135*

47. Lin C. S and Binns M. R (1988). A Superiority Measure of Cultivar Performance for Cultivar x Location Data. *Can. J. Plant Sci. 68; 193-198.*

48. Lin C. S, Binns, M. R and Lefkovitch L. P (1985). Stability Analysis; where do we stand? *Engineering and Statistical Res. Centre, 1 - 770.*

49. Lin C.S. and Thompson B. (1975). An empirical method of grouping genotypes based on a linear function of the genotype –environment interaction. *Heredity, 34,255-263.*

50. Mandel J.(1969). The partitioning of interaction in analysis of variance. *J. Res. Nat. Bur. Standards Sect. B 73B, 309-328*

51. Mandel J (1971). A new analysis of variance model for non-additive data. *Technometrics 13:1 – 18*

52. Mather, Kenneth and Caligari, P.D.S (1974) Genotype x environment interactions. I. Regression of interaction on overall effect of the environment. *Heredity 33, 43 – 59.*

53. Mead, R (1988). The design of experiments, Cambridge; *Cambridge University Press*

54. Menz K.M. (1980). A comparative analysis of wheat adaptation across international environments using statistic dominance and pattern analysis. *Field Crop Research, 3:33 – 41*

55. Mungomery V.E, Shorter, R. and Byth D.E (1974). Genotype x environment interaction and environmental adaptation I. Pattern analysis application to soyabean populations. *Australian Journal of Agricultural Research, 25, 59 – 72.*

56. Patterson H. D. and Thompson, R (1971). Recovery of inter block information when block sizes are unequal. *Brometrika, 58, 445 – 554*

57. Pearson K. (1901). " On lines and planes of closest fit to systems of points in space". *Philosophical Magazine, 6(2), 559-572.*

58. Perkins J.M and Jinks J.L. (1968). Environmental and genotype –environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Hearedity, 23, 525-535.*

59. Piepho H.P (1994). Best Linear Unbiased Prediction (BLUP) for regional trails: a comparison to additive main effects and multiplicative interaction (AMMI) Analysis. *Theoretical and Applied Genetics, 98, 647 – 654*

60. Plaisted R.L and Peterson, L.C (1959). A Technique for evaluating the ability of selection of yield consistently in different locations or seasons. *American Potato Journal 36, 398 – 420*

61. Ramey T.B and Rosielle, A.A (1983). Hass Cluster analysis, a method of grouping genotypes or environments in plant breeding. *Theoretical and Applied Genetics, 66: 131 – 133.*

62. Rao C.R (1964). The use and interpretation of principal component analysis in applied research. *Sankhya A26:329 – 358*

63. Schoenberg, I.J (1938). Remarks to Mauriee Frechet 'Sur la de'finition axomtique d'une classe d'espace instances vectoriellement applicable sur l'esapce d'Hilbert. *Annals of Mathematics (2nd Series), 36, 724 – 732.*

64. Seiller G.J and Stafford R.E (1985). Factor analysis of components in Guar. *Crop Sci. 25, 905 – 908*

65. Shorter R., Byth D.E and Mungomery V.E. (1977). Genotype x environment interactions and environmental adaptation II. Assessment of environment contributions. *Australian Journal of Agricultural Research, 28, 223 – 235*

66. Shukla G.K. (1972). Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity, 29,237-245.*

67. Snedecor G.W, Cochran W.G (1980). Statistical Methods. *7th edn. Lowa State University Press, Ames*

68. Sokal R.R and Michener C.D. (1958). A statistical model for evaluating systematic relationship. University of Kansas Science Bulletin. 38,1409-1438.

69. Sprague G.F and Federer W.T (1951). A Comparison of variance components in corn yield trials. II Error, Year x variety, location x variety and variety components. *Agronomy Journal 43, 535 – 41*

70. Torgerson W.S (1952). Multidimensional scaling .I. Theory and Method. *Psychometrika 17, 401 – 419*

71. Van den Wollenberg A.L. (1977). Redundancy analysis and alternative canonical correlation analysis. *Psychometrika 42:207 – 219*

72. Van Eeuwijk F. A. (1992). Interpreting genotype-by-environment interaction using redundancy analysis. *Theor. Appl. Genet, 85; 89-100*

73. Walton P.D. (1972). Factor analysis of yield in spring wheat (Triticum aetivum L.). *Crop Sci. 12, 731 – 733*

74. Williams W.T. (1976). Pattern analysis on agricultural science. Elsevier Scientific Publishing Company, Amsterdam.

75. Williams E. R. and Wood J. T (1993). Testing the significance of Genotype-environment interaction. *Austral. J. Statist. 35(3), 359-362.*

76. Wricke G (1964). Zur berechung der okoralenz bei sommerweizen und hafer *Zeitschrift fiir Pflanzenuchtung, 52, 127 – 138*

77. Wricke G. (1962). Uber eine methode Zur erfassung der Okologischen streubreite in feldversuchen. *Zeitschrift fier Pflanzen-zuchtung, 47, 92 – 96*

78. Wright A. J. (1971). The analysis and prediction of some two factor interactions in grass breeding. *J. agric. Sci., Camb, 76, 301-306.*

79. Yates F. and Cochran W. G. (1938). The analysis of groups of experiments. *Agric Sci. 28; 556-580.*

80. Young G. and Householder, AS (1938). Discussion of a set of points in terms of Mutual distances. *Psychometrika 3, 19 – 22*

81. Zobel R.W, Wright M.J and Gauch H.G. (1988). Statistical analysis of a yield ntrial. *Agron J. 80, 388-393.*

APPENDIX A

Table 7.0. Sums of squares explained by genotype vectors of residual interaction matrix from the first analysis of yield data consisting of 9 environments and 10 genotypes randomly selected from 45 environments and 40 genotypes respectively (CIMMYT data, 1986-1987).

Genotype	1	2	3	4	5
Sums of squares	18026935	1242867	40168	527317	5930931

Genotype	6	7	8	9	10
Sums of squares	11603335	479	4022756	214833	4184265

Table 7.1 Correlations among the ten genotypes (G1-G10) vectors and with the environmental eigenvector (ef) of the genotype vectors used in Table 7.0.

Geno-types	1	2	3	4	5	6	7	8	9	10	ef
1	1.00										
2	0.57	1.00									
3	-0.07	0.05	1.00								
4	0.25	0.07	-0.16	1.00							
5	-0.62	-0.14	-0.24	-0.03	1.00						
6	-0.83	-0.55	-0.41	-0.20	0.69	1.00					
7	-0.02	-0.27	-0.11	-0.60	0.12	0.00	1.00				
8	0.58	-0.58	0.16	-0.22	0.22	0.61	-0.23	1.00			
9	-0.42	-0.41	0.38	-0.26	-0.40	0.11	-0.11	0.30	1.00		
10	0.53	0.33	-0.14	-0.17	-0.71	-0.53	0.26	-0.67	0.20	1.00	
ef	-0.94	-0.58	-0.08	-0.23	0.74	0.93	0.01	0.71	0.20	-0.70	1.00

Table 7.3. Sums of squares explained by genotype vectors of residual interaction matrix from the second analysis of yield data consisting of 9 environments and 10 genotypes randomly selected from 45 environments and 40 genotypes respectively (CIMMYT data, 1986-1987).

Genotype	1	2	3	4	5
Sums of squares	1798769	179838	521	718364	16161281

Genotype	6	7	8	9	10
Sums of squares	1199	5362695	1526109	6146788	3055542

Table 7.4 Correlations among the ten genotypes (G1-G10) vectors and with the environmental eigenvector (ef) of the genotype vectors used in Table 7.3.

Geno-types	1	2	3	4	5	6	7	8	9	10	ef
1	1.00										
2	-0.38	1.00									
3	-0.61	0.49	1.00								
4	-0.23	0.02	0.11	1.00							
5	-0.47	-0.40	0.07	0.21	1.00						
6	0.37	-0.45	-0.82	-0.03	-0.05	1.00					
7	-0.23	-0.28	-0.42	0.27	0.67	0.31	1.00				
8	-0.25	0.59	0.51	-0.38	-0.55	-0.34	-0.80	1.00			
9	0.36	-0.15	-0.01	-0.58	-0.60	0.12	-0.58	0.42	1.00		
10	0.50	0.27	-0.22	-0.27	-0.66	-0.14	-0.43	0.37	0.10	1.00	
ef	0.51	0.29	-0.01	-0.43	-0.95	-0.02	-0.81	0.64	0.74	0.65	1.00

APPENDIX B

Program[1]

"This program [1] calculates the first eigenvector produced
by the selected genotypes whose numbers are equal
to the values of the scalars y1,y2...y20.

PARAMETERS

y1...y20	Genotype vectors used
effe1	Eigenvectors produced by genotypes with values y1...y20
k	Number of eigenvectors required"

```
units [nvalues=1080]
factor [levels=9]environ
factor[levels=40]genotype
factor[levels=3]rep
open'toy4t';channel=4;filetype=output
open'newdata';channel=2

read[channel=2;end=*]env,gen,repn,yield
"print env,gen,repn,yield"
calculate lyld=log(yield)
```

```
scalar y1,y2,y3,y4,y5,y6,y7,y8,y9,y10,y11,y12,y13,y14,y15,y16,y17,y18,y19,y20\
;value=25,10,31,21,23,3,15,1,37,33,32,13,6,24,2,8,22,18,19,28
subset[condition=gen.eq.y1.or.gen.eq.y2.or.gen.eq.y3.or.gen.eq.y4.or.gen.eq.y5.o
r.gen.eq.y6.or.gen.eq.y7.or.gen.eq.y8.or.gen.eq.y9.or.gen.eq.y10.or.gen.eq.y11.or
.gen.eq.y12.or.gen.eq.y13.or.gen.eq.y14.or.gen.eq.y15.or.gen.eq.y16.or.gen.eq.y
17.or.gen.eq.y18.or.gen.eq.y19.or.gen.eq.y20;setlevels=yes]\
oldvectors=env,gen,repn,yield;newvectors=envir1,genot1,repli1,yiel1

calculate nn=nobservation(yiel1)
set[units=nn]
calculate loc1=envir1
calculate  var1=genot1
calculate  re1=repli1
calculate  yl1=yiel1
factor[levels=20]ge1,variety1
factor[levels=9]en1,locat1
factor[levels=3]repi1

table[classification=en1]effenv1
table[class=locat1,variety1]looking1
table[classification=ge1]effgen1
groups [lmethod=given]loc1;factor=en1
groups [lmethod=given]var1;factor=ge1
```

```
groups [lmethod=given]re1;factor=repi1
treatmentstructure en1*repi1+ge1
anova[print=a;fprob=yes] yl1;r1
akeep terms=en1,ge1;effects=effenv1,effgen1
equate effgen1;egen1
scalar no_g1,no_e1,no_ge1,nv1
calculate v11=ge1
calculate no_g1=max(v11)
calculate v21=en1
calculate no_e1=max(v21)
calculate v31=v11+(v21-1)*no_g1
calculate no_ge1=no_g1*no_e1
factor[levels=no_ge1]ge_fact1
group[lmethod=given] v31;factor=ge_fact1
table [classification=ge_fact1]ge_tab1

variate[nvalues=no_ge1]ge_riate1,amm21,fge1
blockstructure

treatmentstructure ge_fact1
anova[print=a;fprob=yes]r1
akeep terms=ge_fact1;means=ge_tab1
matrix[rows=no_e1;columns=no_g1]ge_mat1
```

```
matrix[rows=no_e1;columns=no_g1]res_mat1

equate ge_tab1;ge_mat1
equate ge_tab1;ge_riate1
scalar nulss1
model ge_riate1
fit[print=*;nomessage=l,r]
rkeep deviance=nulss1
variate[nvalues=no_g1]cg1[1...1],vregg1
variate[nvalues=no_e1]ce1[1...1]
for k=1...1
variate[nvalues=no_e1] eig1[1...no_e1]
matrix[rows=no_g1;columns=no_e1]tge1
matrix[rows=no_e1;columns=1]effe1

matrix[rows=1;columns=no_e1]effet1,efe1[k],dd1[1...1]
matrix[rows=1;columns=1]divisor1,cre21
symmetricmatrix[rows=no_e1] hht1
matrix[rows=no_e1;columns=no_e1]llr1
calculate tge1=transpose(ge_mat1)
calculate hht1=product(ge_mat1;tge1)
lrv[rows=no_e1;columns=no_e1]l;llr1
flrv[print=*]inmatrix=hht1;lrv=l
```

```
equate [oldformat=!((1,-8)9,-1)]lr1;newstructures=eig1
for i=1...1
equate eig1[i];effe1
calculate effet1=transpose(effe1)
print effe1
endfor
endfor

close channel=2
```

Program[2]

"This program [2] gives AMMI analysis of a given data set using the self, complement or average complement eigenvectors any of which can be calculated using program[1])

PARAMETERS

mu	mean yield
me	environment means
mg	genotype means
k	number of multiplicative components required
egen	genotype deviation
een	environment deviation
PCA	total sums of squares by the eigenvector used
vg	sums of squares by each genotype
df	degrees of freedom
ef	eigenvector
dd[k]	environment IPCA scores of the kth eigenvector
cc[k]	genotype IPCA scores of the kth eigenvector"

units [nvalues=5400]

factor [levels=45]environ,location

factor[levels=40]genotype,variety

```
factor[levels=3]rep
variate[nvalues=40]mg,egen
variate[nvalues=45]me,een
table[classification=environ]effenv

table[class=location,variety]looking
open'toy4t.sco';channel=4;filetype=output
open'real';channel=2
scalar mu
read[channel=2;end=*]environ,genotype,rep,yield
calculate mu=mean(yield)
calculate yield=(yield)
"print mu"
table[classification=genotype]effgen

treatmentstructure environ*rep+genotype
anova[print=a;fprob=yes]yield;r
akeep terms=environ,genotype;effects=effenv,effgen
equate effgen;egen
equate effenv;een
calculate mg=mu+egen
calculate me=mu+een
```

```
print mg,me
print egen,een
scalar no_g,no_e,no_ge,nv
calculate v1=genotype
calculate no_g=max(v1)
calculate v2=environ
calculate no_e=max(v2)
calculate v3=v1+(v2-1)*no_g
calculate no_ge=no_g*no_e
factor[levels=no_ge]ge_factor
group[lmethod=given] v3;factor=ge_factor
table [classification=ge_factor]ge_table

variate[nvalues=no_ge]ge_variate,amm2,fge
blockstructure
treatmentstructure ge_factor
anova[print=a;fprob=yes]r
akeep terms=ge_factor;means=ge_table
matrix[rows=no_e;columns=no_g]ge_matrix
matrix[rows=no_e;columns=no_g]res_matrix
equate ge_table;ge_matrix
equate ge_table;ge_variate
scalar null,nulss
```

```
scalar regss,gess,sre2,sre2pct,cr2
model ge_variate
fit[print=*;nomessage=l,r]
rkeep deviance=nulss
variate[nvalues=no_g]cg[1...3],vregg
variate[nvalues=no_e]ce[1...3]
for k=1...3
variate[nvalues=no_e] ge[1...no_e]
variate[nvalues=no_e] eig[1...no_e]
matrix[rows=no_g;columns=no_e]tge
matrix[rows=no_e;columns=1]effe
matrix[rows=1;columns=no_e]effet,dd[1...3]
matrix[rows=1;columns=1]divisor,cre2
symmetricmatrix[rows=no_e] hht
matrix[rows=no_e;columns=no_e]llr
calculate tge=transpose(ge_matrix)
calculate hht=product(ge_matrix;tge)

lrv[rows=no_e;columns=no_e]l;llr
flrv[print=*]inmatrix=hht;lrv=l
equate[oldformat=!((1,-44)45,-1)]llr;newstructures=eig
for i=1...1
equate eig[i];effe
```

```
calculate effet=transpose(effe)
endfor
matrix[rows=1;columns=no_g]regg,regg1,cc[1...3]
scalar PCA,df1
variate[nvalues=45]gev[1...40]
equate T(ge_matrix);!p(gev[ ])
variate[nvalues=45] ef
equate effe;ef
print ef
correlate[print=correlations]gev[],ef
calculate regg=product(effet;ge_matrix)
equate regg;regg1
calculate divisor=product(effet;effe)
equate divisor;gess
calculate regg=regg/gess
equate regg;vregg
calculate vg=(vregg*vregg)*max(rep)
print vg
calculate df1=no_g+no_e-1-2*k
calculate cre2=rtproduct(regg1;regg)
calculate PCA=cre2*max(rep)
print df1,PCA
equate cre2;cr2
```

```
calculate cc[k]=regg/((cr2)**0.25)
calculate dd[k]=effet*((cr2)**0.25)
equate cc[k];cg[k]
equate dd[k];ce[k]
print dd[k]
print cc[k]
calculate cre2=cre2/nulss
calculate nge=ltproduct(effet;regg)
equate nge;;ge
calculate amm2=ge_variate-fge
equate amm2;ge_table
equate ge_table;ge_matrix
equate ge_table;ge_variate
endfor
close channel=2
```

APPENDIX C

THE SVT DATA

Yield of ten maize genotypes in nine environments with four replications in the year 1999.

Each row gives the yield of a particular genotype in a given environment (environ) and replication (rep).

environ	genotype	rep	yield
1	9	1	4574
1	6	1	4857
1	10	1	3482
1	2	1	4164
1	1	1	5109
1	7	1	4506
1	5	1	3506
1	3	1	4979
1	4	1	4968
1	8	1	5369
1	10	2	4385
1	1	2	5662

1	3	2	4518
1	5	2	3479
1	9	2	5479
1	2	2	4037
1	4	2	4718
1	6	2	3895
1	8	2	4968
1	7	2	3907
1	9	3	6373
1	10	3	3666
1	6	3	4947
1	3	3	4518
1	5	3	3272
1	7	3	6518
1	8	3	4813
1	2	3	3885
1	1	3	5813
1	4	3	4838
1	4	4	4409
1	5	4	3338

1	8	4	4855
1	1	4	5150
1	3	4	4560
1	9	4	4662
1	2	4	3672
1	7	4	5042
1	6	4	4536
1	10	4	3874
2	1	1	6593
2	7	1	6069
2	4	1	6752
2	5	1	4907
2	8	1	6831
2	3	1	5524
2	6	1	6858
2	2	1	5922
2	9	1	5815
2	10	1	6336
2	6	2	6816
2	1	2	5987

2	2	2	5686
2	5	2	4247
2	4	2	5882
2	7	2	5827
2	8	2	7058
2	3	2	5590
2	9	2	5943
2	10	2	7115
2	1	3	6723
2	9	3	6179
2	3	3	6121
2	2	3	7036
2	5	3	5805
2	4	3	6500
2	10	3	7459
2	7	3	6632
2	8	3	6889
2	6	3	5451
2	9	4	6905
2	10	4	6567

2	3	4	5429
2	1	4	6908
2	8	4	4983
2	6	4	5879
2	7	4	5158
2	4	4	4744
2	5	4	4762
2	2	4	4717
3	9	1	6640
3	1	1	6808
3	10	1	4813
3	3	1	6403
3	5	1	5286
3	8	1	5482
3	7	1	5678
3	2	1	6031
3	4	1	6154
3	6	1	5847
3	2	2	6725
3	9	2	5837

3	6	2	4777
3	10	2	3739
3	3	2	4624
3	5	2	4274
3	8	2	6162
3	7	2	5826
3	4	2	5501
3	1	2	5983
3	5	3	4286
3	4	3	5454
3	10	3	3602
3	1	3	4312
3	3	3	5749
3	7	3	6152
3	2	3	6505
3	8	3	6193
3	6	3	5823
3	9	3	5421
3	5	4	5233
3	8	4	5954

3	7	4	5903
3	10	4	3848
3	4	4	5286
3	9	4	5908
3	1	4	5545
3	2	4	6099
3	6	4	5770
3	3	4	4624
4	7	1	2530
4	3	1	2983
4	9	1	3021
4	2	1	2121
4	6	1	2770
4	10	1	2177
4	8	1	2962
4	4	1	1909
4	1	1	3075
4	5	1	752
4	1	2	2966
4	4	2	4348

4	7	2	1805
4	10	2	1895
4	2	2	3425
4	5	2	2027
4	8	2	3086
4	6	2	2311
4	3	2	3287
4	9	2	3763
4	9	3	3064
4	4	3	2948
4	6	3	2515
4	8	3	3147
4	5	3	1462
4	3	3	2232
4	7	3	2530
4	1	3	3276
4	2	3	3919
4	10	3	1369
4	3	4	1911
4	2	4	3272

4	5	4	1691
4	7	4	3291
4	10	4	2521
4	1	4	1998
4	6	4	2969
4	4	4	3287
4	9	4	3593
4	8	4	2198
5	10	1	4680
5	7	1	6607
5	8	1	5923
5	2	1	5005
5	3	1	6634
5	9	1	4467
5	1	1	4636
5	4	1	5292
5	5	1	2676
5	6	1	5672
5	5	2	3979
5	8	2	5432

5	9	2	5722
5	7	2	6041
5	3	2	4919
5	1	2	6322
5	6	2	5473
5	10	2	5508
5	2	2	5625
5	4	2	6819
5	5	3	6722
5	6	3	5374
5	10	3	4648
5	2	3	5650
5	7	3	5207
5	3	3	5751
5	8	3	5587
5	1	3	6048
5	4	3	4785
5	9	3	4428
5	3	4	5394
5	1	4	5779

5	9	4	4535
5	8	4	6214
5	10	4	5294
5	2	4	5808
5	5	4	4229
5	4	4	6263
5	6	4	4969
5	7	4	6182
6	9	1	4337
6	1	1	5798
6	6	1	2875
6	2	1	2181
6	3	1	5107
6	5	1	3857
6	8	1	5439
6	10	1	4735
6	7	1	5577
6	4	1	4957
6	5	2	3761
6	2	2	4919

6	8	2	6345
6	3	2	5347
6	6	2	5321
6	1	2	4979
6	7	2	4919
6	10	2	4351
6	4	2	5121
6	9	2	5908
6	6	3	3954
6	9	3	4058
6	1	3	4461
6	4	3	4269
6	5	3	2933
6	2	3	4439
6	3	3	4116
6	7	3	4371
6	10	3	3769
6	8	3	3866
6	8	4	5453
6	4	4	4168

6	7	4	4263
6	1	4	4618
6	3	4	4982
6	6	4	4116
6	10	4	2899
6	9	4	4676
6	5	4	2643
6	2	4	4016
7	7	1	4405
7	6	1	5306
7	2	1	4279
7	10	1	3840
7	8	1	5472
7	5	1	3815
7	4	1	3915
7	1	1	5152
7	3	1	4292
7	9	1	3979
7	2	2	4320
7	6	2	3720

7	8	2	5057
7	3	2	3927
7	4	2	3529
7	1	2	5150
7	7	2	4387
7	5	2	4317
7	9	2	4959
7	10	2	3294
7	1	3	4612
7	3	3	4643
7	10	3	3646
7	2	3	4759
7	4	3	3915
7	6	3	3962
7	5	3	3569
7	8	3	3890
7	7	3	3891
7	9	3	4643
7	9	4	4800
7	10	4	3298

7	4	4	4307
7	1	4	4182
7	7	4	3456
7	2	4	5184
7	5	4	4618
7	3	4	5022
7	6	4	4863
7	8	4	5313
8	3	1	4049
8	5	1	4003
8	4	1	4387
8	2	1	4158
8	1	1	4630
8	7	1	4517
8	9	1	2084
8	10	1	4214
8	8	1	4700
8	6	1	4149
8	9	2	4243
8	1	2	3768

8	5	2	3839
8	6	2	4585
8	2	2	4278
8	8	2	4407
8	10	2	3043
8	4	2	5440
8	3	2	2502
8	7	2	3050
8	2	3	4643
8	1	3	4224
8	8	3	4771
8	7	3	4054
8	6	3	3939
8	5	3	2638
8	3	3	4372
8	4	3	2416
8	10	3	4559
8	9	3	3510
8	4	4	2399
8	3	4	3279

8	10	4	1641
8	9	4	2079
8	6	4	2665
8	5	4	3053
8	8	4	3522
8	7	4	4258
8	1	4	3275
8	2	4	4382
9	4	1	3755
9	8	1	4267
9	2	1	3711
9	9	1	2651
9	5	1	2729
9	1	1	3346
9	3	1	2931
9	7	1	2698
9	10	1	648
9	6	1	3004
9	10	2	858
9	7	2	2497

9	4	2	2439
9	1	2	2560
9	9	2	3413
9	6	2	3795
9	3	2	2201
9	5	2	3499
9	8	2	3238
9	2	2	3413
9	2	3	3648
9	7	3	3393
9	5	3	1931
9	3	3	1707
9	6	3	3268
9	8	3	4209
9	4	3	2560
9	10	3	1476
9	9	3	2051
9	1	3	3520
9	8	4	1813
9	9	4	2453

9	6	4	3333
9	4	4	3075
9	1	4	4053
9	10	4	1792
9	5	4	2347
9	7	4	1687
9	2	4	2987
9	3	4	3162