ISSN: 2410-1397

MODELING GENOTYPE AND ENVIRONMENT INTERACTION FOR PERFORMANCE STABILITY AND ADAPTABILITY OF SUGARCANE CULTIVARS

MASTERS OF SCIENCE IN BIOMETRY

OTIENO VICTOR OUMA

Supervisor:

Dr. Nelson O. ONYANGO

Research Thesis in Biometry

Number 82, 2016

School of Mathematics

College of Biological and Physical Sciences

UNIVERSITY OF NAIROBI

MODELING GENOTYPE AND ENVIRONMENT INTERACTION FOR PERFORMANCE STABILITY AND ADAPTABILITY OF SUGARCANE CULTIVARS

by

Otieno Victor Ouma

A Thesis Submitted to School of Mathematics, University of Nairobi in Partial Fulfillment of the Requirements for the Degree of Masters of Science in Biometry.

September, 2016

Declaration of Authorship

I, Otieno Victor Ouma, declare that this thesis titled, "MODELING GENOTYPE AND ENVIRONMENT INTERACTION FOR PERFORMANCE STABILITY AND ADAPTABILITY OF SUGARCANE CULTIVARS" and the work presented in it are my own. I confirm that:

This work was done wholly or mainly while in candidature for a research degree at this University and has never been presented before in any other institution.

I have acknowledged all main sources of help.

Signed:

_____

Date:

_____

I, Dr. Nelson O. ONYANGO, declare that this thesis titled, "MODELING GENOTYPE AND ENVIRONMENT INTERACTION FOR PERFORMANCE STABILITY AND ADAPTABILITY OF SUGARCANE CULTIVARS" and the work presented in it have been done under my supervision. I confirm that:
This project has been submitted for examination with my approval as University supervisor.

Signed:

_____

Date:

_____

UNIVERSITY OF NAIROBI

Abstract

College of Biological and Physical Science

School of Mathematics

Masters of Science in Biometry

MODELING GENOTYPE AND ENVIRONMENT INTERACTION FOR PERFORMANCE STABILITY AND ADAPTABILITY OF SUGARCANE CULTIVARS

by Otieno Victor Ouma

In last stages of the sugarcane breeding programs, cultivars are evaluated in multiple environments for stability and adaptability that often results in Genotype by Environment interaction (GEI). GEI is a challenge to selection of high performing and stable cultivars. Univariate, Multivariate and Bayesian statistical techniques have been developed to help with interaction problem. The use of different treatment controls in test environments, dropping of poor performing cultivars in earlier stages and missing data occasion by other eventualities results in unbalanced dataset for combined analysis. Statistical techniques for determining cultivars performance, stability, adaptability when GEI is significant like Additive Main Effect and Multiplicative Interaction (AMMI) and related principles like singular value decomposition (SVD) and principal components analysis (PCA) requires balanced data matrix.There are also many GEI matrix imputation techniques producing different values and biases. The objective were statistically evaluate cultivars using AMMI modeling in the presence of GEI, AMMI biplot analysis for performance and stability, compare AMMI stability value (ASV)selection index to yield stability index (YSI) and non-parametric Rank-Sum (RS) index and Compare performances of Expectation maximization- AMMI (EM-AMMI) and Expectation maximization Singular value decomposition (EM-SVD) imputation techniques in imputing genotype and environment two-way data table. Secondary experimental data of 33 cultivars from Mtwapa Series 2006 (MS2006) and seven standards totaling 40 cultivars of the Preliminary variety trials (PVT) with the Randomized Complete Block Design (RCBD) of three replication in the nine test environments (harvest) was used. Individual and combined environment analysis preluded and precipitated AMMI analysis. AMMI modeling uses ANOVA for additive effects and PCA for interaction effect. Error mean squares (EMS) from individual environments were homogeneous allowing their combination for analysis and environment, genotype and GEI effects in combined analysis were significant thus precipitating AMMI analysis. EM-AMMI and EM-SVD produced correlated and non-significantly different imputed values, however data structures differed immensely by PCA and biplot analysis. Using EM-SVD imputed data, Environment effect accounted for 72%, genotypes 6% and GEI 8% while the residual accounted for 13%. Out of nine AMMI models(AMMI0-AMMI8),only AMMI0 and AMMI1 were significant (at $\alpha = 0.05$) with p-values of 2.382E-04 and 7.34E-09 respectively. By Gollobs F-test AMMI1 explain 77.11% of variation in GEI which is patter response present in the GEI sum of squares with 46 degrees of freedom (43.4% of the interaction degrees of freedom) and sufficiently explaining GEI effect and complexity. GEI complexity was simple given AMMI1 showing lower diversity in germplasm Environments were delineated to four harvest groups and ideal cultivars that were stable and high yielding were MS271, Ms326, Ms278, Ms556 and MS395. The commonly selected cultivars by the indices were Ms282 and Ms339 for performance and stability but they also differed slightly in other cultivars. AMMI model

identified interaction patterns, noise and extent of complexity. Through scores, performance, stability, adaptability and test environment delineation and GEI complexity were determined by AMMI1.

Acknowledgements

First and foremost i thank Almighty God for having enable this work from the beginning to the end. I would wish to appreciate Dr. Nelson O. Onyango; my supervisor for having created adequate time for my work and consistently providing advise and guidance.
I am greatly indebted to Kenya Agriculture and Livestock Research Organization - Sugar Research Institute (KALRO-SRI) for having facilitated my study and offering enough time to concentrate and complete the task.
Lastly i thank my parents and siblings for their unwavering support and prayers

Contents

List of Figures

List of Tables

List of Abbreviations

GEI	Genotype by Environment Interaction
AMMI	Additive Main effect and Multiplicative Interaction
SVD	Singular Value Decomposition
PCA	Principal Component Analysis
ASV	AMMI Stability Value
YSI	Yield Stability Index
RS	Rank Sum
EM-AMMI	Expectation Maximization-AMMI
EM-SVD	Expectation Maximization SVD
ANOVA	Analysis of Variance
EMS	Error Mean Squares
KALRO-SRI	Kenya Agriculture and Livestock Research Organization-Sugar Research Institute
MET	Multi Environmental Trials
GEE	Genotype and Genotype by Environmental
DI	Desirability Index
IPCA	Interactive Principal Component Axis
JRA	Joint Regression Analysis
AMMID	AMMI Distance
DFRI	Distributive Free Imputation
RMSPD	Root mean square predictive difference
PRESS	Predictive Residual sum of squares
F	Fishers ratio
GoK	Government of Kenya

Chapter 1

Introduction

1.1 Background

Sugarcane farming in Kenya is mainly in South Nyanza, Nyando, Western Kenya and South Coastal regions. It engages 8,000 people directly, over six million people indirectly and practiced by over 600,000 farmers. The sector contributes 25% of Agricultural Gross Domestic Product and third ranked contributor after horticulture and tea GoK (2015). Sugarcane breeding is undertaken by Kenya Agriculture and Livestock Research Organization-Sugar Research Institute (KALRO-SRI) with hybridization at Mtwapa and subsequent evaluation of cultivars across sugarcane growing regions. Several varieties has been released since 2001, their adoption depends on environment, productivity, and resistance to un-desired biotic and abiotic factor affecting performance and stability in wider environments. Breeding begins with hybridization; a process where parents known to have desired traits are crossed to provide progenies for testing and selection. Breeding objective in Kenya is to develop high yielding, high sucrose cultivars that are resistant to undesired biotic and a biotic factors. In last stage (adaptation stage), narrowed down cultivars are tested in many environments with diverse agroclimatic conditions. Performance in these environments with respect to the standards varieties determines selection and recommendations of cultivars. Adaptation is ability of a genotype to survive in a given environment. In variety trials, cultivars performing better than standard commercial variety are recommended.

Stability is sustained performance (high sucrose and yields) of the cultivar with very little variation if any in different environments. The procedures for quantifying stability vary as reviewed by Tadege, Utta, and Aga (2014). Genotype dynamic stability is variation in its performance given changes in environment that are predictable (agronomic stability). Pereira et al. (2012). Selection efficiency of superior cultivars are affected by environmental, genetics and genotype and environment interaction (GEI), the variations

in cultivars performances in different environments are attributed to the GEI effects Falconer and Mackay (1996). A GEI is differential ranking of genotypes across environments or changes in relative performance of genotypes in different environments Baye, Abebe, and Wilke (2011), it complicates selection process and efficiency by ranking genotypes differently across environments thus reduces accuracy of cultivar recommendation for a target environment.

Analysis of variance (ANOVA) technique evaluates and ranks cultivars. It does mean separation based on trait under investigation. With reference to the standard commercial varieties, good cultivars are identified and recommended. It is good for individual site trials but requires further analysis in multi environmental trials (METs) where GEI effects pose a challenge to identification and recommendation of high performing cultivars. High and consistent performance is the indicator of better adaptability to the test environment.

1.1.1 Statement of the problem

Evaluations of sugarcane cultivar for performance, adaptability and stability in Multi-Environmental Trial (MET) often resulted in GEI that compromises selection efficiency. ANOVA has the ability to determine performance of cultivars but inadequate in determining stability and adaptability in the face of significant GEI. Measured traits are less predictable and cannot be interpreted using genotype and environment alone, more analyses are needed Gauch, Piepho, and Annicchiarico (2008) as cited in Akter et al. (2015). Without GEI, genotypes performances across test environments are interpreted using main effects and ANOVA is adequate. There are conflicting theories as whether main effect are interpreted in the presence of GEI, however what is clear is that GEI necessities stability and adaptability analysis. Apart from regression methods, there are other better methods such as the AMMI, PCA and genotype and genotype by environment (GEE) that are thought to address the GEI better. New varieties release since 2001 only occupies less than 10% of cane surface area and there is need increase coverage. Their evaluation using techniques like AMMI and PCA would help determine true stability and performance in the face of GEI that might be affecting them and subsequently performances. GEI is sign of diversity in test cultivars but must be handled with appropriate techniques that identify best ones for recommendation. The use of AMMI, PCA and GEE requires balance dataset and most METs are characterized with missing /unbalanced data due to use of different cultivars and controls (standards) in across environments and may also

arise from pest and disease destruction of plots. GEI matrix imputation becomes necessary to proceed with the evaluation, however there are so many imputation techniques producing different values and biases creating a problem on the choice of best imputation techniques to adopt.

1.1.2 Main Objective

Utilization of Additive Main Effect and Multiplicative Interaction models and related principals (Principal Component Analysis and Singular Value Decomposition) in selection of sugarcane cultivars given significant Genotype by Environment Interaction and unbalanced data sets.

1.1.3 Specific Objectives

- Statistical evaluation of cultivars using AMMI model with existence of GEI.
- Graphical analysis using AMMI-Biplots for stability and adaptability of superior performing cultivars.
- Compare AMMI Stability Value(ASV) selections with Yield Stability Index (YSI) and non-parametric Rank Sum (RS) index.
- Comparative performances of EM-SVD and EM-AMMI in Imputation of Genotype by Environment interaction data matrix.

1.1.4 Justification and significance of the study

Cultivars are evaluated in MET for their performance and adaptation before recommendation for adoption. Their stability with respect to desired traits is important in the presence of GEI. ANOVA treats GEI superficially hence the need for a robust statistical techniques like AMMI and related techniques (PCA, GEE). Methods of evaluation like joint regression, mixed modeling, Wricke's ecovalence (W_i), Shukla's stability variance (σ^2) and coefficient of determination r_i^2 have weaknesses hence the advocation for AMMI. AMMI model has been used extensively in other crops but limited in sugarcane cultivars evaluation. Development of cultivars that are high performing and stable for specific environments will spar adoption, boast national productivity for the country.

1.1.5 limitation of the study

This study is limited to AMMI modeling as a solution to the GEI problem in evaluation of sugarcane cultivars, AMMI-Biplot analysis for visualization, determination of stabilities, performance, adaptability. The imputation was restricted to the use of EM-SVD and EM AMMI omitting other techniques. The data limitation and assumptions of test makes the crop cycle of sites to represent the environments.

Chapter 2

Literature Review

GEI, performance and stability analysis.
Statisticians and breeders have studied GEI problem as it complicates efficiency in selection of high performing and stable cultivars during evaluation. Statistical techniques; both parametric and non-parametric minimizing GEI effect on cultivars selection exists Silveira et al. (2013), Karimizadeh et al. (2012), Annicchiarico (1997), Mohammadi and Amri (2012), Parmar et al. (2012) and have been used in overcoming GEI problem. Parametric univarivate (linear regression analysis and variance components) and multivariate approaches are based on statistical assumptions and considers the underlying distribution of a dataset Karimizadeh et al. (2012). Pereira et al. (2012) analyzed GEI using curvilinear regression. Multivariate approaches (Additive Main effects and Multiplicative Interactions (AMMI), Principal component analysis (PCA), and genotype plus GEI biplot (GGE) analysis) for GEI are explored by Yan et al. (2007). Multiplicative GEI are complex and should be summarized by two or more stability parameters under univariate, Karimizadeh et al. (2012) but multivariate approaches extract more information from GEI components by exploring the multi-directional aspects Miranda et al. (2009).AMMI analysis is one of the most effective multivariate techniques, the process involves evaluation of cultivars using least square technique in ANOVA for additive effect and PCA for multiplicative effects of cultivars in diverse environments.

Lavoranti, Santos Dias, and Kraznowski (2010) concurred that AMMI model comprehensively analyzes GEI structure in MET, offering better ways of interpretation and understanding of GEI but lament that it lacked ways of assessing stability of its estimates. He proposed bootstrap re-sampling in AMMI modeling and used it to get graphical and numerical analysis of stabilities of Eucalyptus grandis genotypes. Bootstrap coefficient of stability with squared Mahalanobis distance of scores differentiated genotypes and environments while graphical analysis of AMMI biplot gave better understanding and interpretation of yield stability. The proposed AMMI bootstrap eliminated uncertainties

created by low scores in ordinary analyses. However, bootstrapping may have problems as same measurements are re-sampled, bootstrapped performance of genotype and environment may be difficult to interpret. Purchase, Hatting, and Van Deventer (2000) sorted the challenge in AMMI stability issue using the scores to generate the stability values.

Thirty six wheat genotypes form dialle and their parents were evaluated by Rad et al. (2013) in six environments with seed yields per plant being the performance measure under drought and non-drought stress conditions. Unlike Kahram et al. (2013), he used the genotype and genotype x environment interaction (GGE) in characterizing environments and stability. The ASV selected stable crosses while GGE-biplot models combined the six environments to two mega-environments and confirmed the stable and high performing genotypes.

Karimizadeh et al. (2012) used the ANOVA technique to test for interaction effect, stability and performance. Using the type I stability concept, they identified most stable genotypes and types (II, III and IV) stability concept for the most favorable genotypes. Using clusters analysis they clustered the genotypes based on stability properties and mean yield groups. The findings were that regression methods slopes, genotypic stability, H statistic and desirability index (DI) which benefit type II and dynamic stability concept be recommended for GEI studies and yield stability. That was a complete deviation from AMMI by incorporating multivariate aspect in of clustering.

Kahram et al. (2013) evaluated GEI for durum wheat genotypes in moderate region of Iran by applying AMMI analysis and ASV and Rad et al. (2013) evaluated 36 wheat genotypes form dialle and their parents in six environments with seed yields per plant as performance measure under drought and non-drought stress conditions. Unlike Kahram et al. (2013), he used genotype and genotype x environment interaction (GGE) advocated for by Yan et al. (2007) in characterizing environments and stability. The ASV selected cross number 14 (Irena Veery) as stable while GGE-biplot models combined the six environments to two mega-environments and confirmed the stable and high performing genotypes. In environment 3 (F3 population, drought) that had an inbreeding depression effect, hybrid number 17 (S-78-11 Chamran) was best line based on its stability and high yield.

Amira et al. (2013) examined comparative discriminatory abilities of GEE and AMMI

models in selection of performing and stable tropical soybean genotypes. Their concepts were similar to Rad et al. (2013). They evaluated six genotypes in ten environments. AMMI revealed the most variable genotype with high interaction principal component axes (IPCA) and more stable environments for soybean genotypes evaluation. The most promising and stable genotypes across the test locations were identified. Their results showed GGE biplot as superior, effective and informative stability model in mega-environment analysis as compared to AMMI analysis. They showed that AMMI and GGE are applicable in the evaluation of performance and stability of any crop where GEI is present.

Josse et al. (2014) proposal of treating AMMI the Bayesian way as means of solving major over parameterization problem used real plant and simulated data, they ignored issues at prior level but applied the best processing at posterior level to get interpretable inferences using win bugs, open bugs and the Just Another Gibbs Sampler (JAGs) Bayesian software. Other than the issues of performance and stability they suggested a new solution to the estimation of risk of genotypes not exceeding a given performance threshold.

Tadege, Utta, and Aga (2014) reviewed statistical tools identifying better performing genotypes in diverse environments and their relation in describing GEI and cultivars stability. They showed that Shuklas stability variance (σ^2) and Wricke's ecovalence (W_i) were perfectly correlated by spearmans rank correlation. They also showed a highly significant positive rank correlation with coefficient of determination (r_i^2), deviation from regression (S_{di}^2), AMMI stability value (ASV), variance of ranks $(S_i^{(2)})$, mean absolute rank difference $(S_i^{(1)})$ and rank sum (RS), indicating their similarity in cultivar ranking. They grouped the statistical methods as dynamic concept of stability, static concept of stability and yield performance measures using PCA and suggested use of one dynamic concept of stability measure and yield performance measures for efficient cultivar recommendation.

Tadege, Utta, and Aga (2014) results concurred with those of Roostaei, Mohammadi, and Amri (2014) that undertook rank correlation among joint regression analysis (JRA), AMMI analysis, GGE biplot analysis and yield–stability (YSi) statistic in evaluating GEI for 20 winter wheat genotypes in 20 environments for yield and stability. GGE biplot and AMMI analysis were significantly correlated $(P < 0.01)$. AMMI distance (AMMID), regression deviation (S^2di) variance in JRA $(r = 0.83)$ and Shukla stability variance (σ^2) in YSi $(r = 0.86)$ were highly correlated $(P < 0.01)$ indicating that they could be used interchangeably. No correlation existed between yield ranks and stability ranks (AMMID,

S^2di, σ^2, and GGE stability index) showing that they measure static stability and could be used for selection based purely on stability. Yield stability and rank correlation varied among statistical methods.

Hongyu et al. (2014) addressed GEI using AMMI, the effects of genotypes (SSG), GEI signal (GES), and GEI noise (GEN)) sum of squares from combined ANOVA provided preliminary worthiness of AMMI. The SSG is a product of error mean square (EMS) and degrees of freedom (d.f.) for GEI and GES is GEN subtracted from GEI. They postulate that AMMI analysis is appropriate for datasets that have substantial G and GES and more so when the SS for GES is at least as large as that of G. When GEI is buried in noise, with the SS for GEN approximately equal to that for GEI, GEI should be ignored and AMMI analysis becomes inappropriate. That was a significant contribution for pre-determining worthiness of AMMI. However, many studies using AMMI seem not to heed their suggestions and still got good results in GEI analysis.

Degrees of freedom and significance of effects

Using four data sets of different cereal crops to test for the consistency of the significance of components by Gollob's 1968 F-test, F_{GH2} test, F_R test and the heuristic criterion based on the signal-to-noise ratio test, Annicchiarico (1997) found that Gollob's F_{GH2} test appeared more liberal than the F_R test. Dias and Krzanowski (2006) compared Eastment Krzanowski, Gabriel, Gollob, Cornelius and original singular values squared methods for sufficient components determination and found Eastment Krzanowski stable and appropriately behaving with small number of 'important' components, but underestimates when there is a larger number. Cornelius behaved appropriate in the presence of 'important' components, but was less stable than Eastment-Krzanowski. Gollob was similar to Cornelius method but with slightly worse stability and had likelihood of choosing more components in some situations. They preferred Eastment Krzanowski for cross-validation and recommended Cornelius method as F test method. If parsimony is a major concern then the former is preferred otherwise the latter is preferable when large numbers of interaction components are expected.The test of hypothesis about the k^{th} component $H_o : \lambda_k = 0$ using a complete dataset based on sequential sum of squares explained by the multiplicative terms. When the there are many significant IPCs the number explaining 70% proportion of variation and above is used or the Scree plot and application of the elbow rule.

Forkman and Piepho (2014) suggested the use parametric bootstrap methods for selecting principal components in PCA; the GEI data matrix with rows (genotypes) and columns

(environments) is standardized to have zero means. PCA uses the covariance matrix of that GEI matrix. The variances of the computed principal components are proportional to the squared singular values of the matrix. The large singular values indicate important principal components and parametric bootstrap is used to test for their significance. However they proposed that the performance of the method in PCA deserved further study.

2.1 Biplot analysis

AMMI biplot analysis is a multivariate visualization technique showing genotype stability, contribution to complexity of GEI, delineation of environments and narrowing of adaptation of genotypes to environments. It graphical represents genotypes and environments in a biplot by exploiting matrices U, S and V from SVD of GEI in determining positions in the interaction of singular axes (Garcia Pena and Dias, 2009, Hongyu et al. (2014)). GEI is displayed in two-dimension where elements are approximated by the inner product of vectors corresponding to the appropriate genotypes and environments. It investigates the pattern of genotypes response over different environments. It's important in substantially increasing information available from PCA without additional computations.

Biplot analysis delineate mega environment as crossovers between winning genotypes in environments would necessitate subdivision of a growing region into two or more mega-environments for exploitation of narrow adaptations that provides opportunity for substantial increase in yield. Model diagnosis is essential as changes in AMMI model changes the mega environment. Higher-order AMMI models defines a larger number of mega-environments. Model diagnosis enables researchers to distinguish between GES causing actual narrow adaptations and GEN generating spurious complexity (Gauch (2013), Hongyu et al. (2014)).

2.2 Imputation

Missing data challenges in Agricultural research results from pest and diseases destroying plots of experiment, failure in data collection and use of different controls (standards) in different environments especially in multi environment (MET) trials that results in gaps

when environments are merged for analysis. Data imputation overcomes the challenge by replacing the missing with plausible values for valid analysis on a complete data set Bergamo, Dias, and Krzanowski (2008). Existing techniques for missing data imputations have different levels of biasness and effects on the representativeness of final results.

Sugarcane breeding (MET) involves the use of different treatment controls (standards) for different environments. Genotypes may not be uniformly present in all environments as some are dropped in early stages of the breeding program hence resulting in unbalanced dataset. Possible solutions include extracting balanced subset by deleting the environments and genotypes with missing values. Disadvantage is that it removes all the information about the removed genotypes or environments from the analysis and subsequent interpretation. The best solution is to estimate the missing values and conduct analysis on complete data set Paderewski (2013).

Techniques for completing GEI data matrix are referenced with Arciniegas-Alarcón et al. (2010),Arciniegas-Alarcón et al. (2014) and Hourani and El Emary (2009) and Arciniegas-Alarcón et al. (2013). In choosing imputation technique, reality should take precedence as imputation isn't just for the sake of having complete dataset but a dataset reflecting reality. AMMI and related principles (such as PCA, SVD) and Biplot statistical techniques commonly used in MET with GEI require complete dataset Gauch (1992) Yan and Kang cited in M. and R., 2014, Paderewski (2013). Decision on the most efficient imputation method for the prevailing matrix would be a challenge as they produce different results.Troyanskaya et al. (2001) comparison of SVD imputation (SVDimpute), row average and weighted K-nearest neighbors (KNNimpute) using real datasets with some percentage missing and realized that KNNimpute was robust and sensitive than SVDimpute for estimation but generally SVDimpute and KNNimpute were better that row average method.

Yoon, Lee, and Park (2007) extended local least square imputation (LLSimpute) method by using quantile regression and estimated PC of complete subset of similar genotypes in imputing missing values. Their robust least squares estimation combined with PC (RLSP) method was evaluated against LLSimpute, Bayesian principal component analysis (BPCA) and kNNimpute method using normalized root mean squares error (NRMSE). RLSP outperformed the rest but was competitive to the BPCA.

Arciniegas-Alarcón et al. (2014) evaluated the Biplot imputation, distribution free Imputation (DFRI), Gabriel Eigen imputation, Expectation maximization SVD (EM-SVD) and EM-AMMI techniques and found the most efficient methods as EM-SVD and EM-AMMI0

while the least efficient were Biplot and EM-AMMI1. The Gabriel Eigen and DFMI methods consistently lied intermediately between EM-SVD and EM-AMMI. With 10-20% missing values, Gabriel Eigen was better than DFMI, but when the percentage increases to 40% DFMI was preferable. He concluded that EM-SVD was a very competitive alternative to EM-AMMI models with respect to additive model which had the disadvantage of ignoring interaction. EM-AMMI0 or EM-AMMI1; low AMMI model are recommended as associated errors increases with an increase in the number of multiplicative components Arciniegas-Alarcón et al. (2014) and Paderewski and Rodrigues (2014).

EM-SVD and EM-AMMI uses expectation maximization algorithm in imputation process. The points of departure is that EM-SVD initial imputation values are the columns means while EM-AMMI uses estimates calculated by subtracting the overall means effect from the sum of genotype mean effects and environment means effect. Expectation step involved the use of involves root mean square predictive difference (RMSPD) model for the least error with respect to the precision for EM-SVD and Chebycheves distance in EM-AMMI. The Maximization step involves the SVD of complete GEI matrix for EM-SVD and error matrix for EM-AMMI. EM-AMMI models estimates the missing values and the processes are repeated through iteration until convergence based on some set conditions. EM-SVD and EM-AMMI were proven to be efficient as compared to others Arciniegas-Alarcón et al. (2014).

2.3 Summary of literature

GEI in MET is addressed using univariate, multivariate and Bayesian approaches. Most univariate techniques reviewed had weaknesses overcome by multivariate technique. However, Tadege, Utta, and Aga (2014) showed that univariate and multivariate technique were highly correlated in determining stability. GGE increase the account of genotype variation as oppose to ordinary AMMI biplot. Failure of ASV to incorporate the cultivars performance as per (Purchase, Hatting, and Van Deventer, 2000), it is enhanced by the biplot analysis, yield stability index and rank sum test that provide information on performance, stability, adaptability and environment delineation. It is evident that AMMI, AMMI Biplot and ASV are the better ways of handling GEI problem. Their use are more common with other crops but limited in sugarcane which is a perennial crop with a number of harvesting cycles. The Bayesian perspective of the AMMI proposed by Josse et al. (2014) is worth exploring given that now analytical software's are available as

open source. A number of scientists have differed on when the AMMI model is appropriate for GEI analysis. Hongyu et al. (2014) argues that the SS GEI must be sufficiently large as compared to SS genotype. However many studies had not take that into consideration and still had good results. His suggestion that even with the presence of GEI, ANOVA should be used when the interaction is buried in the noise is alright but that has to be confirmed by AMMI itself to validate the use of ANOVA. AMMI should be used to address the GEI problem together with the bi plot analysis, ASV and supplemented by the non-parametric statistic rank sum test to account for performance. Many ways of selecting sufficient component explaining the interaction are confusing as they seem to differ in methodology and interpretation. It would therefore depend on individual researcher. When many components are significant, Cornelius method can be used as it reduces liberality of Gollob's. However with very few significant principal components Gollob's will be able to detect GEI effects.

Chapter 3

Methodology

The chapter provides methods for analysis of nine individual environments (cycles) as ANOVA, combined environments ANOVA, the test for assumptions, AMMI analysis in the presence of GEI and effects of different sources of variations.

3.1 Data; Genotypes, Environments and Design of Experiment

Secondary data variable that were used included test environment, crop cycle, replication, genotypes and yield performance in tones cane per hectare (tch). Genotypes were 33 from Mtwapa Series 2006 (MS2006) stage 5 and seven controls differing across zones. Crop cycle subset represented the nine environments. This arrangement was adopted given the limited data with some sites providing only crop cycle. The nine environments were MuhoroniPC, MumiasPC, MumiasRC1, Nzoia, Sony420pc, Sony527B-PC, Sony527B-RC and West-Kenya-PC. Sites experimental designs were Randomized Complete Block (RCBD) of three replicates (blocks) with differing randomization and number of genotypes.

3.2 Exploratory Analysis

Box plot visualized the mean performance of 33 genotypes,seven controls and nine environments, there minimum and maximum yields, dispersion from mean performance, outliers and whether data was from normally distributed population.

Initially environments were six grouped as specific zones (Muhoroni, Mumias, Nzoia, Sony420, Sony527B and West-Kenya). The analysis process showed violations of the required assumptions as most of the zonal yield data failed normality and homoscedasticity tests. Logarithmic transformation to improve on the assumption failed in normalizing the data and homogeneity of variance also failed. Sub-setting of the zones by crop cycles help in meeting the necessary assumptions.

3.3 ANOVA Statistical methods

The ANOVA was done for the nine individual environments (crop cycle) and combined MET with both environment and genotypes considered fixed.

3.4 Individual environment ANOVA

ANOVA statistical technique compares means of groups with continuous observations where groups are defined by the levels of factors, explanatory variables are categorical and all the elements of design matrix X are dummy variables. The choice of dummy variables could be arbitrary; an important consideration is the optimal choice of specification of X. The structure of data in a block design is as in table 3.1.

TABLE 3.1: Two way data structure for individual environment

		Genotypes					
		G_1	G_2	$\cdots$	G_{18}	Total	Means
Blocks	B_1	$Y_{1,1}$	$Y_{1,2}$	$\cdots$	$Y_{1,18}$	$Y_{1.}$	$\bar{Y_{1.}}$
	B_2	$Y_{2,1}$	$Y_{2,2}$	$\cdots$	$Y_{2,18}$	$Y_{2.}$	$\bar{Y_{2.}}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	B_3	$Y_{3,1}$	$Y_{3,2}$	$\cdots$	$Y_{3,18}$	$Y_{3.}$	$\bar{Y_{3.}}$
Total		$Y_{.1}$	$Y_{.2}$	$\cdots$	$Y_{.18}$		$Y_{3,18}$
Means		$\bar{Y_{.1}}$	$\bar{Y_{.2}}$	$\cdots$	$\bar{Y_{.18}}$		$\bar{Y_{...}}$

ANOVA for 18-genotypes and controls, replicated 3 times for the determination of genotype, replication, experimental error and the total effects. The model is given as equation

3.1.

$$Y_{ij} = \mu + G_i + B_j + e_{ij} \tag{3.1}$$

Where Y_{ij}- the response (yield) of genotype i in block j, μ - the overall mean, G_i - the i^{th} genotype effect $(i = 1, 2, \cdots, 16 and 18)$, B_j - the j^{th} block $(j = 1, 2, 3)$, ε_{ij} - experimental error. Genotype and block effects are fixed and the conditions of the model are $\sum g_i = 0$ and $\sum b_j = 0$ where $\sum g_i = b_j \sum \bar{g}_i - u$ and $\sum b_j = g_{ij} \sum \bar{b_j} - u$. The assumptions are that $Y_{ij} \sim (u, \sigma^2)$ and $\varepsilon_{ij} \sim (0, \sigma^2)$, $i = (1, 2, \cdots, 16 and 18)$ and $j = (1, 2, 3)$.

Normality of yields (Y) and error terms were tested using the QQplot and Shapiro Wilk test. Transformation (i.e. logarithmic) was used for data from non-normal distributed population, failure to which Central Limit (CL) theorems and laws of large number was applied for sample more than 30 for both individual test site and the combined MET. Graphics includes histogram, quartile-quartile plot and plotting of fitted value against the residual for scatter and patterns checking.

3.5 Parameter estimation of individual site ANOVA

Least square estimation (LSE) technique estimates the effect of the sources of variation for a two way structure. The sum of squares are formulated as shown in the ANOVA table 3.2 with grand mean given by;

$$\bar{Y_{...}} = \sum_{i=1}^{16 or 18} \sum_{j=1}^{3} \frac{Y_{ij}}{N} = \bar{Y}.$$

TABLE 3.2: ANOVA table for individual environment

Source of variation	d.f.	Sum of Squares (SS)	MS	F
Genotype	$g-1$	$b_j \sum_{i=1}^{16or18} (\bar{Y_{i.}} - \bar{Y_{..}})^2$	$\frac{SSG}{g-1}$	$\frac{MSG}{MSE}$
Block	$b-1$	$g_i \sum_{i=1}^{3} (\bar{Y_{.j}} - \bar{Y_{..}})^2$	$\frac{SSB}{b-1}$	$\frac{MSB}{MSE}$
Error	$bg-b-g+1$	$\sum_{i=1}^{16or18} \sum_{j=1}^{3} (Y_{ij} - \bar{Y_{i.}} - \bar{Y_{.j}} + \bar{Y_{..}})^2$	$\frac{SSE}{d.f.}$	
Total	$bg-1$	$\sum_{i=1}^{16or18} \sum_{j=1}^{3} (Y_{ij} - \bar{Y_{..}})^2$		

The hypotheses test on genotype effects $H_o : g_1 = g_2 = \cdots = g_{16or18}$ vs. $H_a : g_i \neq g_j \ (i \neq j)$ and block effect $H_o : b_1 = b_2 = b_3$ vs. $H_a : b_i \neq b_j \ (i \neq j)$ are tested using the F-test, significance determined by comparing computed F-value to the critical F value $(p < 0.05)$ at $\alpha = 0.05$. (Table 3.2). The Error means Squares (EMS) determines whether environments are merged for combined ANOVA.

3.6 Test for homogeneity of error variance

Combining data from all test environments requires homogeneity of error variances. For two environments EMS, F-test is sufficient while for nine environments EMS like in this case, a chi-square test (Bartlett's test) is appropriate. F-value is ratio of two $\frac{S_1^2}{S_2^2}$ variances, with larger variance as numerator and the smaller variance as denominator. It tests homogeneity of EMS under the hypothesis $H_o : \sigma_1^2 = \sigma_2^2$ vs $H_a : \sigma_1 \neq \sigma_2^2$. If$F_c > F_{\alpha,df1,df2}$, Ho is reject and EMS are heterogeneous otherwise homogeneous. The study had nine environments and therefore Bartlett's test (equation 3.2) was applied.

$$B = \frac{(N-K)\ln(S_p^2) - \sum_{i=1}^{k}(n_i - 1)\ln(S_i^2)}{1 + \frac{1}{3(k-1)}(\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{N-K})} \sim \chi^2_{\alpha,n-1} \tag{3.2}$$

Where $N = \sum_{i=1}^{k} n_i$ and $S_p^2 = \frac{1}{N-K}\sum(n_i - 1)S_i^2$ being the pooled estimate of the variance. $K = 9$ is number of test environments. The hypothesis tested $H_o : \sigma_1^2 = \sigma_2^2 = \cdots \sigma_9^2$ vs $H_a : \sigma_i \neq \sigma_j^2$. If the computed Bartlett's statistics $B > \chi^2_{\alpha,n-1}$, reject H_o and declare heterogeneity of the EMS. If $B < \chi^2_{\alpha,n-1}$, fail to reject the H_o and so homogeneity exists and the MET data is combined for analysis.

3.7 Combined ANOVA

Environment and genotype (both fixed), blocks nested within every environment, GEI and experimental error are the sources of variation in combined ANOVA. Significant GEI complicates genotype recommendation in terms of performance, stability and adaptability. GEI problems are addressed through AMMI analysis. Conditions for AMMI analysis are discussed in studies by Gauch (1992) and Gauch (2013) as cited by Hongyu et al. (2014). The overall effect, genotype effect, environment effect and GEI effect are computed from the two way table of the GEI means. Combined ANOVA had 33 genotypes and seven controls as (G), nine environments as (E) and three replicates (blocks) as (B), the model being equation 3.3.

$$Y_{ijk} = \mu + E_i + B(E)_{jk} + G_j + (GE)_{ij} + e_{ijk} \tag{3.3}$$

Where Y_{ijk}- Yield response variable, μ - overall mean effect, G_j - the j^{th} genotype effect $(j = 1, 2, \cdots, 40)$, E_i -i^{th} environment effect $(j = 1, 2, \cdots, 9)$, B_k - the k^{th} block $(k = 1, 2, \cdots, 3)$, e_{ijk} - experimental error. Conditions of the model are $\sum G_i = 0$, and $\sum E_j = 0$, $\sum(GE)_{ij} = 0$ The assumptions are that $Y_{ij} \sim (u, \sigma^2)$ and $\varepsilon_{ij} \sim (0, \sigma^2)$ $i = (1, 2, \cdots, g)$ and $j = (1, 2, \cdots, b)$, $E_i \sim Nid(0, \sigma_E^2)$, $GE_{ij} \sim Nid(0, \sigma_{GE}^2)$, $e_{ij} \sim Nid(0, \sigma_e^2)$. The normality assumptions are confirmed as in individual site analysis.

3.8 Combined ANOVA Parameter estimation

Least square estimation (LSE) technique estimates genotype, block and interaction effect from the two way GEI table. The effect of environment, genotype, GEI and total are $\bar{y}_{.j} - \bar{y}_{..}$, $\bar{y}_{i.} - \bar{y}_{..}$, $y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$, $y_{ij} - \bar{y}_{..}$ respectively. The degrees of freedom, sum of

squares, mean squares and variance ratio are as in table 3.3.

TABLE 3.3: Combined ANOVA table for all environments

Source of variation	**d.f.**	**Sum of Squares (SS)**	**MS**	**F**
Environment(E)	(e-1)	$g_i \sum\sum(\bar{y}_{.j.} - \bar{y}_{...})^2$	MSEnv	$\frac{MSEnv}{MSE}$
Block(Environments) B(E)	b(e-1)	$k\sum_{g-1}^{40}\sum_{b=1}^{3}(y_{ij.} - y_{...})^2$	MSB(E)	$\frac{MSB(E)}{MSE}$
Genotype(G)	(g-1)	$b_j \sum\sum(\bar{y}_{i..} - \bar{y}_{...})^2$	$\frac{SSG}{g-1}$	$\frac{MSG}{MSE}$
Genotype by Environment Interaction (GEI)	(g-1)(e-1)	$\sum_{g=1}^{40}\sum_{b=1}^{3}(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$	$\frac{SSB}{b-1}$	$\frac{MSB}{MSE}$
Error(e)	diff	diff	$\frac{SSE}{(bg-b-g+1)}$	
Total	(beg-1)	$\sum_{g=1}^{40}\sum_{b=1}^{3}\sum_{k=1}^{9}(y_{ij} - \bar{y}_{..})^2$		

3.9 Environment and Genotype effects

The environment effect tests differences between the test environments under the hypothesis $H_o : E_1 = E_2 = \cdots = E_9$ vs. $H_a : E_i \neq E_j \ (i \neq j)$ and the genotypic effect tests difference between the genotypes under the hypothesis $H_o : g_1 = g_2 = \cdots = g_{40}$ vs $H_a : g_i \neq g_j \ (i \neq j)$. With F-test at $\alpha = 0.05$, the H_o is rejected with a p-value < 0.05. The main effects are interpreted when there is no significant GEI.

3.10 GxE Interaction effect

The GEI effect tests whether genotypes performances changes significantly from one test environment to the other. Figures 3.1a and 3.1b illustrates the GEI for two genotypes in two test environments.

In figure 3.1a, G1 performs relatively high than G2 in environment one (E1) but low in two (E2) showing GEI. In figure 3.1b, G1 performs relatively high in both environment than G2 hence no GEI. GEI affects selection efficiency of the genotypes but also offers an opportunity for breeders and statisticians to identify different environments creating homogeneous regions and best cultivars with sustained performance irrespective of a

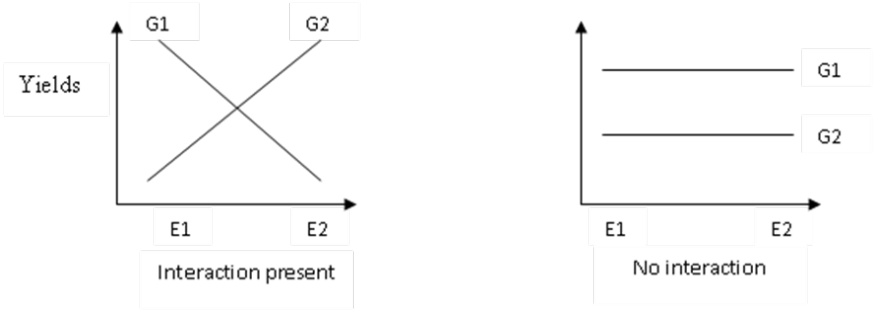

FIGURE 3.1: Illustration of GEI

change in environment. GEI significance requires AMMI modeling since ANOVA treats it superficially and fails to identify the interactions complexity.

3.11 AMMI modeling

AMMI modeling combines ANOVA and PCA for additive and multiplicative components respectively. The ANOVA uses LSE techniques to estimate the main effects while PCA uses the SVD technique to partition multiplicative component into individual variables by analyzing GEI matrix further into a non-random (pattern) and random (noise) parts. Analysis is applied to two-way GEI tables from MET with the assumptions and conditions of the ANOVA holding. AMMI model is defined in equation 3.4 and re-written in matrix notation in equation 3.5.

$$Y_{ij} = u + G_i + E_i + (GE)_{ij} + e_{ij} \tag{3.4}$$

Where Y_{ij} -Yield - a continuous response variable, μ-the overall mean, G_i -i^{th} genotype main effect corresponding to row factor i, $i = 1, 2 \cdots, 40)$,E_j -j^{th} environment main effect corresponding to j, $(j = 1, 2, \cdots, 9)$, $(GE)_{ij}$-i^{th} genotype and j^{th} environment interaction effect and e_{ij}-residuals.

$$Y = UI_gI'_e + GI'_g + I_eE' + X + e \tag{3.5}$$

$$\text{Y=}\begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,9} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,9} \\ \vdots & \vdots & \ddots & \vdots \\ y_{40,1} & y_{40,2} & \cdots & y_{4,09} \end{bmatrix}$$

U- is a scalar representing grand mean, I_g- (gx1) vector with elements of 1, I_e- (ex1) vector with elements of 1, $G - [G1, G2, \cdots, Gg]$- a gx1 vector of genotypes mean effects, $E - [E1, E2, \cdots, Ee]$- a ex1 vector of environmental mean effects

$$X\text{=}\begin{bmatrix} GE_{1,1} & GE_{1,2} & \cdots & GE_{1,9} \\ GE_{2,1} & GE_{2,2} & \cdots & GE_{2,9} \\ \vdots & \vdots & \ddots & \vdots \\ GE_{40,1} & GE_{40,2} & \cdots & GE_{40,9} \end{bmatrix}$$

a gxe (40x9) matrix with interaction effects elements of the i^{th} genotype and j^{th} environment and $\varepsilon = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_e]$ - a vector of the error terms.

3.12 Singular value decomposition (SVD) of the multiplicative component

Let the multiplicative term $(GE)_{ij}$ be a mxn (40x9) matrix X $(m \geq n)$ then $X = \sum_{i=1}^{k} \lambda_i u_i v_i = U_r \Lambda_r V_r'$ with $UU' = I$ and $VV' = I$ where U-matrix of orthogonal eigen vector associated with k (k=8) eigen values of XX'. V –matrix of orthogonalized eigen vector of $X'X$. Λ - gxe (8x8) diagonal matrix with the elements (i, j) where λ_i are singular values of the matrix X. If the matrix X is of rank r (r=8) then there are r (8) positive constants $\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_r$, r orthogonal m x 1 unit vectors $\mu_1, \mu_2, \cdots, \mu_r$ and r orthogonal k x 1 unit vectors $v_1, v_2, \cdots, v_r$. $U_r = [u_1, u_2, \cdots, u_r]$,$V_r = [v_1, v_2, \cdots, v_r]$ and Λ_r is an r x r (8x8) diagonal matrix with diagonal entries λ_i'. Baker (2005).

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

With

$$U_n = \left[\alpha_1, \alpha_2, \cdots, \alpha_n \right] = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{g1} & \alpha_{g2} & \cdots & \alpha_{ge} \end{bmatrix}$$

$$V_n = \left[\gamma_1, \gamma_2, \cdots, \gamma_n \right] = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{g1} & \gamma_{g2} & \cdots & \gamma_{ge} \end{bmatrix}$$

$$\Lambda_n = \left[\lambda_1, \lambda_2, \cdots, \lambda_n \right] = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

The normalization and orthogonal constraints; $I_g'U = I_e'V = O$, where O is a 1xn vector of zeros ($O = [0, 0, \cdots, 0]$ and $U'U = V'V = I_n = U_nU_n' = V_n'V_n$. I_n is an identity matrix (a matrix by its transpose gives identity matrix of 1's). n=8.

Given that $X = U\Lambda V' = \sum_{i=1}^{n} \lambda_i \alpha_i \gamma_i' = U_n \Lambda_n V_n'$. Substituting in the AMMI model 3.5 would give equation 3.6 in matrix which is the same equation 3.7

$$Y = UI_gI_e' + GI_g' + I_eE' + U\Lambda V' + e \tag{3.6}$$

$$Y_{ij} = \mu + G_i + E_j + \sum_{k=1}^{k'} \lambda_i \alpha_i \gamma_i + \theta_{ij} \tag{3.7}$$

$\theta \sim N(0, \sigma^2)$; $i = 1, 2, \cdots, S$, and constraints $\sum_{i=1}^{g} \alpha_{ik} = \sum_{j=1}^{e} \gamma_{jk} = 0$, $\sum_{i=1}^{g} \alpha_{ik}^2 = \sum_{j=1}^{e} \gamma_{jk}^2 = 1$ and $\sum_{i=1}^{g} \alpha_{ik}\alpha_{jk'} = \sum_{i=1}^{e} \gamma_{ik}\gamma_{ik'} = 0 \forall k \neq k', k = (1, 2, \cdots, n)$

Where Y_{ij}, μ, G_i and E_j are defined as in equation 3.4

α_{ik}- PC score for axis k (genotype)

γ_{jk}- PC score for axis k (environment)

θ_{ij}- the experimental error

n - Maximum number of the multiplicative terms n=rank(X)=min(g-1,e-1)
λ_k- kth singular value of the matrix X. It's also the square root of the Eigen value of the covariance matrix XX' and they are ordered $(\lambda_1, \lambda_2, \cdots, \lambda_n)$ implying $(\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_n)$.

μ, G_i, and E_j are additive parameters of AMMI model and λ_k, α_{ik} and γ_{jk} are multiplicative parameters. Interaction has pattern and noise, and the pattern is used for bi-linear modeling excluding noise and that change AMMI model to equation 3.8;

$$Y_{ij} = \mu + G_i + E_j + \sum_{k=1}^{s} \lambda_k \alpha_{ik} \gamma'_{jk} + \rho_{ij} + \theta_{ij} \tag{3.8}$$

Where; Y_{ij}, μ, G_i and E_j are defined as in equation 3.4. GEI is represented by the factors;
λ_k - a unique value of the kth interaction principal component analysis (IPCA), $(k = 1, 2, \cdots, p)$, where p is the maximum no. of estimable components. It is the PCA k^{kt} axis Eigen value or positive Eigen value of GEI matrix (ordered).
α_{ik} - is a singular value for the ith genotype in the kth IPCA. (i^{th} genotype PCA scores for the PCA axis k).
γ_{jk} - is a unique value of the jth environment in the kth IPCA. (j^{th} environment PCA scores for the PCA axis k.
θ_{ij} - the error for the GEI or AMMI residue ($residual - noise$ present in the data).
k - the characteristic non-zero roots, $k = [1, 2, \cdots, min(G-1, E-1)]$.
s - the number of multiplicative terms adequately addressing the GEI.
$\rho_{ij} = \sum_{k=1}^{k-s} \lambda_k \alpha_{ik} \gamma_{jk}$ - being the noise not accounted for by the multiplicative component.

AMMI model 3.8 is a family of models constituting the AMMI chain of models depending on the numbers of interactive term used. The chains of models is broken down in table 3.4.

TABLE 3.4: AMMI family of models

AMMI number	**Model**	**Comments**
AMMI-0	$Y_{ij} = \mu + G_i + E_j + \theta_{ij}$	Simplest and estimates additive main effects without interaction
AMMI-1	$Y_{ij} = \mu + G_i + E_j + \lambda_1\alpha_{j1}\gamma_{j1} + \rho_{ij} + \theta_{ij}.$	Combine main effect of AMMI-0 and interaction effects of first multiplicative terms
AMMI-2	$Y_{ij} = \mu + G_i + E_j + \lambda_1\alpha_{j1}\gamma_{j1} + \lambda_2\alpha_{j2}\gamma_{j2} + \rho_{ij} + \theta_{ij}.$	Combine main effect of AMMI-0 and interaction effects of first two multiplicative terms
⋮	⋮	⋮
AMMI-F	$Y_{ij} = \mu + G_i + E_j + \sum_{k=1}^{s} \lambda_k\alpha_{ik}\gamma'_{jk} + \rho_{ij} + \theta_{ij}.$	The saturated AMMI model (Full model)

3.13 AMMI model Parameters estimation

The least square estimation (LSE) technique fits parameters of AMMI model (equation 3.8). μ, G_i and E_j are estimated using the two-way ANOVA of interaction means $(\bar{Y}_{gxe})$. Equation 3.8 if reduced to $Y_{ij} = \mu + G_i + E_j + Z_{ij}$, Z_{ij} becames residual having patterns, noise and GEI (Z_{gxe}). Z_{ij} can be partitioned as $Z_{ij} = Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$, where y_{ij} - observations, $\bar{y}_{i.}$ -genotypes mean effect, $\bar{y}_{.j}$-environments mean effect and $\bar{y}_{..}$ -the grand mean effect. The interaction terms are estimated by SVD of Z_{ij} giving; λ_k- estimated by k^{th} singular value of Z.
α_{ik}- estimated by the i^{th} element of the left singular vector $\alpha_{k(gx1)}$.
γ_{jk}- estimated by j^{th} element of the right singular vector $\gamma_{k(gx1)}$.

The expectation of Z_{ij} will be $\hat{Z_{ij}} = Y_{ij} - \hat{e_j} - \hat{g_i} - \hat{\mu}$. The parameters $\lambda_k, \alpha_{ik} and \gamma_{jk}$ can be used to re-compute $\hat{Y_i}j = \sum_{k=1}^{k'} \hat{\lambda_k}\hat{\alpha_{ik}}\hat{\gamma_{jk}}$. Therefore $\alpha^*_{ik} = \lambda^c_k\hat{\alpha_{ik}}$ is the i^{th} genotype PCA score for the n^{th} axis and $\gamma^*_{ik} = \lambda^{1-c}_k\hat{\gamma_{jk}}$ is the n^{th} PCA score for the j^{th} environment and c is a scaling constant varying between 0 and 1. When genotype and environment are of equal

importance, the scaling constant takes the value 0.5. The estimates ($\hat{u}$, $\hat{G_i}$, $\hat{E_j}$ and $\hat{Z_{ij}}$) for the additive and interaction parameters (μ, G_i, E_j and Z_{ij}) are $\hat{u} = \bar{y_{..}}$, $\hat{G_{i.}} = \bar{y_{i..}} - \bar{y_{...}}$, $\hat{E_j} = \bar{y_{.j}} + \bar{y_{..}}$ and $Z_{ij} = y_{ij} - \bar{y_{i.}} - \bar{y_{.j}} + \bar{y_{..}}$

3.14 Estimation of the multiplicative effects using SVD

The residual $Z = \hat{U}\hat{\Lambda}\hat{V'} = \sum_{i=1}^{n} \hat{\lambda_i}\hat{\alpha_i}\hat{\gamma_i'} = \hat{U_n}\hat{\Lambda_n}\hat{V_n'}$ where; $\hat{U_n} = [\hat{\alpha_1}, \hat{\alpha_2}, \cdots, \hat{\alpha_n}]$ has n orthogonal units Eigen vector of ZZ' as the columns, $\hat{V_n} = [\hat{\gamma_1}, \hat{\gamma_2}, \cdots, \hat{\gamma_n}]$ -which has n orthogonal units Eigen vector of $Z'Z$ as the columns and $\hat{\Lambda_n}$ - an nxn diagonal matrix of estimated singular values $\hat{\lambda}_n > 0, n = 8$.

3.15 The sum of squares for the AMMI model

The effects and sum of squares for the environment and genotype main effects and multiplicative components $(GE)_{ij} = Z_{ij} + error)$ are $\bar{y_{.j}} - \bar{y_{..}}$, $\bar{y_{i.}} - \bar{y_{..}}$ and $y_{ij} - \bar{y_{i.}} - \bar{y_{.j}} + \bar{y_{..}}$ for the effects respectively and $g_i \sum(\bar{y_{.j}} - \bar{y_{..}})^2$, $\sum e_j(\bar{y_{i.}} - \bar{y_{..}})^2$ and $\sum\sum(y_{ij} - \bar{y_{i.}} - \bar{y_{.j}} + \bar{y_{..}})^2$ for the sum of squares respectively.

3.15.1 Sum of squares for the principal components

The sum of squares derivation for the multiplicative and error components are as in the first appendix. The principal component sum of squares is given by $S_k = \sum_{k=1}^{g}\sum_{k=1}^{e}(\hat{\lambda_k}\hat{\alpha_{ik}}\hat{\gamma_{jk}})^2$. Since $\hat{\lambda_k}$ is a constant, it is factored out such that $S_k = \hat{\lambda_k}^2 \sum_{k=1}^{g}\sum_{k=1}^{e}\hat{\alpha_{ik}}\hat{\gamma_{jk}}$. The sum of a product is the same as the product of the sum and therefore $S_k = \hat{\lambda_k}^2(\sum_{k=1}^{g}(\hat{\alpha_{ik}})^2(\sum_{j=1}^{e}\hat{\gamma_{jk}})^2$. From the constraint of the AMMI model $(\sum_{k=1}^{g}(\hat{\alpha_{ik}})^2 = (\sum_{j=1}^{e}\hat{\gamma_{jk}})^2 = 1$. Therefore $S_k = \hat{\lambda_k}^2 * 1 = \hat{\lambda_k}^2$ for $k = 1, 2, \cdots, min(g-1, e-1)$. With r replicates then $S_k = r\hat{\lambda_k}^2 * 1 = r\hat{\lambda_k}^2$ for $k = 1, 2, \cdots, min(g-1, e-1)$. (g=40, e=9)

3.16 Degrees of Freedom and Optimal number of the interactive Principle component (Model diagnostic and selection)

Degrees of freedom are assigned by Gollob's (1968) F-test are $(g-1)+(e-1)+(2k-1)$, with g-levels of genotypes, e-levels of the environments and k-the PCA level. The rank of GEI matrix is $s = min(g-1, e-1)$ where all s components provides saturated AMMI model. Fewer components sufficiently explaining the GEI are the optimal numbers of IPC required, the remainders being noise. The m $(m < s)$ components sufficiently explaining GEI gives a truncated AMMI model. Not all AMMI models best explain the GEI complexity and Gollob's technique is used in the selection where the models that accounts for the biggest variation $(> 70\%)$ is selected.

3.17 AMMI Biplot Analysis

Two AMMI biplots (Mean yields vs. PC1 and PC1 vs. PC2.) visualizes environment, genotypes, GEI, performance, stability and adaptability. Decomposition of GEI provides singular (eigen) values and eigenvector for the IPCs of genotypes and environments. Singular values are split for genotypes and environments eigenvectors in the second biplot. Mean yields vs. PC1 contrasts multiplicative terms and additive main effects while PC1 vs. PC2 indicating levels of GEI in data.

In singular-value partitioning, $\alpha^*_{ik} = \lambda^c_k \hat{\alpha_{ik}}$ (i^{th} genotype PCA score for the n^{th} axis) and $\gamma^*_{jk} = \lambda^{1-c}_k \hat{\gamma_{jk}}$ (the n^{th} PCA score for the j^{th} environment). Position of the i^{th} genotype is the i^{th} genotype scores $(\lambda^{\frac{1}{2}}\alpha_{i2}\lambda^{\frac{1}{2}}\alpha_{i1})$ while that of the j^{th} environment is j^{th} environment score $(\lambda^{\frac{1}{2}}\gamma_{j2}\lambda^{\frac{1}{2}}\gamma_{j1})$ on the biplot. The interaction effect of the i^{th} genotype in environment j is given by the projection of genotype position on the line of the environmental vector which has the slope $\frac{\lambda^{\frac{1}{2}}\alpha_{i2}}{\lambda^{\frac{1}{2}}\alpha_{i1}}$.

1^{st} biplot interpretation

- The distance from the origin determines the magnitude of interaction effect
- Angle between the i^{th} genotype and the j^{th} environment determines the strength of interaction.

- Acute angles, right angle and obtuse angles show positive, negligible and negative interactions respectively.

2^{nd} biplot interpretation

- Genotypes and environment points on the x-axis have similar interaction for the PC1 while those on the y-axis have similar interactions for the PC2.
- The genotypes in the 3rd quadrant have negative interaction along the PC1 and PC2.
- Genotypes and environments in the 1st quadrant have positive interaction with both PC1 and PC2 while the ones in the 2nd and 4th quadrant have different signs for interaction on the PC1 and PC2.
- Numerically genotypes or environment with higher scores on either PC1 or PC2 or both, whether negative or positive have higher contribution to the interaction and vice versa.
- Genotype and environment points close to the origin are contributing small to the interaction and are estimated by the additive main effects terms only.

3.18 Stability and adaptability analysis

3.18.1 AMMI stability value

The ASV (equation 3.9) by Purchase, Hatting, and Van Deventer (2000) is computed from IPCA1 and IPCA2 scores. Minimum ASV indicates most stable genotypes.

$$ASV = \sqrt{\frac{IPCA1_{ss}}{IPCA2_{ss}} IPCA^2_{score} + IPCA2^2_{score}} \tag{3.9}$$

3.18.2 Yield Stability Index (YSI) and Rank-Sum test (RS)

Yield Stability Index (YSI) and Rank-Sum test (RS) are extension of ASV incorporating yield performance by equation 3.10 and 3.11.

$$YSI = RASV + RY \tag{3.10}$$

Where; RASV - is the rank of ASV and RY - genotypes average yield ranks

YSI incorporate both average yield and stability. its Low value shows desirable genotypes.

$$RS = RM + Sd(RS) \tag{3.11}$$

Meaning Rank $sum(RS) = Ranksaverage(R) + Standarddeviationofrank(SDR)$. It incorporates both yield and stability in a single non-parametric index. Genotypes with the least RS values are stable with high yields. Standard deviation of rank (SDR) was measured as:

$$S_i^2 = \frac{\sum_{j=1}^{m}(R_{ij} - \bar{R}_{i.})}{i-1}$$

Where; R_{ij}- is the rank of X_{ij} within the j^{th} environment, (R) - is the mean rank across all environments for the i^{th} genotype and SDR= $(S_i^2)^{0.5}$.

3.19 Matrix Imputation

The mathematical framework are covered by Troyanskaya et al. (2001) and Arciniegas-Alarcón et al. (2014) for SVD imputation for single case of missing value. For multiple missing values like in this case, modification is done by initially imputing all missing values by the column means to give a complete GEI matrix that is standardized (mean centering the columns with m_j and dividing by standard deviation s_j). The imputation of the standardized GEI matrix for each cell corresponding to an original missing value is made using $x_{11}^m = X_{.1}^T V D^{-1} U^T X_{.1}$ where after imputing all the missing, the GEI matrix is reverted to its original scale using $x_{ij} = m_j + s_j \hat{x}_{ij}^m$. This process is iterated until it achieves convergence which is stability in successive imputed value.Imputation process depends on equation $X_{11} = \sum_{k=1}^{m} u_k d_k v_k^T$. The iteration process is done to achieve convergence based on some specified value Alter, Brown, and Botstein (2000) as cited by Troyanskaya et al. (2001) for expectation maximization method to arrive at the final estimates. Each missing value in GEI is estimated using the algorithm and then the procedure is repeated on the

newly generated matrix until the total change in the matrix falls below the precision desired (may be 0.01).

EM-AMMI imputations are based on the AMMI models and uses cross validation procedure together with root means square predictive difference (RMSPD) for possible principal components to get the best imputation model. Convergence of the imputation process occurs when maximum changes in the predicted cell are less than 0.001 but can be adjusted depending ones desire. EM-AMMI0 converges automatically with number of iterations limited to 1000. The AMMI model is defined as equation 3.8 with all the conditions holding, it's shortened as $Y_{ij} = \mu + G_i + Ej + Z_{ij}$ where Z_{ij} is residual having patterns and noise and contains the interaction Z_{gxe}.

The maximum number of interactive principal components (max.IPC) is the minimum of either rows of GEI or its columns minus 1 while the possible number of principal components for imputation is given by $PC.nb = min((g-1),(e-1))$. When maximum interactive principal component is less than the number of PC requested for in imputation, the maximum interactive PC is used. The initial estimate values for imputing are the sum of the GEI genotypes mean effects and environments mean effects subtracted the overall mean effects which provides a new complete GEI matrix with estimates for the missing data.

Once the conditions are met if the PC needed for imputation ≥ 1 then SVD is done on the new matrix $(SVD(newmatrix) = UDV')$ in order to get new interaction adjusted matrix through dimension reduction, otherwise the interaction adjusted becomes zero (0).

EM-SVD and EM-AMMI both impute using expectation maximization process. The points of departure is that EM-SVD initial imputation values are columns means while EM-AMMI uses estimates calculated by subtracting the overall means effect from the sum genotype mean effects and environment means effect. The expectation and maximization step involves determining the most stable imputed values by repeating the process through iteration until convergence based on some set conditions are in table 3.5.

3.19.1 AMMI imputation using EM-AMMI

The procedure for undertaking an EM-AMMI imputation; Additive parameters are set initially by computing the overall mean, genotype mean and environment mean from

TABLE 3.5: EM-SVD and EM-AMMI imputation

Areas	EM-SVD	EM-AMMI
Initial imputation with external values	Impute all the missing values with the column means	Impute all the missing values with estimated values
Maximization step	Maximization of the complete matrix of rank k using singular value decomposition $SVD(X) = U_k \Lambda_k V_k^T$	Maximization by estimating the parameters of the model $r_{ij} = Y_{ij} - \mu - G_i - E_j$
The model in the process	The model is the Root mean square error $RMSE = \sum (w_{ij} - \hat{w}_{ij})^2$	Model is used in the imputation $Y_{ij} = \mu + G_i + E_j + Z_{ij}$
Expectation step	$w_{ij} = w_{ij} if ij$ $\in R, U_k \Sigma_k V_k$ otherwise, $w_{ij} = err + [u_k \Sigma_k v_k]_{ij}$ Implying that $err_{ij} = w_{ij} - \sum \sigma_k u_{ki} v_{kj}$	The chebychevs distance determines the expectation whereby if the difference between imputations is less that precision, the process repeats $D_{chebyshev}(p, q) = max\|p_i - q_i\|$
Processing	The algorithm alternates between SVD computation (maximization) and expectation $(= err_{ij} + [u_k \Sigma_k V_k]_{ij})$ until convergence	The process is iterated until convergence

the observed data.The residual of observed cells are initialized as $r_{ij} = \hat{Y_{ij}} - \hat{g} - \hat{e} + \mu$ .The interactions for the missing portion are initially set to zero. The initial multiplicative parameters of the AMMI are obtained from the SVD of the matrix of the residuals (r_{ij}). The values that missed are filled by the appropriate AMMI estimates and a normal procedure carried on. The algorithms for EM-AMMI are given by Paderewski (2013) as follows;

Step 1 Users can impute missing with any value initially. otherwise, initial imputation values are computed as overall mean plus main effects of rows (genotypes) and main effects of columns (environments) to fill the missing.

Step 2 The parameters of the AMMI model are estimated.

Step 3 The adjusted means are obtained from AMMI model with n principal components.

Step 4 The missing cells are filled with the adjusted means.

Step 5 If the maximum change in these values (Chebyshev distance in the two iteration steps) is larger than assumed precision, the steps 2 - 5 are repeated. Otherwise, the algorithm stops.

3.19.2 EM-SVD algorithm

The EM-SVD is well covered by Arciniegas-Alarcón et al. (2014) and the algorithm for imputation takes the following steps;

Step 1 Let $I = \{i, j\} : X_{ij}$ isn't missing be the full set of all observed values.

Step 2 For $1 < j < p$, let U_j be the mean of non-missing values in column j of A. Set U_i to zero if all column missing.

Step 3 Define $X^{(0)}$ by $X^{(0)} = X_{ij}$ if (I,j) $\in I$, U_j, otherwise

Step 4 Initialize the iteration count $N \leftarrow 0$

Step 5 Maximization: Compute SVD for $X_K^{(N)} = \sum_{i=1}^{p} d_i^{(N)} U_i^{(N)} V_i^{(N)T}$ and let $X_K^{(N)}$ and let X_K^N denote SVD truncated to k terms $X_K^{(N)} = \sum_{i=1}^{p} d_i^{(N)} U_i^{(N)} V_i^{(N)T}$.

Step 6 Expectation: Define nxp matrix $A^{(N+1)}$ as X^{N+1} as $X_{ij}^{N+1} = X_{ij}$ if (I,j) $\in$ I, $X_{k=ij}^N$ otherwise

Step 7 Set the residual sum of squares $RSS^(N) = ||X - X_K^N||_{F,I}^2$. If $||RSS^(N) - RSS^(N-1)|| <$ than some predefined small value then and the output X_K^N contains the missing values. Otherwise increase $N \leftarrow N + 1$ and return to step 5.

3.19.3 Comparison of EM-AMMI and EM-SVD imputation process and values

Availability, code complexity, iteration to convergence and imputation ability
Availability was evaluated in term of accessing of utility packages and codes for each technique. Code complexity, number of iteration for set conditions.

Efficiency of the technique Runtime
Processes time and system time are function R software determines efficiency of code processing. Used are Proc.time ;a stop watch function with timings are based on user time, system time and elapsed time which relating to code execution, use of the central processing unit (CPU) and time difference respectively. Faster processing time is an measure of efficiency.

Correlation and significant difference of the imputed value

Correlation coefficient $\rho_{xy} = \frac{Cor(r_x,r_y)}{\sigma_x\sigma_y}$ measures the degree to which the two imputation techniques variable movements are associated. Values range between -1 and +1 and closeness to either extreme shows strong correlation in that direction while those closer to zero indicate weak correlation.Student t test for the imputed values by the two techniques test for significant difference between the values imputed $H_o : \mu_1 = \mu_2)$ vs. $H_a : \mu_1 \neq \mu_2$. The test statistics $t = \frac{(\bar{x}_1-\bar{x}_2)}{se(\bar{x}_1-\bar{x}_2)}$ and standard error $se(\bar{x}_1 - \bar{x}_2) = \sqrt{S_p^2(1/n_1 - 1/n_2)}$ where; $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ - is the pooled variance, n-sample size and $n_1 + n2 - 2$ is the degrees of freedom. If the computed $t > criticalt$ at$\alpha = 0.05$ and $n_1 + n_2 - 2$ d.f reject the null hypothesis and there is a significant difference in the imputed value by the two techniques.

Error minimization in imputation

Imputation accuracy was determined using the predictive residual sum of squares (PRESS); $\frac{1}{np}\sum\sum(x_{ij} - \bar{x}_{ij}^m)$ and works by averaging the estimated errors in the imputation. It estimates accuracy of imputing 47% of missing and predicting 53% of the present data using imputed values. The least PRESS values give the best techniques.

Data evaluation using Principal component analysis and Biplot.

The complete GEI matrices imputed by EM-SVD and EM-AMMI techniques are subjected to PCA to check on data structure, correlations and dimensionality. It describes the variation in a set of correlated variables $x_i(i = 1, 2, \cdots, 9)$ in terms of a new set of uncorrelated variables y_i which are linear combination of x_i variables. The 1st PC of observations $y_1 = a'_{ij}x_1$ is linear combination $y_1 = a_1x_1 + a_2x_2 + \cdots + a_qx_q$ whose sample variance is greatest subject to $a_1a'_1 = 1$. The 2nd PC $y_2 = a_1x_1$ is linear combination $y_2 = a_1x_1 + a_2x_2 + \cdots + a_qx_q$ which has the greatest variance subject to $a_2a_2 = 1$ and $a_2a'_1 = 0$. The j^{th} PC being the linear combination $y_j = a_1x_1 + a_2x_2 + \cdots + a_qx_q$ which has the greatest subject to $a_ja'_j = 1$ and $a_ja_i = 0$. The choice of appropriate number of PC explaining variation GEI are based on elbow rule, number of PC with variance of 1 and above and number PC accounting for at least 70% of the total variation.

Chapter 4

Results

4.1 Exploratory analysis of the data

4.1.1 Varietal means performance

The performance of cultivars varied with different mean performances as shown by boxplot (figure 4.1). Outliers were observed for MS271 cultivar on both extremes. Cultivars Ms282, Ms271 and Ms800 had best performances with mean yields of 138.9, 125.66 and 122.61 tch respectively. The least performing was control CO421 with mean yield of 21.9tch. Cultivars Ms166, Ms30, Ms302, Ms303, N14, CO945, MS866 had the biggest variation as oppose to CO421, D8484, CB32-22, Ms830. The actual mean performances are in table 6.3 in appendices. Small variations show consistency in performance. The order of genotypes in x-axis are as in table 6.3.

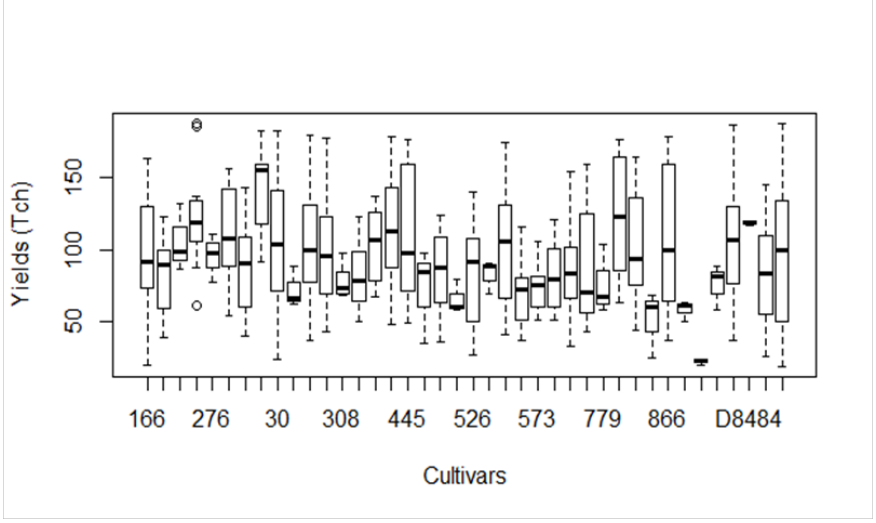

FIGURE 4.1: Genotypes boxplot

4.1.2 Environmental mean performance

Outliers were recorded for MuhoroniPC and MumiasPC and environmental means performances were 89.30, 59.69, 120.02, 61.76, 149.46, 84.30, 155.00 and 92.00 tch for Chemelil, MuhoroniPC, MumiasPC, MumiasRC1, Nzoia, Sony420pc, Sony527BPC, Sony527BRC and West Kenya PC respectively. The crop cycles for the different environments were significantly different as indicated by the mean performances in table 6.4 in appendix. Sony527BRc had the best performance for all the cultivars while Mumias PC had the least performance (table 6.4).

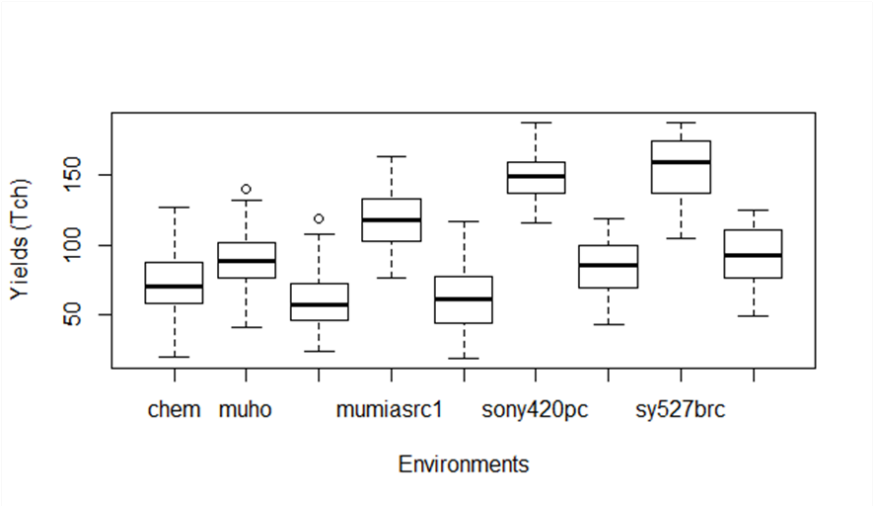

FIGURE 4.2: Environments mean performance(tch)

4.2 Environment ANOVA

ANOVA for the various environments (crop-cycles) were done with block and varieties (cultivars) as the main effects. Cultivars showed significant difference in Nzoia, Chemelil, West Kenya, Sony527brc and MumiasRC but not in Sony527B, Muhoroni, MumiasPC and Sony420PC (table 4.2). The lack of consistence in performance of cultivars across environments created an uncertainty of performances, stabilities and the adaptabilities, thus further analysis undertake for clarity. Cultivars and controls mean separations in individual environments are in table 6.7 and 6.8 in appendices.

4.3 Test for the ANOVA assumptions

4.3.1 Normality of the response variable (yield) and the error term

Yield response and error terms for all environments come from normally distributed population as indicated by Shapiro Wilks test and qqplot (figures 6.1-6.7 in appendices) normality test except for Sony527brc (table 4.1). Combined environments yield data was slightly sigmoidal with a Shapiro Wilks statistic of W = 0.97557, and a P-value = 5.444e-07. Yield transformation (i.e logarithmic) failed in (Shapiro Wilk statistic of W = 0.96741 and p-value = 1.309e-08) and the Central limit theorem was applied given sample was n=462. It equally applied to Sony527brc as $n > 30$. Normality assumption for error terms were checked confirmed using qqplot of fitted values and residual for all environments. The scatters were uniform and showed no relationship between fitted values and residual.

TABLE 4.1: Normality of the response variable (yield) by environment

Environments	N	Shapiro Wilk statistic (W)	P-value	Normality
Nzoia	54	0.98821	0.8714	Normal
Sony527B	48	0.97598	0.97598	Normal
Chemelil	54	0.99041	0.9424	Normal
Muhoroni	48	0.99126	0.9751	Normal
Mumias	54	0.96742	0.1481	Normal
West Kenya	54	0.96213	0.08585	Normal
Sony527brc	48	0.93088	0.007349	Not normal*
Sony420PC	48	0.97478	0.384	Normal

*Applying central limit theorem (CLT) stating that, based on certain conditions, the arithmetic mean of a sufficiently large number of sample of independent random variables whose expected values are well defined and approximately normally distributed finite variance is irrespective of the distribution. The sample size was more than 30, hence sufficient.

TABLE 4.2: Individual environment Analysis

Environment	Source	d.f.	SS	MS	F	P-value
Nzoia	Block	2	863.3	431.64	1.3939	0.261936 ns
	Variety	17	15848.7	932.28	3.0105	0.003039 **s
	Error	34	10529.0	309.68		
Sony527B	Block	2	2404.1	1202.04	5.0346	0.01302 *s
	Variety	15	6880.4	458.69	1.9212	0.06246 ns
	Error	30	7162.7	238.76		
Chemelil	Block	2	6806.1	3403.0	17.7632	5.231e-06 ***s
	Variety	17	16469.2	968.8	5.0568	2.967e-05 ***s
	Error	34	6513.7	191.6		
Muhoroni	Block	2	1853.3	926.64	2.2213	0.1260 ns
	Variety	15	6333.5	422.23	1.0122	0.4694ns
	Error	30	12514.7	417.16		
MumiasPC	Block	2	270.1	135.05	0.2828	0.7554 ns
	Variety	17	4696.8	276.28	0.5786	0.8845ns
	Error	34	16234.0	477.47		
West Kenya	Block	2	1539.0	769.51	2.4144	0.10460 ns
	Variety	17	10807.0	635.71	1.9946	0.04252 *s
	Error	34	10836.0	318.72		
Sony527brc	Block	2	416.8	208.40	0.5305	0.59375 ns
	Variety	15	12660.9	844.06	2.1485	0.03639 *s
	Error	30	11785.8	392.86		
Sony420PC	Block	2	272.8	136.38	0.5731	0.56985 ns
	Variety	15	7055.0	470.34	1.9763	0.05479 .ns
	Error	30	7139.6	237.99		
MumiasRC	Block	2	874.9	437.46	2.0037	0.1504419 ns
	Variety	17	15074.7	886.75	4.0616	0.0002487 ***s
	Error	34	7423.0	218.32		

4.3.2 Homoscedasticity test.

Bartletts test had both block and environments as homogeneous grouping. With 2 degrees of freedom and Bartletts statistic of 1.2547 the p-value was 0.534 for block while with 8 degrees of freedom and statistic of 7.2249, the environments p-value was 0.5126

hence homogeneous. This allows the merging of data for combined ANOVA.

4.4 Combined Analysis of variance (ANOVA)

The ANOVA for genotype, environment and GEI effects were significant $(p < 0.05)$ indicating different behaviour of genotypes across environments thus making it important for a study to enable identification of the magnitudes of that interaction with environment. This is achieved through further analysis of the interaction effect to enable efficient selection of the best performing cultivars that are stable and adaptable. In the analysis output 71.76% of the total sum of squares attributed to environmental effects, 5.53% to genotypic effects, and 7.91% to GEI effects. Large environment sum of squares (SS) indicated large diversity with big differences among environmental means causing most of the variation in cultivars performance. The GEI SS was 1.43 times larger than that for genotypes, indicating bigger differences in genotypes responses across environments (table 4.3).The overall mean performance and difference amongst them and the controls are in tables 6.9 in appendices.

TABLE 4.3: Combined ANOVA

$Sources$	$D.f.$	$SumSq$	$Mean_{S}q$	F	$(Pr > F)$
Environment	8	511,415	63,927	205.6691	$< 2.2E-16 * **$
Variety	39	39,446	1,011	3.254	$6.176E-09 * **$
Environment(block)	18	15,300	850	2.7347	$2.345E-04 * **$
GxE Interaction	106	56,380	532	1.7112	$2.382E-04 * **$
Residuals	290	90,139	311		

Parameters estimation of the model; All environments contributed positively and significantly $(p < 0.05)$ to overall yields except for Mumias PC and Nzoia whose parameters are not different from zero. Cultivars and control Ms271, Ms279, Ms302, Ms313, Ms535, and D8484 contributed significantly and positively to the yields as oppose to Ms282, Ms300, Ms326, Ms556 and N14 that had negative contributions. Contributions of the rest were not different from zero. The blocks nested in the environment whose effect were positive to the yield were block 2 in Chemelil and West Kenya, block 3 in Chemelil and Sony527Pc. The rest had no significant effects hence the blocking was important. GEI contributions to the yield were positive for Ms759 in Mumias Rc1 and Nzoia, Ms77 in Mumias Rc1 and CO945. CO945 interaction with Sony527Bpc and Ms866 in Mumias Rc1 were negative.

The effects of the rest of the interaction were not significant. The parameter and their significance are shown in table 6.10-6.15 in appendix. The model for the parameters estimation was significant $(p = 2.2e - 16)$, the residual standard error of 18.68 on 290 degrees of freedom. The R^2 was 0.8735 and $Adj.R^2$ of 0.7989 which 79.86% of the variability in the yields was explained by the model. The d.f of freedom for the GEI were adjusted by 206 to account for the imputation.

4.5 Comparative Imputation on the GEI matrix

4.5.1 SVD Imputation and EM-AMMI imputed matrix

Two-way GEI table 4.4 was the data matrix of 40 genotypes and 9 environments with 57% missing data and is of rank k=8 (k=min (n-1,p-1)). EM-AMMI0 was the only possible model used in imputation as given by minimum genotype data present in rows (n) and environments data present in columns (p); min(n-1, p-1)-1 was zero hence the main effects (PC=0) were used in imputation. EM-SVD imputation depended on the lowest rank of the GEI and the complete subset matrix of data. The lowest rank of the matrix was rank 1.

4.5.2 Packages requirement, Efficiency in runtime, number of iteration for convergence

EM-SVD packages are not directly within the CRAN's but are archived hence making availability an uphill task without proper link. The dependencies are also many (table 6.16 and 6.17 in appendices) occupying space of the disk. Nonetheless, the fact that they exist whether in archives or elsewhere gives it upper hand. EM-AMMI packages aren't available on the CRAN's and the codes were developed by Paderewski and Rodrigues (2014) and are available online. Based on efficiency of execution of the generated codes for the two techniques, user-time and elapsed time are as indicated in table 6.16 and 6.17 in appendices. EM-AMMI0 took the least time in all the aspects and convergence earlier than the EM-SVD with few numbers of iterations and therefore one would prefer it to EM-SVD for processing efficiency. However the codes for EM-AMMI could be complex for novice person in programming R codes.

4.6 Correlation and significant difference of the imputed value

The imputed values by the two techniques had a very strong positive correlation (correlation coefficient of 0.937) showing that the imputations were related and in same direction. The paired t-test for the imputed values of the two techniques showed no significant difference given critical t = -1.7186, with df = 205 and p-value = 0.0872. The variances of the two techniques were not significantly different and equal of n=206 the pooled variance was used. Based on two results one would pick of the techniques with minor differences in the results given that there no significant difference between the EM-SVD and EM-AMMI imputation techniques for the prevailing GEI data matrix.

4.6.1 Predictive Residual Sum of squares-(PRESS)

EM-SVD and EM-AMMI0 uses cross validation procedure where one or a number of the data point are left out and the techniques applied to predict them. The errors in the prediction are used in determining the best imputation technique where the one with the smallest PRESS is better. The PRESS value in the prediction of the 47% data that was available yielded 55.18 and 118.86 for EM-AMMI0 and EM-SVD respectively. The imputed data were then used to predict the data that had been available and the error in the estimation calculated to give the PRESS where the method with the least PRESS value of the two is the best. EM-AMMI behaved in a similar manner in the first imputation of the missing data where it could not impute beyond the principal component one. Given that the cultivar Ms302 had all data present and the coding for EM-AMMI doesn't allow complete missing row or columns a one real value was use in the 1st column of that row. EM-AMMI using the additive components had a PRESS of 55.18 while EM-SVD had a PRESS value of 118.86.

4.6.2 GEI data matrices evaluation

The complete GEI dataset for the techniques were evaluated using the principal component analysis (PCA) and biplot. The correlation between environments shows the possibility of undertaking PCA. EM-AMMI complete data matrix had high correlations between environments; PCA could be useful in reducing its dimensionality unlike EM-SVD

whose complete data matrix had very low correlations in environments. The first PC accounts for 79.8% of total variation in the original data and the first two PCs account for 85.2% of total variation for the EM-AMMI0. In its scree plot, only component 1 is selected since its variance is greater than 1 (average variance). The resulting linear combination is;

$$Z1 = -0.349_{E1} - 0.353_{E2} - 0.328_{E3} - 0.318_{E4} - 0.320_{E5} - 0.323_{E6} - 0.337_{E7} - 0.347_{E8} - 0.323_{E9}$$

It explains 79.8% of the original variation. The variables in this component are relatively of equal importance (loadings are in the same range on average). In EM-SVD case, the first 5 components accounted for 76.8% of original total variation and revealed inefficiency of PCA in dimension reduction on this particular dataset as large numbers of components are retained. Principal components 1-5 from the scree plot are selected since their variance is greater than 1 (average variance), however, lacked the "elbow" in the scree plot. Figs 4.3

$$\begin{aligned}
Z_1 &= 0.000_{E1} + 0.588_{E2} + 0.499_{E3} - 0.000_{E4} - 0.475_{E5} - 0.000_{E6} - 0.000_{E7} - 0.405_{E8} - 0.323_{E9} \\
Z_2 &= 0.144_{E1} + 0.143_{E2} - 0.320_{E3} - 0.496_{E4} - 0.331_{E5} - 0.327_{E6} - 0.414_{E7} - 0.345_{E8} - 0.321_{E9} \\
Z_3 &= 0.518_{E1} - 0.160_{E2} + 0.000_{E3} - 0.179_{E4} + 0.124_{E5} - 0.431_{E6} + 0.482_{E7} + 0.266_{E8} + 0.406_{E9} \\
Z_4 &= 0.153_{E1} - 0.433_{E2} + 0.320_{E3} - 0.231_{E4} - 0.177_{E5} - 0.336_{E6} - 0.483_{E7} + 0.426_{E8} + 0.273_{E9} \\
Z_5 &= 0.594_{E1} + 0.000_{E2} + 0.000_{E3} + 0.384_{E4} - 0.101_{E5} - 0.382_{E6} - 0.343_{E7} - 0.248_{E8} - 0.401_{E9}
\end{aligned}$$

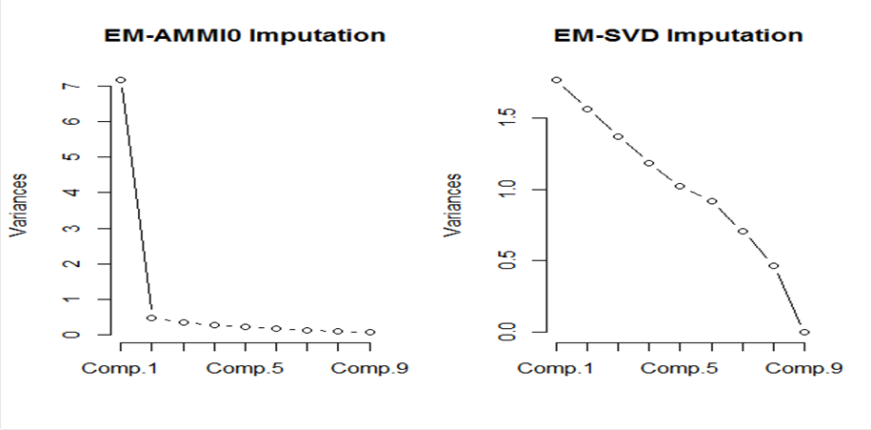

FIGURE 4.3: Screeplots for EM-AMMI and EM-SVD imputations

4.6.3 Biplot analysis

Biplots projects multivariate datasets by showing the variance covariance structure of the variables in this case genotype or environments, values of observations on variables and Euclidean distances between observations in the multidimensional space as quantities of

data matrix Kohler, Luniak, et al. (2005). EM-AMMI0 imputed values had three major groupings corresponding to the environments that formed V-shape in figure 4.4a. They display the variances and covariance of the environments and distances between them, length of vector from origin to the coordinates representing environment and genotype variances. Correlation between the environments or genotypes are reflected by the angle between two corresponding environments or genotypes vectors where smaller angles show greater correlations.

The three groupings of environments were group 1 (environments 1, 2, 3, 5, 6, 7 and 8), environment 4 and environment 9 appearing on x axis, 2nd and 3rd quadrants respectively and acute angles between them showing high correlations. EM-SVD produced diverse enviroments of six groupings. The major group had environments (1,4,5,6,7), environments 5,8,2,3 and 9 formed individual groupings. However, environments 1,2,4,5,6,7 and were highly correlated by having acute angles between them and appearing in first quadrant figure 4.4b. The environments numbering represents Chemelil (1),Muhoroni (2), MumiasPC (3),MumiasRc1 (4), Nzoia (5),Sony420Pc (6),Sony527BPc (7), Sony527BRc (8) and West Kenya (9).

The two techniques produced different data matrices described through PCA and biplot analysis. EM-AMMI would be better if one is investigating a single characteristic behavior as all variations are accounted for by the first PC while EM-SVD would be good if one is investigating multiple behavior that may be represented by the five PC.

4.7 AMMI modeling

The existence of GEI makes it difficult establishing superior and stable cultivars. AMMI analysis reveals patterns between cultivar and test environments through SVD or PCA techniques that identify those patterns, separate them from the noise that exists in GEI matrix and enabling realization of multiplicative effect of cultivars and environments. The GEI data matrix had 57% missing data shown with the NA cells (table 4.4). As per the description of Paderewski and Rodrigues (2014) the data was missing completely at random since at least every row and column had missing data and Imputation was undertaken using EM-SVD to proceed with the complete analysis.

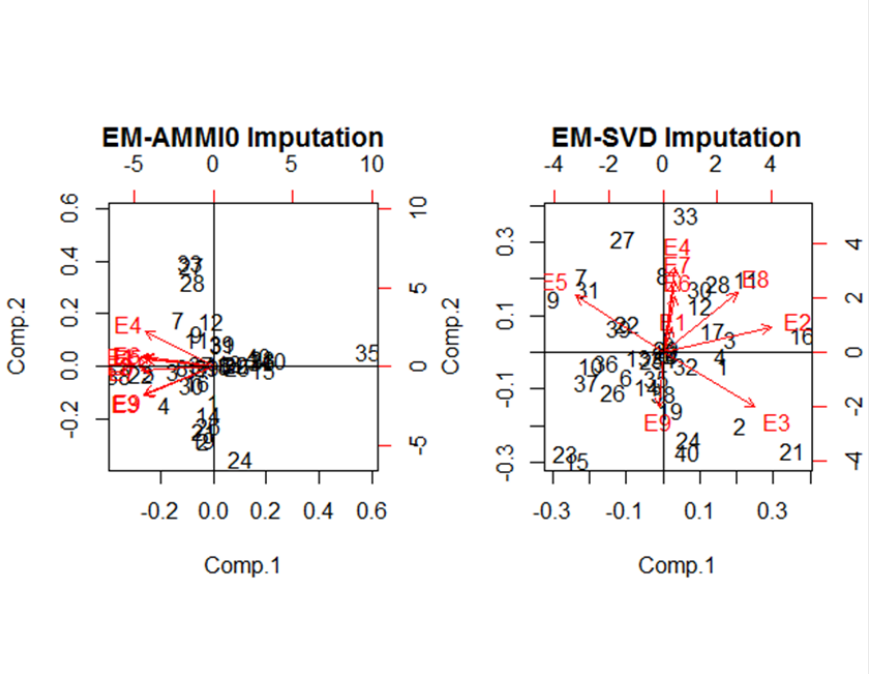

FIGURE 4.4: Biplots EM-AMMI and EM-SVD for data structures

4.8 Matrix Imputation

4.8.1 EM-AMMI imputation and EM-SVD Imputation

Expectation maximization AMMI (EM-AMMI) and SVD Imputation were used in imputing the GEI matrix. The imputation didn't converging beyond PC zero but still yielded the same imputation figures for EM-AMMI imputation (table 6.16 and 6.17 in appendices). Convergence at PC zero is automatic with the precision set at 0.009. EM-SVD imputation method imputed the missing values in table 4.4 attaining to convergence at the 583rd iteration (maximum iteration set at 1000) producing a complete data matrix in table 4.5 that was used in the AMMI analysis. that EM-SVD is the most powerful technique as it was able to impute, attain convergence and values detecting the GEI that existed.

Varieties	Chemelil	Muhoroni	Mumias PC	Mumias rc1	Nzoia	Sony 420pc	Sony 527BPc	Sony 527brc	West Kenya
166	NA	97.36	55.72	106.48	40.85	149.28	93.19	160.69	89.39
172	NA	NA	74.05	107.63	52.35	NA	NA	NA	96.47
270	NA	105.8333	NA	NA	NA	NA	NA	NA	NA
271	103.33	NA	NA	NA	NA	131.53	92.57	186.50	114.36
276	95.14	NA	NA	NA	NA	NA	NA	NA	NA
278	77.50	NA	NA	NA	NA	142.49	87.36	134.86	NA
279	NA	NA	50.48	131.65	85.34	NA	NA	NA	NA
282	NA	NA	NA	NA	NA	165.85	100.70	150.16	NA
30	NA	NA	49.85	121.46	85.92	159.67	85.42	148.32	94.56
300	NA	71.94	NA	NA	NA	NA	NA	NA	NA
302	89.72	101.11	61.07	125.10	53.43	147.24	94.93	166.17	78.57
303	86.18	92.08	61.45	116.37	NA	159.79	79.44	156.89	65.25
308	71.67	83.47	NA	NA	NA	NA	NA	NA	NA
313	83.96	NA	61.07	99.41	62.53	NA	NA	NA	92.82
326	NA	86.53	NA	NA	NA	122.72	76.71	125.79	NA
339	66.81	115.70	NA	NA	NA	144.03	91.32	160.58	NA
445	NA	NA	66.32	118.13	58.85	165.37	76.53	167.22	87.25
446	72.08	NA	NA	NA	NA	NA	NA	NA	NA
448	67.22	89.31	62.69	113.66	60.63	NA	NA	NA	114.57
508	65.83	NA	NA	NA	NA	NA	NA	NA	NA
526	68.68	94.17	83.15	121.04	43.17	NA	NA	NA	91.68
535	NA	NA	NA	NA	82.67	NA	NA	NA	NA
556	NA	75.42	NA	NA	NA	136.00	65.19	135.67	NA
569	53.47	NA	62.70	92.34	48.62	NA	NA	NA	90.36
573	61.25	87.91	NA	NA	NA	NA	NA	NA	NA
739	NA	78.75	64.56	101.28	78.03	NA	NA	NA	97.58
759	NA	NA	57.21	138.21	87.40	NA	NA	NA	67.64
77	NA	NA	57.21	152.67	53.90	NA	NA	NA	84.24
779	75.97	NA	NA	NA	NA	NA	NA	NA	NA
800	68.75	NA	NA	NA	NA	146.24	100.77	174.69	NA
801	NA	81.39	48.23	115.43	NA	146.20	94.31	153.98	78.20
830	NA	NA	NA	NA	50.70	NA	NA	NA	NA
866	NA	92.22	45.11	158.80	56.20	160.81	81.94	173.81	98.38
CB38-22	57.92	NA	NA	NA	NA	NA	NA	NA	NA
CO421	21.94	NA	NA	NA	NA	NA	NA	NA	NA
CO617	NA	75.69	NA	NA	NA	NA	NA	NA	NA
CO945	NA	NA	65.32	121.36	81.70	162.18	61.67	150.16	96.12
D8484	NA	NA	NA	NA	NA	NA	NA	NA	118.55
KEN83-737	NA	NA	48.23	119.38	NA	NA	NA	NA	NA
N14	NA	NA	NA	NA	29.39	152.03	66.74	134.51	NA

TABLE 4.4: Two way table of GEI means

Cultivars	Chemelil	Muhoroni	Mumias PC	Mumias RC1	Nzoia	Sony420 PC	Sony527 BPC	Sony527 RC	West Kenya
166	75.60	97.36	55.72	106.48	40.85	149.28	93.19	160.69	89.39
172	75.72	92.34	74.05	107.63	52.35	148.30	83.85	154.42	96.47
270	86.79	105.83	66.52	135.03	68.91	169.98	96.10	176.98	100.95
271	103.33	102.92	64.69	131.31	67.02	131.53	92.57	186.50	114.36
276	95.14	116.02	72.92	148.02	75.54	186.33	105.35	194.01	110.66
278	77.50	86.90	54.62	110.87	56.58	142.49	87.36	134.86	82.89
279	85.55	104.32	50.48	131.65	85.34	167.55	94.73	174.46	99.51
282	80.53	98.20	61.73	125.30	63.95	165.85	100.70	150.16	93.67
30	78.47	95.69	49.85	121.46	85.92	159.67	85.42	148.32	94.56
300	59.00	71.94	45.22	91.79	46.85	115.55	65.33	120.31	68.62
302	89.72	101.11	61.07	125.10	53.43	147.24	94.93	166.17	78.57
303	86.18	92.08	61.45	116.37	60.45	159.79	79.44	156.89	65.25
308	71.67	83.47	53.46	108.51	55.38	136.59	77.23	142.23	81.12
313	83.96	90.17	61.07	99.41	62.53	144.82	81.88	150.79	92.82
326	64.19	86.53	49.20	99.87	50.97	122.72	76.71	125.79	74.66
339	66.81	115.70	60.14	122.07	62.30	144.03	91.32	160.58	91.26
445	79.69	97.17	66.32	118.13	58.85	165.37	76.53	167.22	87.25
446	72.08	87.90	55.25	112.15	57.24	141.17	79.82	146.99	83.84
448	67.22	89.31	62.69	113.66	60.63	152.53	86.23	158.81	114.57
508	65.83	80.28	50.46	102.43	52.28	128.94	72.90	134.25	76.58
526	68.68	94.17	83.15	121.04	43.17	150.97	85.35	157.19	91.68
535	104.08	126.91	79.77	161.92	82.67	203.83	115.24	212.23	121.05
556	66.05	75.42	50.62	102.76	52.44	136.00	65.19	135.67	76.82
569	53.47	79.15	62.70	92.34	48.62	127.12	71.87	132.36	90.36
573	61.25	87.91	51.92	105.38	53.78	132.66	75.00	138.12	78.78
739	73.87	78.75	64.56	101.28	78.03	144.67	81.79	150.63	97.58
759	81.53	99.42	57.21	138.21	87.40	159.67	90.27	166.26	67.64
77	84.57	103.13	57.21	152.67	53.90	165.64	93.65	172.46	84.24
779	75.97	92.65	58.24	118.21	60.33	148.80	84.13	154.93	88.37
800	68.75	98.19	61.72	125.28	63.94	146.24	100.77	174.69	93.66
801	73.77	81.39	48.23	115.43	58.57	146.20	94.31	153.98	78.20
830	63.85	77.87	48.94	99.35	50.70	125.06	70.70	130.21	74.27
866	83.66	92.22	45.11	158.80	56.20	160.81	81.94	173.81	98.38
CB38-22	57.92	70.63	44.40	90.11	45.99	113.43	64.13	118.11	67.37
CO421	21.94	26.76	16.82	34.14	17.43	42.98	24.30	44.75	25.53
CO617	62.07	75.69	47.58	96.58	49.29	121.57	68.73	126.58	72.20
CO945	77.94	95.04	65.32	121.36	81.70	162.18	61.67	150.16	96.12
D8484	101.92	124.29	78.12	158.58	80.93	199.61	112.86	207.84	118.55
KEN83-737	74.04	90.28	48.23	119.38	58.79	145.00	81.98	150.97	86.11
N14	67.94	82.85	52.08	105.71	29.39	152.03	66.74	134.51	79.03

TABLE 4.5: EM-SVD imputed two-way table of GEI means

4.9 AMMI modeling results

AMMI is family chain of models ranging from the lowest, AMMI0 to the saturated model AMMI8. The number of AMMI models are based on rank k (k=min (n-1, p-1) of GEI matrix with p= 9 environments and n= 40 cultivars and controls. AMMI0 and AMMI1 models were significant at $\alpha = 0.05$ with p-values of 2.382E-04 and 7.34E-09 respectively. The biggest variation was accounted for by environment effect (72%), genotypes 6% and GEI 8% while residual accounted 13%. The rest (2%) were accounted for by the blocks nested within environments. That outcome was consistent with most studies findings in literature. The IPCA1 accounting for 77.11% of the GEI was sufficient in explaining the interaction effect under Gollob's method of assigning the degrees of freedom. AMMI3-AMMI8 models were part of the noise. The sum of squares 56,380 due to GEI (SSGEI) corresponds to the Eigen values. The presence of noise (inexplicable variation) inflated it and thus adjusted through SVD. The sum of squares due to genotype (SSG) and Environment (SSE) were 39,446, and 511,415 respectively. In the decomposition of GEI, only the first IPC was significant ($p < 0.05$) by Gollob (1968) F test, and explain 77.11% of the variation of the SSGEI which was a pattern response in SSGEI with 46 degrees of freedom (43.4% of the interaction degrees of freedom). Given that most of the IPC were not significant, GEI complexity was a simple one and explained by AMMI1 (more extreme interaction complexity would be explained by so many IPCs (table 4.6). Complexity of GEI could be affected by the type crop, diversification of gene pool form where cultivars are drawn and environmental conditions.

Sources	**D.f.**	**Sum Sq**	**Mean Sq**	**F**	$P-value\ (Pr(>F))$
Environment	8	511415	63927	205.6691	<2.2E-16***
Variety	39	39446	1011	3.254	6.176E-09***
Environment(block)	18	15300	850	2.7347	2.345E-04***
GxE Interaction	106	56380	532	1.7112	2.382E-04***
IPC1	46	43473.21	945.0698	3.040531	7.34E-09***
IPC2	44	13676.15	310.8216	0.999992	4.78E-01
IPC3	42	10004.7	238.2071	0.766373	8.51E-01
IPC4	40	8575.705	214.3926	0.689756	9.22E-01
IPC5	38	5152.592	135.5945	0.436242	9.99E-01
IPC6	36	4620.077	128.3355	0.412888	9.99E-01
IPC7	34	3572.909	105.0856	0.338087	1.00E+00
IPC8	32	2228.632	69.64476	0.224065	1.00E+00
Residuals (Noise)	206	-34923.6	-169.532	-0.54543	1.00E+00
Residuals	290	90139	311		

TABLE 4.6: AMMI ANOVA

4.9.1 Model diagnostics: Choice of the optimal AMMI model

The optimal multiplicative component determined by Gollob's (1968) technique based on the approximate F test at $\alpha = 0.05$ was IPCA1 given that it was significant $(p < 0.05)$ and accounted for 77.11% of the variation of the SSGEI which was a pattern response present in SSGEI with 46 degrees of freedom (43.4% of the interaction degrees of freedom) and above the threshold of 70%. The entire GEI corresponds to each chain of the AMMI model family that is AMMI0 model with 106 d.f. Removing 46 d.f. and sum of squares assigned to first axis (IPC1) has AMMI1 model sufficiently explaining the interaction. Removing 42 d.f. and the sum of squares of second axis (IPC2) leaves AMMI2 is the remainder and so on up to the AMMI8 model (table 4.6). The 43,473.21 SSIPCA1 is very close to the SSGEI making AMMI1 as the best model explaining interaction complexities. The first singular axis has biggest % of pattern which reduces gradually up to the last axis (IPCA). The subsequent axes corresponding to AMMI2-AMMI8 increases noise retention as most of the pattern of SSGEI are captured in IPCA1.

4.9.2 PC1 and the yield Biplot

Environment and varietal scores (table 6.5 and 6.6 in appendix) are useful in AMMI biplot analysis, establishing ASV and the Rank-Sum test.The biplot graphics analyzes the dispersion of genotypes, environments and interaction. AMMI1 biplot contains variations of principal additive effects of genotypes and environments in horizontal axis (x-axis) and the variation of multiplicative effects of GEI on vertical axis (y-axis). Figure 4.5 of IPCA1 vs. means yields.

Genotypes or environments whose values are closer to the origin of axis (IPCA1) provide a smaller contribution to the GEI than those that are further away. D8484, Ms866, Ms77, Ms535, Ms270, Ms276, Ms325, CB-38-22 contribute more to the interaction and are least stable. Muhoroni contribution to GEI was small with an intermediate contribution by Chemelil, West Kenya and Sony527Bpc and a high contribution by Mumias (both PC and RC), Nzoia, Sony420pc, and Sony527BRC. Environments of Sony527Brc, Sony420pc and Mumiasrc averages recorded above the overall averages (97tch), indicating that they were favourable harvesting cycle for obtain high means (figure 4.5).

Genotypes Ms282, Ms271 and Ms339 display a productivity above the general mean and are more stable as they appear exactly on the x-axis indicating that they are associated with better adaptability and stability. However, not all genotypes with high mean productivity were stable as indicated in fig 4.5.

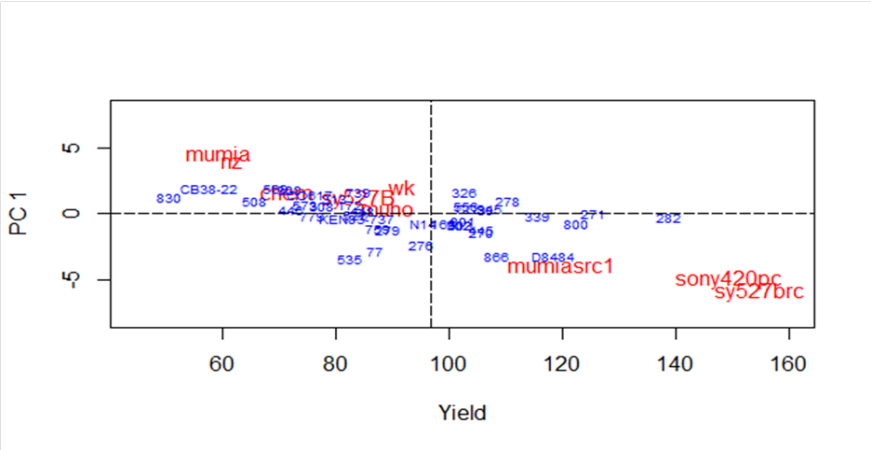

FIGURE 4.5: Yield vs PC1 biplot

4.9.3 PC1 and PC2 Biplot

AMMI2 biplot visualizes the multiplicative effects of the GEI contained in the first two IPCs (PC1 and PC2). The first singular axis captures the highest percentage of patterns of GEI which was (77%) of SSGEI. The second axis exhausted patterns in GEI (100%) and surpassed by including noise. However the scores of genotypes and environments were plotted up to the second axis. Most genotypes were stable but not productive under prevailing environments (harvest) as they appear close to the origin and some being below the mean productivity. Genotype closer to a given environment, the well adapted to them. Most of the cultivars were concentrated around the origin and closer to Muhoroni, Chemelil and Sony527BPC indicating more adaptability to those environments (figure 4.5). Small angles among the genotypes and environments vectors within the same quadrants show similarities amongst them genotypes and environments while vectors in the opposite quadrants show differences in genetic make up among the corresponding cultivars (Figure 4.6).

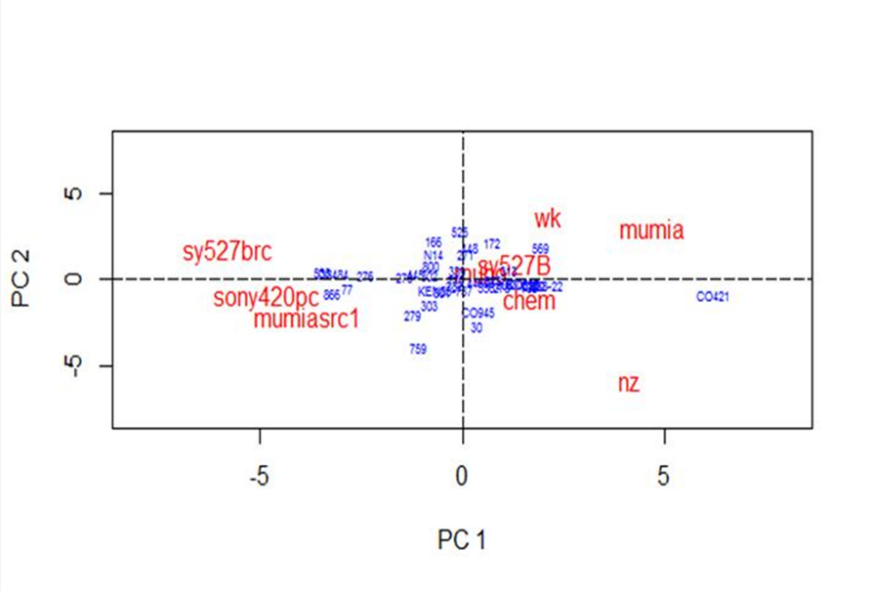

FIGURE 4.6: PC1 vs PC2

4.10 Stability and performance measure using the AMMI stability value (ASV), Yield Stability Index (YSi) and the Non parametric Rank Sum test (RS)

The cultivars showed significant differences in yield performance(tch). Ms282, Ms339, Ms271, CO945(control), Ms565, Ms801, Ms800, Ms302, Ms303, N14(control), Ms30, Ms278, Ms166, Ms445, Ms270, D8484(control) and Ms866 gave the best mean yields well above the overall mean performance of 97 tch as compared to the rest across the environments.

The IPCA scores of cultivars in AMMI model are indicators of the stability of a genotype over environment Purchase, Hatting, and Van Deventer, 2000. The lowest IPCA1 scores were observed for cultivar Ms271 followed by Ms526 and Ms779. Therefore Ms271 was the most stable cultivar with the mean yield of 126tch higher than grand mean of 97tch. The least stable were CO421, Ms535, Ms866 and D8484.

AMMI stability value (ASV) using the scores of the IPCA1 and IPCA2 and their sum of squares selected cultivars Ms779, Ms339 Ms282, Ms446, KEN83-737 and Ms271 as the most stable given the relatively low values and the least stable as CO421, Ms535, Ms866 and D8484 with large values of the ASV. IPCA1 scores and ASV agreed on the Ms271, Ms779 as among the most stable and CO421, Ms535, Ms866 and D8484 as the least stable but slightly differed on some.
Rank-sum (RS) introduced cultivar Ms282 (RS=3.41), Ms339 (RS=5.62) and Ms271 (RS=6.83) as the most stable and high yielding given the lowest values and Ms866, D8484, Ms830 and Ms535 as the least stable as they had higher values. The slight difference is attributed to Rank sum test factoring in the yield performance (table 4.7).

Varieties	PC1score	PC2score	Genotype Yield (tch)	ASVi	Rank ASV	Rank yld	YSIi	RS
282	-0.2071	-0.2502	138.9011	0.7042	3	1	4	3.41
339	-0.1184	0.5655	115.6860	0.6793	2	5	7	5.62
271	0.0653	1.5360	125.6580	1.5500	6	2	8	6.83
CO945	0.3860	-1.8344	105.5000	2.2069	11	11	22	11.00
556	0.5979	-0.3759	103.0692	1.9375	9	12	21	12.62
801	-0.5067	-0.6652	102.5324	1.7428	7	14	21	15.45
800	-0.7646	0.8475	122.6108	2.5739	14	3	17	16.28
302	-0.8044	0.2680	101.9285	2.5710	13	16	29	16.62
303	-0.8172	-1.4089	102.1817	2.9553	18	15	33	18.62
N14	-0.7076	1.5012	95.6667	2.7043	15	18	33	18.62
30	0.3508	-2.6704	106.4581	2.8939	17	8	25	18.86
278	0.9711	-0.3790	110.5533	3.1102	19	6	25	21.69
166	-0.7117	2.2955	99.1200	3.2229	21	17	38	21.83
448	0.1986	1.9474	84.6817	2.0472	10	23	33	25.69
445	-1.1710	0.3954	105.6667	3.7433	24	10	34	26.90
279	-1.2222	-1.9819	89.1578	4.3614	26	20	46	27.24
526	-0.0692	2.8586	83.6483	2.8670	16	26	42	28.07
KEN83-737	-0.3878	-0.5516	83.8067	1.3504	5	25	30	29.14
172	0.7148	2.1611	82.6225	3.1358	20	28	48	29.66
313	1.1502	0.5373	79.9580	3.6954	23	29	52	30.24
759	-1.0730	-3.9534	87.6150	5.2215	29	21	50	30.66
270	-1.4194	0.1709	105.8333	4.5150	27	9	36	30.73
739	1.6741	-0.3606	84.0400	5.3337	30	24	54	31.24
CO617	1.4856	-0.2320	75.6933	4.7282	28	32	60	32.83
308	0.6046	-0.1485	77.5700	1.9276	8	30	38	34.56
326	1.6963	-0.1569	102.9375	5.3944	31	13	44	34.73
300	1.8471	-0.2822	71.9433	5.8783	32	35	67	35.62
569	1.9401	1.9177	69.4960	6.4583	34	36	70	36.41
779	-0.1485	-0.0053	75.9733	0.4721	1	31	32	37.21
573	0.7505	0.0006	74.5817	2.3857	12	33	45	37.35
276	-2.4009	0.3071	95.1367	7.6379	35	19	54	38.31
77	-2.8338	-0.5050	87.0025	9.0222	36	22	58	38.90
CB38-22	1.9738	-0.2996	57.9167	6.2815	33	38	71	39.04
CO421	6.2020	-0.8860	21.9433	19.7347	40	40	80	40.00
508	1.0434	-0.1706	65.8333	3.3210	22	37	59	40.11
446	0.3091	-0.0688	72.0800	0.9849	4	34	38	40.21
535	-3.4501	0.4503	82.6667	10.9761	39	27	66	41.49
830	1.2764	-0.2031	50.7033	4.0624	25	39	64	41.90
D8484	-3.1980	0.4176	118.5500	10.1742	37	4	41	43.83
866	-3.2263	-0.7882	108.4096	10.2858	38	7	45	44.42

TABLE 4.7: ASV, YSI and RS

Chapter 5

Findings and Conclusion

5.1 Findings

EM-AMMI0 was efficient in all aspects of execution of the codes by utilizing lesser user and system times as compared to EM-SVD. It attained convergence much faster than EM-SVD with a difference of 115 iterations. The imputed values in both techniques were strongly positive correlated with correlation coefficient of 0.937 and were equally not significantly different from one another as indicated by paired t-test $(p > 0.05)$.

EM-AMMI0 had the least predictive residual sum of squares (PRESS) value of 55.18 compared to EM-SVD's 118.86. Given the prevailing GEI matrix with the missing values, EM-AMMI0 was comparatively better imputation technique.

The two techniques produced completely different data structures. The environments under EM-AMMI0 were strongly correlated unlike EM-SVD that had diverse environments. The first two PC under EM-AMMI0 accounted for 85.2% of the total variation in the data while EM-SVD had the first five PC's accounted for 76.8% and an inconclusive scree plot.

In the performance of 40 genotypes in nine environments,environments effect accounted for 71.76% of the total sum of squares, 5.53% to genotypic effects, and 7.91% to GEI effects which conformed to most of the findings in literature. Large environmental sum of squares indicated diversity hence large differences in their means performance caused most of the variation in cultivars performance.

The AMMI model that sufficiently explained the main effect and GEI was AMMI1, thus the GEI complexity was simple. The low order AMMI (AMMI1) for the GEI defines a

small number of mega-environments hence the biplot analysis showing four mega environments which is a good for the seed producers.

The lowest IPCA1 score as an indicator of stability was observed for cultivars Ms271, Ms526 and Ms779 in that order. relying on IPCA alone as IPCA2 had some noise would imply that Ms271 was the most stable cultivar with the mean yield of 126tch higher than grand mean of 95tch. The least stable would therefore be CO421, Ms535, Ms866 and D8484.

The most stable cultivars by ASV were Ms779, Ms339 and Ms282 and Ms446 while the least stable included CO421, Ms535, Ms866 and D8484. AMMI stability value index determines stability but doesn't incorporate yield performance. Rank-sum (RS) introduced cultivar Ms282 (RS=3.41), Ms339 (RS=5.62) and Ms271 (RS=6.83) as the most stable and Ms866, D8484, Ms830 and Ms535 as the least stable.

5.2 Conclusion

EM-AMMI was a better imputation technique than the EM-SVD. However the final data structure would determine the technique as both produces different structure with different levels of correlation among environments.

Imputation of GEI is a challenging task and given that every techniques uses different ranks and models and are bound to change with the change in the GEI matrix tests on error minimization need to be conducted prior to decision the technique to used.

The choice of AMMI1 as the optimal model indicted a simple complexity of GEI with four mega-environments (harvest) delineated Environmental effect was the most predominant source of variation followed by GEI and genotype effect. GEI effect was five times higher than genotypic effect and influenced the difference among genotype.

According to Crossa, Gauch, and Zobel (1990), lower order AMMI models are indications of weaker germplasm. There would be need of thorough evaluation of the gene pool to ascertain the finding

Most cultivars were stable appearing closer to the origin (fig 4.5 and 4.6) due to lower variance. Different stability indices rated cultivars stability slightly differently in terms of

order but agreed on most of the tops and the bottoms in general.

5.3 Further areas of research

Bayesian evaluation of GEI as prior cultivars parentage performance distributions and experimental data provides posterior distribution for prediction of cultivars performance, stability and adaptability.

Review of all GEI imputation techniques for efficiencies and biases.

Bibliography

Akter, Anowara et al. (2015). "GGE Biplot Analysis for Yield Stability in Multi-environment Trials of Promising Hybrid Rice (Oryza sativa L.)" In: *Bangladesh Rice Journal* 19.1, pp. 1–8.

Alter, Orly, Patrick O Brown, and David Botstein (2000). "Singular value decomposition for genome-wide expression data processing and modeling". In: *Proceedings of the National Academy of Sciences* 97.18, pp. 10101–10106.

Amira, JO et al. (2013). "Relative discriminating powers of GGE and AMMI models in the selection of tropical soybean genotypes". In: *African Crop Science Journal* 21.1, pp. 67–73.

Annicchiarico, P (1997). "Additive main effects and multiplicative interaction (AMMI) analysis of genotype-location interaction in variety trials repeated over years". In: *Theoretical and Applied Genetics* 94.8, pp. 1072–1077.

Arciniegas-Alarcón, Sergio et al. (2010). "An alternative methodology for imputing missing data in trials with genotype-by-environment interaction". In: *Biometrical Letters* 47.1, pp. 1–14.

Arciniegas-Alarcón, Sergio et al. (2013). "Deterministic imputation in multienvironment trials". In: *ISRN Agronomy* 2013.

Arciniegas-Alarcón, Sergio et al. (2014). "Imputing missing values in multi-environment trials using the singular value decomposition: An empirical comparison". In: *Communications in Biometry and Crop Science* 9.2, pp. 54–70.

Baker, Kirk (2005). "Singular value decomposition tutorial". In: *The Ohio State University* 24.

Baye, Tesfaye M, Tilahun Abebe, and Russell A Wilke (2011). "Genotype–environment interactions and their translational implications". In: *Personalized medicine* 8.1, pp. 59–70.

Bergamo, Genevile Carife, Carlos Tadeu dos Santos Dias, and Wojtek Janusz Krzanowski (2008). "Distribution-free multiple imputation in an interaction matrix through singular value decomposition". In: *Scientia Agricola* 65.4, pp. 422–427.

Crossa, J, HG Gauch, and R W_ Zobel (1990). "Additive main effects and multiplicative interaction analysis of two international maize cultivar trials". In: *Crop Science* 30.3, pp. 493–500.

Dias, Carlos Tadeu dos Santos and Wojtek Janusz Krzanowski (2006). "Choosing components in the additive main effect and multiplicative interaction (AMMI) models". In: *Scientia Agricola* 63.2, pp. 169–175.

Falconer, Douglas S and Trudy FC Mackay (1996). "Introduction to quantitative genetics". In: *Harlow: Longman Group Ltd*.

Forkman, Johannes and Hans-Peter Piepho (2014). "Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models". In: *Biometrics* 70.3, pp. 639–647.

Gauch, Hilbert Gilbert (1992). "Statistical analysis of regional yield trials: AMMI analysis of factorial designs". In: *Academic Journals*.

Gauch, Hugh G (2013). "A simple protocol for AMMI analysis of yield trials". In: *Crop Science* 53.5, pp. 1860–1869.

Gauch, Hugh G, Hans-Peter Piepho, and Paolo Annicchiarico (2008). "Statistical analysis of yield trials by AMMI and GGE: Further considerations". In: *Crop Science* 48.3, pp. 866–889.

GoK (2015). *Statistical Abstract 2015*. Kenya National Bureau of Statistics (KNBS).

Hongyu, Kuang et al. (2014). "Statistical analysis of yield trials by AMMI analysis of genotype× environment interaction". In: *Biometrical Letters* 51.2, pp. 89–102.

Hourani, Mou'ath and Ibrahiem MM El Emary (2009). "Microarray missing values imputation methods: Critical analysis review". In: *Comput Sci Inf Syst ComSIS* 6.2, pp. 165–190.

Josse, Julie et al. (2014). "Another look at Bayesian analysis of AMMI models for genotype-environment data". In: *Journal of Agricultural, Biological, and Environmental Statistics* 19.2, pp. 240–257.

Kahram, Azarakhsh et al. (2013). "Genotype× environment interaction analysis for grain yield of durum wheat new genotypes in the moderate region of Iran using AMMI model". In: *World Journal of Agricultural Sciences* 9.3, pp. 298–304.

Karimizadeh, Rahmatollah et al. (2012). "Univariate stability analysis methods for determining genotype× environment interaction of durum wheat grain yield". In: *African Journal of Biotechnology* 11.10, pp. 2563–2573.

Kohler, Ulrich, Magdalena Luniak, et al. (2005). "Data inspection using biplots". In: *Stata Journal* 5.2, p. 208.

Lavoranti, Osmir José, Carlos Tadeu dos Santos Dias, and Wojtek J Kraznowski (2010). "Phenotypic stability via ammi model with bootstrap re-sampling Estabilidade fenotípica através da reamostragem "bootstrap" no modelo AMMI". In: *Pesquisa Florestal Brasileira*.

M., Mustapha and Bakari H. R. (2014). "Statistical evaluation of genotype by environment interactions for grain yield in Millet (penniisetum glaucum (L) R. Br)". In: *The International Journal of Engineering and Science (IJES)* 3.9, pp. 7–16.

Miranda, Glauco Vieira et al. (2009). "Multivariate analyses of genotype x environment interaction of popcorn". In: *Pesquisa Agropecuária Brasileira* 44.1, pp. 45–50.

Mohammadi, Reza and Ahmed Amri (2012). "Analysis of genotype× environment interaction in rain-fed durum wheat of Iran using GGE-biplot and non-parametric methods". In: *Canadian Journal of Plant Science* 92.4, pp. 757–770.

Paderewski, Jakub (2013). "An R function for imputation of missing cells in two-way data sets by EM-AMMI algorithm". In: *Communications in Biometry and Crop Science8* 2, pp. 60–69.

Paderewski, Jakub and Paulo Canas Rodrigues (2014). "The usefulness of EM-AMMI to study the influence of missing data pattern and application to Polish post-registration winter wheat data". In: *Australian Journal of Crop Science* 8.4, p. 640.

Parmar, DJ et al. (2012). "Non-Parametric Methods for Interpreting Genotype× Environment Interaction of Rice Genotypes (Oryza sativa L.)" In: *J. Ric. Res* 5, pp. 33–39.

Pereira, Dulce Gamito Santinhos et al. (2012). "Analyzing genotype-by-environment interaction using curvilinear regression". In: *Scientia Agricola* 69.6, pp. 357–363.

Purchase, JL, Hesta Hatting, and CS Van Deventer (2000). "Genotype× environment interaction of winter wheat (Triticum aestivum L.) in South Africa: II. Stability analysis of yield performance". In: *South African Journal of Plant and Soil* 17.3, pp. 101–107.

Rad, MR Naroui et al. (2013). "Genotype environment interaction by AMMI and GGE biplot analysis in three consecutive generations of wheat (Triticum aestivum) under normal and drought stress conditions". In: *Australian Journal of Crop Science* 7.7, p. 956.

Roostaei, Mozaffar, Reza Mohammadi, and Ahmed Amri (2014). "Rank correlation among different statistical models in ranking of winter wheat genotypes". In: *The crop journal* 2.2, pp. 154–163.

Silveira, Luís Cláudio Inácio da et al. (2013). "AMMI analysis to evaluate the adaptability and phenotypic stability of sugarcane genotypes". In: *Scientia Agricola* 70.1, pp. 27–32.

Tadege, Muluken Bayable, Habtamu Zelleke Utta, and Amsalu Ayana Aga (2014). "Association of statistical methods used to explore genotype environment interaction (GEI) and cultivar stability". In: *African Journal of Agricultural Research* 9.29, pp. 2231–2237.

Troyanskaya, Olga et al. (2001). "Missing value estimation methods for DNA microarrays". In: *Bioinformatics* 17.6, pp. 520–525.

Yan, Weikai et al. (2007). "GGE biplot vs. AMMI analysis of genotype-by-environment data". In: *Crop science* 47.2, pp. 643–653.

Yoon, Dankyu, Eun-Kyung Lee, and Taesung Park (2007). "Robust imputation method for missing values in microarray data". In: *BMC bioinformatics* 8.Suppl 2, S6.

Chapter 6

Appendices

6.1 Sum of squares for the multiplicative and error component

Let Z be the residual having multiplicative terms and error. SVD of $Z = \hat{U}\hat{\Lambda}\hat{V}'$ $\sum\sum e^2 = \sum\sum(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = trace(ZZ') = trace(Z'Z)$. For any square matrices A and B, $trace(AB) = trace(BA)$.
Since $Z = \hat{U}\hat{\Lambda}\hat{V}'$, $ZZ' = [\hat{U}\hat{\Lambda}\hat{V}'][\hat{U}\hat{\Lambda}\hat{V}'] = [\hat{U}\hat{U}'\hat{\Lambda}\hat{\Lambda}'\hat{V}\hat{V}']$,but $\hat{U}\hat{U}' = I$ and $\hat{V}\hat{V}' = I$ are identity matrices and therefore $ZZ' = [I_n'\hat{\Lambda}'\hat{\Lambda}_n I_n]$ = trace$(\hat{\Lambda}'\hat{\Lambda}_n)$.
Hence trace (ZZ') =trace $(Z'Z)$ = ZZ' = trace $\hat{\Lambda_n'}\hat{\Lambda_n}$.
$\hat{\Lambda_n}$- is a diagonal matrix, then trace $\hat{\Lambda_n'}\hat{\Lambda_n}$ = trace$\hat{\Lambda_n}^2 = [\lambda_1^2 > \lambda_2^2 > \lambda_3^2 > ... > \lambda_n^2]$ = $\sum_{k=1}^{n} \lambda_k^2$.
Therefore the sum of squares for the interaction component (GE) =trace ($Z'Z = \sum_{k=1}^{n} \lambda_k^2$)

6.2 AMMI ANOVA Table

TABLE 6.1: AMMI ANOVA Table

	Gollob's			**Cornelius**		
Source	**d.f.**	**SS**	**MS**	**d.f.**	**SS**	**MS**
Environment (E)	e-1	SSE	MSE	-	-	-
Rep(Environment) R(E)	e(r-1)	SSR(E)	MSR(E)	-	-	-
Genotype (G)	g-1	SSG	MSG	-	-	-
GxE interaction	(e-1)(g-1)	SSGxE	MSGxE	-	-	-
IPCA1	V1=g+e-1-(2*1)	$r\hat{\lambda_1}^2$	MSIPCA1	(g-1-1)*(e-1-1)	$\sum_{k=2}^{f} \lambda_k^2$	SSIPCA1/d.f.
IPCA2	V2=g+e-1-(2*2)	$r\hat{\lambda_2}^2$	MSIPCA2	(g-1-2)*(e-1-2)	$\sum_{k=3}^{f} \lambda_k^2$	SSIPCA2/d.f.
IPCA3	V3=g+e+1-(2*3)	$r\hat{\lambda_1}^2$	MSIPCA3	(g-1-3)*(e-1-3)	$\sum_{k=4}^{f} \lambda_k^2$	SSIPCA3/d.f.
⋮	⋮	⋮	⋮	⋮	⋮	⋮
IPCAF	Vf=g+e-1-(2*f)	$r\hat{\lambda_F}^2$	MSIPCAF	(g-1-f)*(e-1-f)	$\sum_{k=F+1}^{f} \lambda_k^2$	SSIPCAF/d.f.
Residual	(g-1)(e-1)-sum (vk)	-SS_{G-residual}	MSR	-	-SS_{C-residual}	MSR
Experimental error	e(g-1)(r-1)	SS error mean	MSE	-	-	-
Total	ger-1	-	-	-	-	-

6.3 EM-AMMI imputed GEI matrix

Cultivars	Chemelil	Muhoroni	Mumias PC	Mumias RC1	NzoiaPC	Sony 420PC	Sony527 BPC	Sony527 BRC	West Kenya
166	73.02	97.36	55.72	106.48	40.85	149.28	93.19	160.69	89.39
172	76.48	95.01	74.05	107.63	52.35	151.61	86.44	157.14	96.47
270	87.31	105.83	70.41	130.74	72.12	162.43	97.27	167.97	100.54
271	103.33	108.39	72.96	133.29	74.67	131.53	92.57	186.50	114.36
276	95.14	113.59	78.16	138.49	79.87	170.19	105.02	175.72	108.29
278	77.50	87.65	52.21	112.55	53.92	142.49	87.36	134.86	82.34
279	85.38	103.91	50.48	131.65	85.34	160.51	95.34	166.04	98.61
282	83.66	102.19	66.75	127.09	68.46	165.85	100.70	150.16	96.88
30	79.27	97.80	49.85	121.46	85.92	159.67	85.42	148.32	94.56
300	53.42	71.94	36.52	96.85	38.23	128.54	63.38	134.08	66.65
302	89.72	101.11	61.07	125.10	53.43	147.24	94.93	166.17	78.57
303	86.18	92.08	61.45	116.37	58.98	159.79	79.44	156.89	65.25
308	71.67	83.47	51.39	111.72	53.10	143.42	78.25	148.95	81.52
313	83.96	93.58	61.07	99.41	62.53	150.17	85.00	155.70	92.82
326	56.87	86.53	39.97	100.30	41.68	122.72	76.71	125.79	70.10
339	66.81	115.70	61.93	122.26	63.64	144.03	91.32	160.58	92.06
445	78.48	97.01	66.32	118.13	58.85	165.37	76.53	167.22	87.25
446	72.08	90.54	55.10	115.43	56.81	147.13	81.96	152.66	85.23
448	67.22	89.31	62.69	113.66	60.63	152.62	87.46	158.16	114.57
508	65.83	84.29	48.86	109.19	50.56	140.88	75.72	146.42	78.98
526	68.68	94.17	83.15	121.04	43.17	151.59	86.43	157.12	91.68
535	97.85	116.38	80.94	141.28	82.67	172.97	107.81	178.50	111.07
556	57.00	75.42	40.10	100.43	41.81	136.00	65.19	135.67	70.23
569	53.47	83.11	62.70	92.34	48.62	139.71	74.54	145.24	90.36
573	61.25	87.91	48.40	108.74	50.11	140.43	75.27	145.96	78.53
739	75.42	78.75	64.56	101.28	78.03	150.55	85.38	156.08	97.58
759	81.48	100.01	57.21	138.21	87.40	156.60	91.44	162.13	67.64
77	80.86	99.39	57.21	152.67	53.90	155.99	90.82	161.52	84.24
779	75.97	94.43	59.00	119.33	60.70	151.02	85.86	156.56	89.12
800	68.75	99.70	64.27	124.60	65.98	146.24	100.77	174.69	94.40
801	70.53	81.39	48.23	115.43	55.34	146.20	94.31	153.98	78.20
830	65.88	84.41	48.98	109.31	50.70	141.01	75.84	146.54	79.11
866	82.31	92.22	45.11	158.80	56.20	160.81	81.94	173.81	98.38
CB38-22	57.92	76.37	40.94	101.27	42.65	132.97	67.80	138.50	71.07
CO421	21.94	40.40	4.97	65.30	6.67	96.99	31.83	102.53	35.09
CO617	57.17	75.69	40.27	100.60	41.98	132.29	67.13	137.83	70.40
CO945	78.32	96.84	65.32	121.36	81.70	162.18	61.67	150.16	96.12
D8484	105.33	123.86	88.43	148.76	90.13	180.45	115.29	185.99	118.55
KEN83-737	70.55	89.08	48.23	119.38	55.35	145.67	80.51	151.20	83.77
N14	58.03	76.56	41.13	101.46	29.39	152.03	66.74	134.51	71.26

TABLE 6.2: EM-AMMI imputed GEI matrix

6.4 Cultivars and controls mean performance

TABLE 6.3: Cultivars and controls mean performance

Cultivars and Controls	Yield (tch)	Cultivars and Controls	Yield (tch)
166	99.12	526	83.65
172	82.62	535	82.67
270	105.83	556	103.07
271	125.66	569	69.50
276	95.14	573	74.58
278	110.55	739	84.04
279	89.16	759	87.62
282	138.90	77	87.00
30	106.46	779	75.97
300	71.94	800	122.61
302	101.93	801	102.53
303	102.18	830	50.70
308	77.57	866	108.41
313	79.96	CB38-22	57.92
326	102.94	CO421	21.94
339	115.69	CO617	75.69
445	105.67	CO945	105.50
446	72.08	D8484	118.55
448	84.68	KEN83-737	83.81
508	65.83	N14	95.67

6.5 Environments mean performance

TABLE 6.4: Environments mean performance

Environment	**Yield (tch)**
Chemelil	71.52
MuhoroniPC	89.31
MumiasPC	59.69
MumiasRC1	120.02
NzoiaPC	61.76
Sony420PC	149.46
Sony527BPC	84.30
Sony527BRC	155.00
West KenyaPC	92.00

6.6 Test for Normality assumptions by plots

Individual environment yield Normality test using QQplot and fitting of the residual vs. fitted values for normality assumption

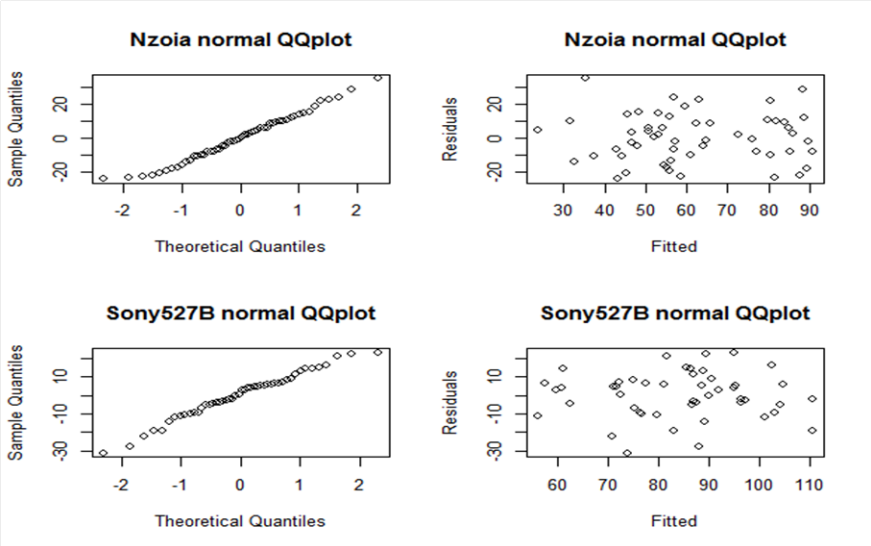

FIGURE 6.1: Normality tests for Nzoia and Sony524B residuals

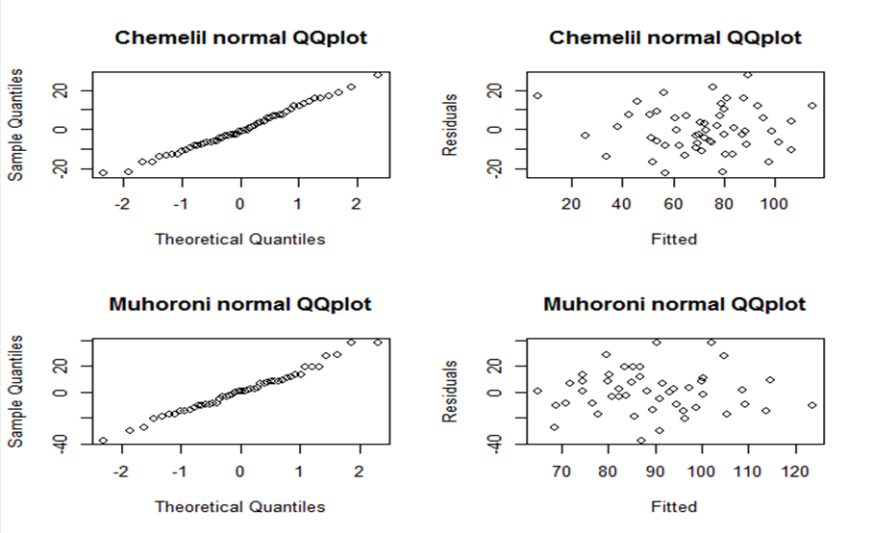

FIGURE 6.2: Normality tests for Chemelil and Muhoroni residuals

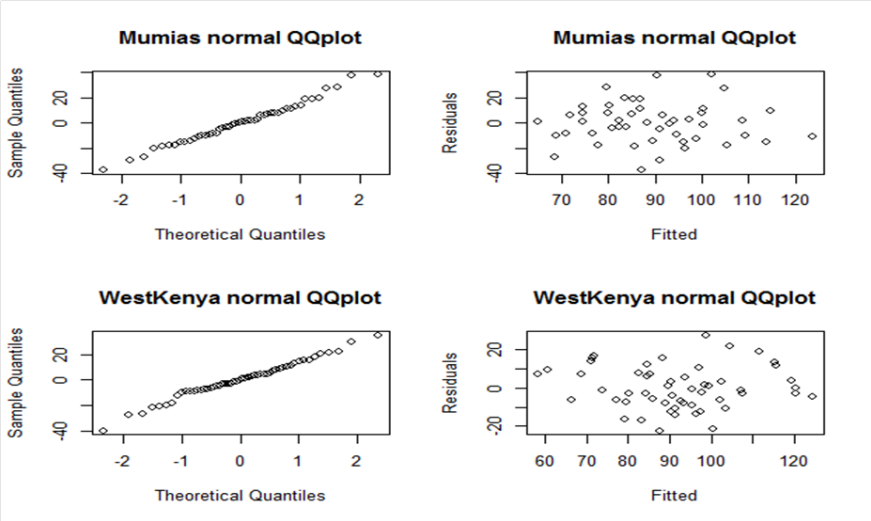

FIGURE 6.3: Normality tests for MumiasPC and West Kenya residuals

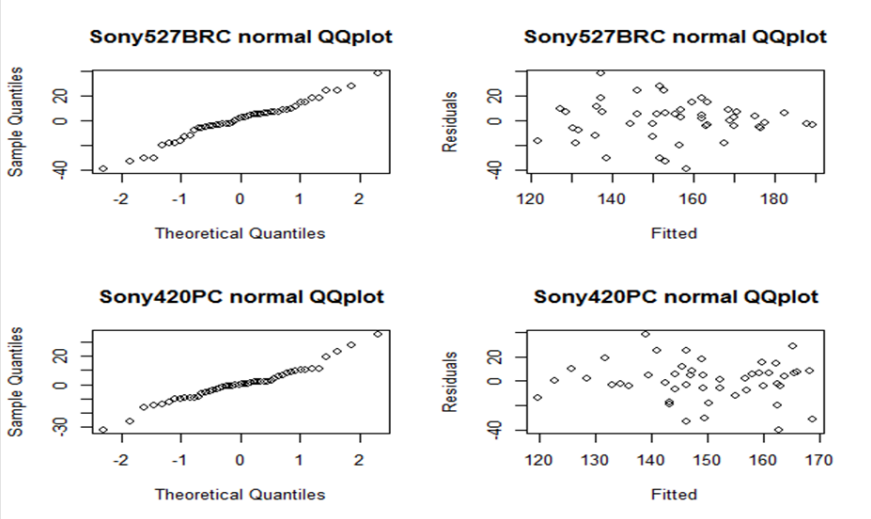

FIGURE 6.4: Normality tests for Sony527BRC and Sony420PC residuals

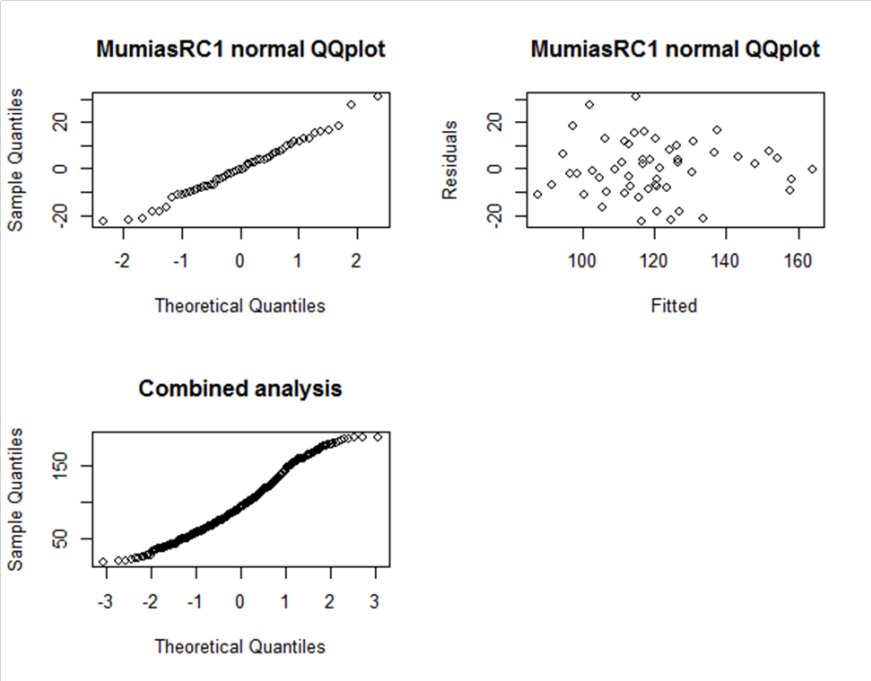

FIGURE 6.5: Normality tests for MumiasRC1 and All environments residuals

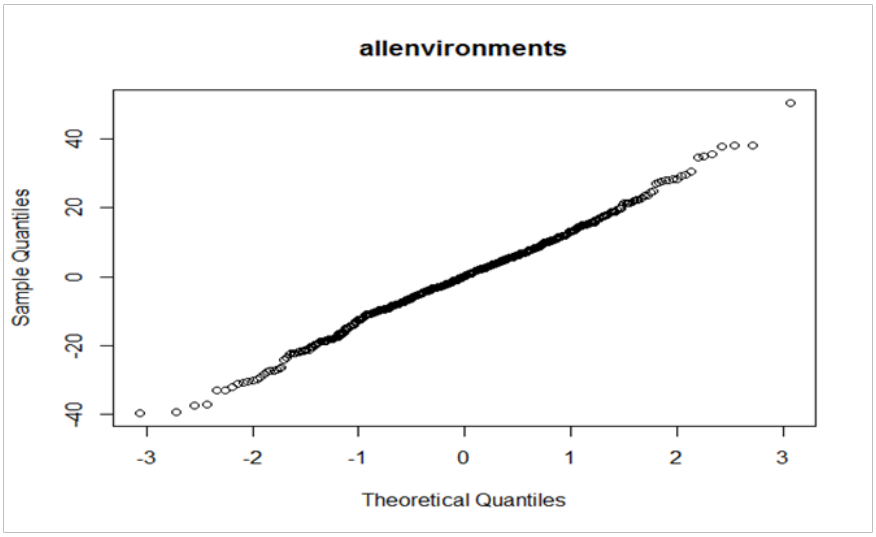

FIGURE 6.6: Normality tests for all environments by qqnormplot

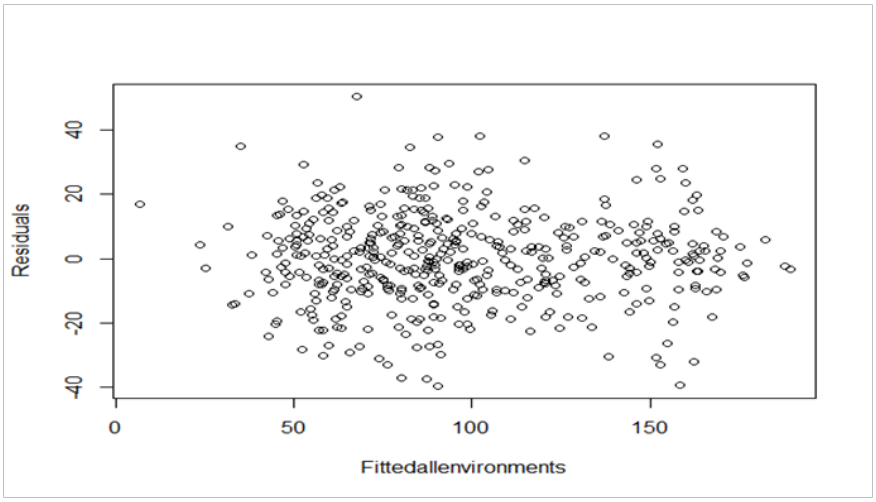

FIGURE 6.7: Normality tests for all environments residuals

6.7 Environmental Score under saturated model

TABLE 6.5: Environmental Score

	Environment scores							
Environments	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**
Chemelil	1.6740	-1.0768	2.0373	0.8175	-2.6511	-3.2088	3.2595	0.8130
Muhoroni	0.4615	0.4677	1.3205	1.3769	0.9632	0.9080	-1.8033	4.3070
Mumiasrc1	4.7285	3.0327	-1.0347	3.1211	-1.2228	-0.9241	-2.5292	-1.8309
Mumiasrc1	-3.8194	-2.1283	2.4788	0.0708	-3.2586	3.1727	-0.9920	-1.2007
Nzoia	4.1377	-5.8167	-1.6762	-2.3227	1.2756	-0.5213	-1.2141	-0.4402
Sony420pc	-4.8341	-0.9016	-4.9314	3.0409	0.8240	0.3088	1.5524	0.0186
Sony527BPC	1.3126	0.9392	3.2635	0.8958	4.0451	1.5599	2.0059	-1.5544
Sony527BRC	-5.7993	1.7805	0.9737	-2.7229	1.1751	-3.2518	-1.7685	-0.6192
West Kenya	2.1385	3.7033	-2.4316	-4.2774	-1.1505	1.9566	1.4893	0.5068

6.8 Varietal Score under saturated AMMI model

	Principal Components							
Cultivars	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
166	-0.7117	2.2955	0.7244	0.5752	1.5690	-0.5631	1.2154	1.1051
172	0.7148	2.1611	-0.8476	0.5617	-0.1391	-0.6882	-0.2515	-0.3456
270	-1.4194	0.1709	-0.2848	-0.1141	0.1861	-0.0969	-0.0062	0.0343
271	0.0653	1.5360	2.8401	-3.7689	-1.8630	-2.3122	0.1727	0.5929
276	-2.4009	0.3071	-0.5028	-0.1806	0.3235	-0.1628	-0.0061	0.0607
278	0.9711	-0.3790	0.3924	0.9457	0.1192	0.7418	1.7417	-0.0713
279	-1.2222	-1.9819	-0.5496	-1.7440	1.2613	-0.0801	0.5292	0.8272
282	-0.2071	-0.2502	-0.4014	1.3075	0.9714	1.6402	1.7499	-0.3177
30	0.3508	-2.6704	-1.5255	-1.0277	0.6825	0.9696	0.8608	0.7601
300	1.8471	-0.2822	0.4406	0.1071	-0.2698	0.1232	-0.0076	-0.0537
302	-0.8044	0.2680	2.4672	1.1787	0.1031	-1.3603	0.2446	0.3098
303	-0.8172	-1.4089	0.0070	2.5899	-0.1446	-2.4991	-0.0761	-0.1219
308	0.6046	-0.1485	0.1918	0.0103	-0.2530	-0.1555	0.2567	-0.2114
313	1.1502	0.5373	-0.7437	-0.3334	0.3631	-1.8928	1.2262	0.7512
326	1.6963	-0.1569	0.9816	0.4627	0.3524	0.8970	0.0161	1.0598
339	-0.1184	0.5655	1.0628	-0.2134	1.5190	1.4379	-2.3581	2.5509
445	-1.1710	0.3954	-1.4638	0.9606	-0.4338	-1.6329	-0.8056	0.4542
446	0.3091	-0.0688	0.0989	0.0029	-0.0548	0.0198	-0.0073	-0.0123
448	0.1986	1.9474	-1.9088	-2.0536	0.3890	1.2484	0.3563	-0.6116
508	1.0434	-0.1706	0.2620	0.0526	-0.1573	0.0693	-0.0076	-0.0321
526	-0.0692	2.8586	-0.2693	1.8668	-0.8401	0.5347	-1.8830	-1.5571
535	-3.4501	0.4503	-0.7367	-0.2529	0.4708	-0.2342	-0.0063	0.0886
556	0.5979	-0.3759	-0.8604	0.1142	-0.8877	-0.3959	0.0475	-0.4058
569	1.9401	1.9177	-1.2614	-0.4328	0.3542	0.7875	-1.0437	-0.5411
573	0.7505	0.0006	0.1054	0.0754	0.4119	0.7072	-0.8975	0.6207
739	1.6741	-0.3606	-1.9880	-1.6324	0.8220	-1.2225	0.2169	-1.8246
759	-1.0730	-3.9534	0.9217	0.8253	0.8146	-0.6690	-1.9166	-0.9997
77	-2.8338	-0.5050	1.7943	1.1932	-1.2989	1.2762	-0.1975	-0.4383
779	-0.1485	-0.0053	-0.0027	-0.0281	0.0090	-0.0109	-0.0071	0.0001
800	-0.7646	0.8475	1.2753	-1.2358	2.0385	0.4176	-1.4967	-1.2444
801	-0.5067	-0.6652	0.9826	0.0221	1.5457	-0.1007	1.3320	-1.7720
830	1.2764	-0.2031	0.3138	0.0684	-0.1900	0.0847	-0.0075	-0.0383
866	-3.2263	-0.7882	1.1040	-1.5701	-3.1411	1.9397	0.5586	-0.7882
CB38-22	1.9738	-0.2996	0.4686	0.1156	-0.2874	0.1320	-0.0079	-0.0572
CO421	6.2020	-0.8860	1.4074	0.4019	-0.8776	0.4168	-0.0097	-0.1709
CO617	1.4856	-0.2320	0.3603	0.0826	-0.2193	0.0989	-0.0074	-0.0439
CO945	0.3860	-1.8344	-3.3092	-0.4389	-2.1368	-0.3176	-1.4320	1.0201
D8484	-3.1980	0.4176	-0.6796	-0.2345	0.4345	-0.2168	-0.0055	0.0823
KEN83-737	-0.3878	-0.5516	0.3807	-0.5040	-0.1016	0.5437	0.4966	0.3811
N14	-0.7076	1.5012	-1.2475	2.2448	-1.4450	0.5253	1.4232	0.9602

TABLE 6.6: Varietal Score

6.9 Individual Environments (harvests) cultivars performance for Nzoia PC, Sony527BPC, ChemelilPC, Muhoroni and MumiasPC

No	Nzoia		Sony527Bpc		Chemelil		Muhoroni		MumiasRc	
	Cultv	yield	Cultv	yield	Cultv	yield	Cultv	yield	Cultv	yield
1	759	87.397	800	100.77	271	103.33	339	115.70	526	158.80
2	30	85.923	282	100.70	276	95.14	270	105.83	172	152.67
3	279	85.343	302	94.93	302	89.72	302	101.11	445	138.21
4	535	82.667	801	94.31	303	86.18	166	97.36	CO945	131.65
5	CO945	81.697	166	93.19	313	83.96	526	94.17	739	125.10
6	739	78.027	271	92.57	278	77.50	866	92.22	569	121.46
7	313	62.527	339	91.32	779	75.97	303	92.08	448	121.36
8	448	60.633	278	87.36	446	72.08	448	89.31	303	121.04
9	445	58.847	30	85.42	308	71.67	573	87.91	302	119.38
10	866	56.200	866	81.94	800	68.75	326	86.53	313	118.13
11	77	53.897	303	79.44	526	68.68	308	83.47	759	116.37
12	302	53.433	326	76.71	448	67.22	801	81.39	77	115.43
13	172	52.347	445	76.53	339	66.81	739	78.75	166	113.66
14	830	50.703	N14	66.74	508	65.83	CO617	75.69	279	107.63
15	569	48.617	556	65.19	573	61.25	556	75.42	30	106.48
16	526	43.170	CO945	61.67	CB38-22	57.92	300	71.94	801	101.28
17	166	40.850			569	53.47			KEN83-737	99.41
18	N14	29.393			CO421	21.94			866	92.34
	Mean	61.759		84.30		71.52		89.31		120.02
	CV	28.494		18.33		19.35		22.87		12.31
	MSerror	309.680		238.76		191.60		417.16		218.32
	LSD	29.200		25.77		22.97		34.06		24.52

TABLE 6.7: Individual environment cultivars performance (a)

6.10 Individual Environments (harvests) cultivars performance for WestKenyaPC, Sony527BRC, Sony420PC, Muhoroni and MumiasRC

TABLE 6.8: Individual environment cultivars performance (b)

	West Kenya		Sony527BRC		Sony420Pc		MumiasRc	
No	Cultv	yield	Cultv	yield	Cultv	yield	Cultv	yield
1	D8484	118.55	271	186.50	282	165.85	866	158.80
2	448	114.57	800	174.69	445	165.37	77	152.67
3	271	114.36	866	173.81	CO945	162.18	759	138.21
4	866	98.38	445	167.22	866	160.81	279	131.65
5	739	97.58	302	166.17	303	159.79	302	125.10
6	172	96.47	166	160.69	30	159.67	30	121.46
7	CO945	96.12	339	160.58	N14	152.03	CO945	121.36
8	30	94.56	303	156.89	166	149.28	526	121.04
9	313	92.82	801	153.98	302	147.24	KEN83-737	119.38
10	526	91.68	282	150.16	800	146.24	445	118.13
11	569	90.36	CO945	150.16	801	146.20	303	116.37
12	166	89.39	30	148.32	339	144.03	801	115.43
13	445	87.25	556	135.67	278	142.49	448	113.66
14	77	84.24	278	134.86	556	136.00	172	107.63
15	302	78.57	N14	134.51	271	131.53	166	106.48
16	801	78.20	326	125.79	326	122.72	739	101.28
17	759	67.64					313	99.41
18	303	65.25					569	92.34
		92.00		155.00	Mean	149.46	Mean	120.02
		19.41		12.79	CV	10.32	CV	12.31
		318.72		392.86	MSerror	237.99	MSerror	218.32
		29.62		33.05	LSD	25.72	LSD	24.52

6.11 Overall Mean performance of the cultivars

TABLE 6.9: Overall performance and mean separation

No	Cultivar	Mean yield	difference	No	Cultivar	Mean yield	difference
1	282	138.90	a	21	759	87.62	fgh
2	271	125.66	ab	22	77	87.00	fgh
3	800	122.61	bc	23	448	84.68	fgh
4	D8484	118.55	bcd	24	739	84.04	fgh
5	339	115.69	bcd	25	KEN83-737	83.81	fghi
6	278	110.55	cd	26	526	83.65	ghi
7	866	108.41	d	27	535	82.67	ghij
8	30	106.46	de	28	172	82.62	ghij
9	270	105.83	def	29	313	79.96	hij
10	445	105.67	def	30	308	77.57	hij
11	CO945	105.50	def	31	779	75.97	hijk
12	556	103.07	def	32	CO617	75.69	hijk
13	326	102.94	def	33	573	74.58	hijk
14	801	102.53	def	34	446	72.08	hijk
15	303	102.18	def	35	300	71.94	hijk
16	302	101.93	def	36	569	69.50	ijk
17	166	99.12	def	37	508	65.83	ijk
18	N14	95.67	efg	38	CB38-22	57.92	jk
19	276	95.14	efgh	39	830	50.70	k
20	279	89.16	fgh	40	CO421	21.94	l
	Mean	97.01	CV	18.18	MSerror	311	

6.12 Parameters estimation for the combined Environments ANOVA

TABLE 6.10: Parameters estimations (a)

Parameters	Estimate	Std.Error	t value	$Pr(>\|t\|)$	
(Intercept)	39.5987	17.9537	2.206	0.0282	*
envirmuho	56.6844	20.9498	2.706	0.00722	**
envirmumia	18.5641	20.9154	0.888	0.3755	
envirmumiasrc1	66.1539	20.9154	3.163	0.00173	**
envirnz	3.5119	20.9154	0.168	0.86677	
envirsony420pc	112.5575	20.9498	5.373	1.59E-07	***
envirsy527B	49.23	20.9498	2.35	0.01945	*
envirsy527brc	122.4757	15.221	8.046	2.20E-14	***
envirwk	42.8293	20.9154	2.048	0.04149	*
variety172	7.0733	14.395	0.491	0.62353	
variety270	8.4733	14.395	0.589	0.55657	
variety271	48.5767	20.3576	2.386	0.01767	*
variety276	40.38	20.3576	1.984	0.04825	*
variety278	22.7467	20.3576	1.117	0.26477	
variety279	44.4933	14.395	3.091	0.00219	**
variety282	-10.5333	14.395	-0.732	0.46492	
variety30	5.17	14.395	0.359	0.71974	
variety300	-25.4167	14.395	-1.766	0.07851	.
variety302	34.9667	20.3576	1.718	0.08693	.
variety303	31.4233	20.3576	1.544	0.12378	
variety308	16.91	20.3576	0.831	0.40686	
variety313	29.2033	20.3576	1.435	0.1525	
variety326	-34.9033	14.395	-2.425	0.01593	*
variety339	12.05	20.3576	0.592	0.55437	
variety445	-2.1467	14.395	-0.149	0.88156	
variety446	17.3233	20.3576	0.851	0.3955	
variety448	12.4667	20.3576	0.612	0.54076	

TABLE 6.11: Parameters estimation(b)

Parameters	Estimate	Std.Error	t value	$Pr(>\|t\|)$	
variety508	11.0767	20.3576	0.544	0.58679	
variety526	13.9233	20.3576	0.684	0.49456	
variety535	41.8167	14.395	2.905	0.00396	**
variety556	-25.0233	14.395	-1.738	0.08321	.
variety569	-1.2867	20.3576	-0.063	0.94965	
variety573	6.4933	20.3576	0.319	0.74998	
variety739	8.19	14.395	0.569	0.56983	
variety759	-21.75	14.395	-1.511	0.13189	
variety77	-5.1533	14.395	-0.358	0.72061	
variety779	21.2167	20.3576	1.042	0.29819	
variety800	13.9967	14.395	0.972	0.3317	
variety801	-11.1967	14.395	-0.778	0.43731	
variety830	9.8533	14.395	0.684	0.49421	
variety866	8.9867	14.395	0.624	0.53293	
varietyCB38-22	3.16	20.3576	0.155	0.87675	
varietyCO421	-32.8133	20.3576	-1.612	0.10808	
varietyCO617	-21.6667	14.395	-1.505	0.13337	
varietyCO945	6.7267	14.395	0.467	0.64064	
varietyD8484	29.1567	14.395	2.025	0.04373	*
varietyKEN83-737	12.9	14.395	0.896	0.37092	
varietyN14	-26.1833	14.395	-1.819	0.06996	.
envirchem:block2	18.6461	5.8767	3.173	0.00167	**
envirmuho:block2	9.1681	6.2332	1.471	0.14242	
envirmumia:block2	-1.9328	5.8767	-0.329	0.74248	
envirmumiasrc1:block2	-3.7983	5.8767	-0.646	0.51857	
envirnz:block2	-7.8794	5.8767	-1.341	0.18104	
envirsony420pc:block2	-2.8012	6.2332	-0.449	0.65347	
envirsy527B:block2	-1.2531	6.2332	-0.201	0.84081	
envirsy527brc:block2	1.3275	6.2332	0.213	0.8315	

TABLE 6.12: Parameter estimations (c)

Parameters	**Estimate**	**Std.Error**	**t value**	$Pr(>\|t\|)$	
envirwk:block2	12.9706	5.8767	2.207	0.02809	*
envirchem:block3	26.8278	5.8767	4.565	7.39E-06	***
envirmuho:block3	-5.9375	6.2332	-0.953	0.34161	
envirmumia:block3	-5.4056	5.8767	-0.92	0.35843	
envirmumiasrc1:block3	5.9806	5.8767	1.018	0.30968	
envirnz:block3	1.0978	5.8767	0.187	0.85195	
envirsony420pc:block3	-5.8375	6.2332	-0.937	0.34979	
envirsy527B:block3	14.3469	6.2332	2.302	0.02206	*
envirsy527brc:block3	-5.4806	6.2332	-0.879	0.37999	
envirwk:block3	7.9256	5.8767	1.349	0.17851	
envirmumia:variety172	11.26	20.3576	0.553	0.58061	
envirmumiasrc1:variety172	-5.9267	20.3576	-0.291	0.77116	
envirnz:variety172	4.4233	20.3576	0.217	0.82814	
envirsony420pc:variety271	-66.3233	24.9329	-2.66	0.00825	**
envirsy527B:variety271	-49.2	24.9329	-1.973	0.04941	*
envirsy527brc:variety271	-22.7667	20.3576	-1.118	0.26435	
envirwk:variety271	-23.6133	24.9329	-0.947	0.34439	
envirsony420pc:variety278	-29.5333	24.9329	-1.185	0.23718	
envirsy527B:variety278	-28.5767	24.9329	-1.146	0.25268	
envirsy527brc:variety278	-48.58	20.3576	-2.386	0.01766	*
envirmumia:variety279	-49.7333	20.3576	-2.443	0.01516	*
envirmumiasrc1:variety279	-19.32	20.3576	-0.949	0.34339	
envirsony420pc:variety282	27.1067	20.3576	1.332	0.18406	
envirsy527B:variety282	18.0367	20.3576	0.886	0.37636	
envirmumia:variety30	-11.04	20.3576	-0.542	0.58803	
envirmumiasrc1:variety30	9.8133	20.3576	0.482	0.63014	
envirnz:variety30	39.9033	20.3576	1.96	0.05094	.
envirsony420pc:variety30	5.2233	20.3576	0.257	0.79769	
envirsy527B:variety30	-12.9433	20.3576	-0.636	0.52541	

TABLE 6.13: Parameters estimations (d)

Parameters	Estimate	Std.Error	t value	$Pr(>\|t\|)$	
envirsy527brc:variety30	-17.54	20.3576	-0.862	0.38962	
envirmuho:variety302	-31.2133	24.9329	-1.252	0.21162	
envirmumia:variety302	-29.61	24.9329	-1.188	0.23597	
envirmumiasrc1:variety302	-16.3467	24.9329	-0.656	0.51258	
envirnz:variety302	-22.3833	24.9329	-0.898	0.37007	
envirsony420pc:variety302	-37.0033	24.9329	-1.484	0.13886	
envirsy527B:variety302	-33.2267	24.9329	-1.333	0.18369	
envirsy527brc:variety302	-29.4833	20.3576	-1.448	0.14862	
envirwk:variety302	-45.7933	24.9329	-1.837	0.06728	.
envirmuho:variety303	-36.7033	24.9329	-1.472	0.14208	
envirmumia:variety303	-25.69	24.9329	-1.03	0.3037	
envirmumiasrc1:variety303	-21.5367	24.9329	-0.864	0.38842	
envirsony420pc:variety303	-20.9067	24.9329	-0.839	0.40243	
envirsy527B:variety303	-45.1733	24.9329	-1.812	0.07105	.
envirsy527brc:variety303	-35.2267	20.3576	-1.73	0.08462	.
envirwk:variety303	-55.5633	24.9329	-2.229	0.02661	*
envirmuho:variety308	-30.7967	24.9329	-1.235	0.21776	
envirmumia:variety313	-23.8467	24.9329	-0.956	0.33965	
envirmumiasrc1:variety313	-36.2733	24.9329	-1.455	0.1468	
envirnz:variety313	-7.5267	24.9329	-0.302	0.76296	
envirwk:variety313	-25.7767	24.9329	-1.034	0.30207	
envirmuho:variety326	24.0733	20.3576	1.183	0.23797	
envirsony420pc:variety326	8.35	20.3576	0.41	0.68199	
envirsy527B:variety326	18.42	20.3576	0.905	0.36631	
envirmuho:variety339	6.2867	24.9329	0.252	0.80111	
envirsony420pc:variety339	-17.2967	24.9329	-0.694	0.48841	
envirsy527B:variety339	-13.9233	24.9329	-0.558	0.57698	
envirsy527brc:variety339	-12.1633	20.3576	-0.597	0.55065	
envirmumia:variety445	12.75	20.3576	0.626	0.53161	

TABLE 6.14: Parameters estimation (e)

Parameters	Estimate	Std.Error	t value	$Pr(>\|t\|)$	
envirmumiasrc1:variety445	13.7967	20.3576	0.678	0.49849	
envirnz:variety445	20.1433	20.3576	0.989	0.32326	
envirsony420pc:variety445	18.2433	20.3576	0.896	0.37092	
envirsy527B:variety445	-14.52	20.3576	-0.713	0.47627	
envirsy527brc:variety445	8.68	20.3576	0.426	0.67015	
envirmuho:variety448	-20.52	24.9329	-0.823	0.41118	
envirmumia:variety448	-5.49	24.9329	-0.22	0.82588	
envirmumiasrc1:variety448	-5.2867	24.9329	-0.212	0.83223	
envirnz:variety448	7.3167	24.9329	0.293	0.76938	
envirwk:variety448	12.7133	24.9329	0.51	0.61051	
envirmuho:variety526	-17.1167	24.9329	-0.687	0.49294	
envirmumia:variety526	13.5067	24.9329	0.542	0.58843	
envirmumiasrc1:variety526	0.64	24.9329	0.026	0.97954	
envirnz:variety526	-11.6033	24.9329	-0.465	0.64201	
envirwk:variety526	-11.6333	24.9329	-0.467	0.64115	
envirmuho:variety556	3.08	20.3576	0.151	0.87985	
envirsony420pc:variety556	11.7467	20.3576	0.577	0.56438	
envirsy527B:variety556	-2.9767	20.3576	-0.146	0.88385	
envirmumia:variety569	8.2667	24.9329	0.332	0.74046	
envirmumiasrc1:variety569	-12.8567	24.9329	-0.516	0.60649	
envirnz:variety569	9.0533	24.9329	0.363	0.71679	
envirwk:variety569	2.2533	24.9329	0.09	0.92805	
envirmuho:variety573	-15.94	24.9329	-0.639	0.52312	
envirmuho:variety739	-26.8	20.3576	-1.316	0.18906	
envirmumia:variety739	0.6533	20.3576	0.032	0.97442	
envirmumiasrc1:variety739	-13.39	20.3576	-0.658	0.51123	
envirnz:variety739	28.9867	20.3576	1.424	0.15556	
envirmumia:variety759	23.2467	20.3576	1.142	0.25443	
envirmumiasrc1:variety759	53.4767	20.3576	2.627	0.00908	**

TABLE 6.15: Parameters estimations (f)

Parameters	Estimate	Std.Error	t value	$Pr(>\|t\|)$	
envirnz:variety759	68.2967	20.3576	3.355	0.0009	***
envirmumia:variety77	6.6433	20.3576	0.326	0.74441	
envirmumiasrc1:variety77	51.34	20.3576	2.522	0.01221	*
envirnz:variety77	18.2	20.3576	0.894	0.37206	
envirsony420pc:variety800	-17.0367	20.3576	-0.837	0.40335	
envirsy527B:variety800	-6.4233	20.3576	-0.316	0.75259	
envirmuho:variety801	-4.7733	20.3576	-0.234	0.81478	
envirmumia:variety801	3.7133	20.3576	0.182	0.85539	
envirmumiasrc1:variety801	20.1433	20.3576	0.989	0.32326	
envirsony420pc:variety801	8.1167	20.3576	0.399	0.6904	
envirsy527B:variety801	12.31	20.3576	0.605	0.54586	
envirsy527brc:variety801	4.4833	20.3576	0.22	0.82585	
envirmuho:variety866	-14.1233	20.3576	-0.694	0.48839	
envirmumia:variety866	-19.59	20.3576	-0.962	0.3367	
envirmumiasrc1:variety866	43.3367	20.3576	2.129	0.03412	*
envirnz:variety866	6.3633	20.3576	0.313	0.75483	
envirsony420pc:variety866	2.5433	20.3576	0.125	0.90066	
envirsy527B:variety866	-20.2367	20.3576	-0.994	0.32102	
envirsy527brc:variety866	4.13	20.3576	0.203	0.83938	
envirmumia:varietyCO945	2.8733	20.3576	0.141	0.88785	
envirmumiasrc1:varietyCO945	8.1533	20.3576	0.401	0.68908	
envirnz:varietyCO945	34.12	20.3576	1.676	0.09481	.
envirsony420pc:varietyCO945	6.18	20.3576	0.304	0.76167	
envirsy527B:varietyCO945	-38.2533	20.3576	-1.879	0.06124	.
envirsy527brc:varietyCO945	-17.26	20.3576	-0.848	0.39723	
envirmumia:varietyKEN83-737	-20.3833	20.3576	-1.001	0.31753	
envirnz:varietyN14	14.7267	20.3576	0.723	0.47002	
envirsony420pc:varietyN14	28.9333	20.3576	1.421	0.15632	
envirsy527B:varietyN14	-0.27	20.3576	-0.013	0.98943	

TABLE 6.16: EM-SVD and EM-AMMI points of comparison

No.	Areas of Comparison	EM-SVD	EM-AMMI
1	R Package required	It requires many packages alongside their utilities some of which are not available directly on the CRAN but can be accesses from archives which is a long process. The main packages are; 'bvc', Imput.svd and cv.SVDImpute Other utility packages that must be loaded; Gbm, survival, lattice, splines, parallel, gbm 2.1.1, TimeProjection, lubridate, timeDate, Matrix and locfit	The package is not in the CRAN and the codes are generated which are a bit complex
2	Complexity of codes generation	Simple if packages are available	Complex- Absence of specific CRAN package for EM-AMMI imputation requires codes generation
3	Choice of interactive components for imputation	cv.SVDImpute determines the best rank for imputation. High ranked GEI matrix gives options of imputing with low rank	Number of interactive principal for imputation determined by min ((genotypes nos. of data elements available, environments no. of elements available)
5	Attaining convergence	Converges up to rank 2 of the GEI matrix	Converges only for the for the 1st PC
6	No of iterations in imputing the missing	189 at the lowest rank of 1 89,946 at rank 2 and gives unrealistic figures (even negatives)	74 iteration with the 0PC

TABLE 6.17: EM-SVD and EM-AMMI points of comparison continued

No.	Areas of Comparison	EM-SVD	EM-AMMI
7	No. of iterations in predicting the non-missing using imputed figures	166 for the rank 1	60 for the 0 PC EM_AMMI takes lesser iterations
8	Code execution efficiency (run time)	For the Missing User time -4.69 seconds System time -0.30 seconds Elapsed time -5.35 seconds	For the Missing User time -0.16 seconds System time -0.14 seconds Elapsed time -0.30 seconds With confirmation process User time -1.17 seconds System time -0.10 seconds Elapsed time -2.68 seconds
		For the none missing User time -0.04 seconds System time -0.02 seconds Elapsed time -0.05 seconds	For the none missing User time -0.17 seconds System time -0.05 seconds Elapsed time -0.22 seconds With confirmation process User time -1.01seconds System time -0.11 seconds Elapsed time -1.95 seconds
9	Model selection based on significant IPCA	AMMI1	AMMI0
10	PRESS values for error in imputation	118.86 at rank one	55.18