
# LONGITUDINAL AND GENOTYPE BY ENVIRONMENT INTERACTION ANALYSIS OF ARABICA COFFEE BEAN YIELD IN SOUTHWEST ETHIOPIA: APPLICATION OF LINEAR MIXED MODEL

By

Tarekegn Argaw

A Thesis Submitted to the Department of Statistics, College of Natural Sciences, Jimma University, as a Partial Fulfillment of the Requirements for the Degree of Master of Science (MSc) in Biostatistics.

October, 2016

Jimma, Ethiopia

# LONGITUDINAL AND GENOTYPE BY ENVIRONMENT INTERACTION ANALYSIS OF ARABICA COFFEE BEAN YIELD IN SOUTHWEST ETHIOPIA: APPLICATION OF LINEAR MIXED MODEL

By

Tarekegn Argaw

Advisor: Girma Taye (PhD)

Co-advisor: Dechasa Bedada (MSc)

October, 2016

Jimma, Ethiopia

**DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES JIMMA UNIVERSITY**

As thesis research advisors, we herby certify that we have read the thesis prepared by TAREKEGN ARGAW under our guidance, which is entitled "**Longitudinal and Genotype by Environment Interaction Analysis of Arabica Coffee Bean Yield in Southwest Ethiopia: Application of Linear Mixed Model**", in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

| | | |
|---|---|---|
| Girma Taye (PhD) | _____ | _____ |
| Advisor | Signature | Date |
| Dechasa Bedada(MSc) | _____ | _____ |
| Co-advisor | Signature | Date |

As the members of the board of examiners of MSc thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

| | | |
|---|---|---|
| _____ | _____ | _____ |
| Name of Chairman | Signature | Date |
| _____ | _____ | _____ |
| Name of Major Advisor | Signature | Date |
| _____ | _____ | _____ |
| Name of Co-advisor | Signature | Date |
| _____ | _____ | _____ |
| Name of Internal Examiner | Signature | Date |
| _____ | _____ | _____ |
| Name of External Examiner | Signature | Date |

## STATEMENT OF THE AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate. Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Tarekegn Argaw

Date: _____

Signature: _____

Jimma University, Jimma

# DEDICATION

This thesis is dedicated to my family especially my mother, Tsehaynesh Tadesse, who were with me at the time of my Happiness and Terrible throughout my study!!!

# BIOGRAPHYCAL SKETCH

The author, Tarekegn Argaw was born on June 8, 1990 at Jamma Wereda, South Wollo Zone, Amhara, Ethiopia. He attended primary school at Elshama Elementary School, starting from February 1997. He has attended secondary and High school education at Degollo Preparatory School. Tarekegn joined Addis Ababa University in September 2009 and graduated with BSc degree in Statistics in July 2011. After his graduation, he Joined Ethiopian Institute of Agricultural Research and worked as junior researcher at JARC for three years. He then joined Jimma University in September 2014, to pursue postgraduate studies of MSc Degree in Biostatistics.

# ACKNOWLEDGEMENTS

First and foremost, I am indebted to the Almighty God because of whose full mercy and grace I could complete my study and to all my friends.

I would like to express my deepest and heart full gratitude to my major adviser, Dr. Girma Taye, for his constructive comments and excellent academic advices, during proposal and thesis writing. I am sincerely grateful to my Co-Adviser Mr. Dechasa Bedada, for his very kind and constructive advices, comments encouragements and appreciation during my entire work on this thesis.

I would like to extend my gratitude to Mr. Ashenafi Ayano, Senior Researcher and Center Director at Jimma Agricultural Research Center for his very kind and positive response up on my request for data, and sharing of the professional knowledge and experience that could possibly helped me in my thesis. I am also indebted to Dr. Taye Kufa and Mr. Tadesse Benti, Senior Researchers at JARC for their consistent support and sharing of the professional knowledge and experience throughout the course of the study.

I would like to thank the host Jimma University and my sponsor Ethiopian Institute of Agricultural Research for providing me to attend my training, and Jimma Agricultural Research center for providing me with all relevant secondary data used in this study.

My sincere thanks also go to my friends, classmate's and all staff member of department of statistics of Jimma University for their unreserved knowledge sharing.

Last but not least, I would like to thank my parents who have invested all their life to hold up me. My parents, I would like to thank you vastly for your sweet words at every time that gave me a long power and anticipate.

# ABSTRACT

**Background:** Arabica coffee (*Coffea arabica* L.*)* is the fine flavored, aromatic type makes up 60-65% of total production and usually fetches the highest prices. Arabica Coffee is the most important and backbone of Ethiopian economy, which accounts for an average 60% of export earnings. Coffee is a perennial crop which can be harvested multiple times of years, and it is known to be affected with a characteristic biennial, which is more pronounced in the species Arabica coffee. The immediate objective of this study was to analyze Arabica coffee bean yield longitudinally by using Linear Mixed Model (LMM), and to assess its Genotype by Environment interaction (GEI). Coffee Bean Yield (CBY), Coffee Yield, and Yield are used interchangeably in this document.

**Methods:** The data for this study came from coffee variety field trials conducted by Jimma Agricultural Research Center (JARC) over several years. The trial was conducted in south west Ethiopia across coffee growing areas (Jimma, Agaro, and Metu). The experimental design of the trial was RCBD with 4 replications and 17 Arabica coffee genotypes. A complete CBY data set of these coffee growing areas which had been collected during 2005-2011 was considered in this study. Exploratory Data Analysis (EDA) and LMM were employed for longitudinal analysis, whereas combined ANOVA and AMMI model were used for GEI analysis. All analyses were done with the help of R statistical package.

**Results:** The LMM results revealed that the heterogeneous variance function (varIdent(t)) and autoregressive order three (AR3) were, respectively, found to give better fit to the variance and correlation structure among measurements of CBY. Biennial interacts significantly with location and genotype. The estimated variance of random effect of block associated with intercept and biennial were $\hat{\sigma}^2$ $(b_{0j}) = (221.81)^2$ and $\hat{\sigma}^2$ $(b_{3j}) = 145.24^2$, respectively. The result also showed significant location by linear and quadratic time effect interactions. Estimates of quadratic time effects for Jimma, Agaro, and Mutu were, respectively, -151.51, -66.05, and -4, whereas estimates of linear time effects for these locations were 158.92, 158.92, and 31.08, respectively. The combined analysis of variance revealed that the genotype, environment, and GEI effects are highly significant (P-values<0.001). GEI accounted for 16.2% of the total sum of squares and was about 2 times larger than that of genotypes. The AMMI procedure revealed that AMMI-5 was the best truncated AMMI model that can sufficiently explain the information contained in GEI. The first three interaction principal components (IPC1, IPC2 and IPC3) retained by Gollob's F-test for graphical display accounted for 64.2% of GEI.

**Conclusion:** The measurements of CBY that are obtained from Arabica coffee tree over time induce an autocorrelation which is known as serial correlation. There is initially an increasing and gradually a decreasing trend in Arabica CBY over time years with linear rate of growth. There is also a differential response of genotypes and environments in the presence and absence of biennially. The major factor that influence yield performance of Arabica coffee in Ethiopia is the environment, and among 17 Arabica coffee genotypes, G1, G2, G3, G7, G8, G9 and G12 have the best performance with G1, G2, G3, G8 and G12 being relatively stable across the test environments. It was recommended to use information from longitudinal and GEI analysis to investigate the effect of time and biennial and the association between genotype and environment in Arabica CBY.



**Key Words:** Arabica Coffee, Biennia, Clustered Longitudinal Data, GEI, LMM

# LIST OF ACRONYMS

| | |
|---|---|
| AIC | Akaike Information Criterion |
| AMMI | Additive Main Effects and Multiplicative Interaction |
| ANCOVA | Analysis of Covariance |
| ANOVA | Analysis of Variance |
| AR1 | Autoregressive Order One |
| BIC | Bayes Information Criterion |
| BLUP | Best Linear Unbiased Prediction |
| CBD | Coffee Berry Disease |
| CBY | Coffee Bean Yield |
| CSA | Central Statistical Agency |
| ECEA | Ethiopia Commodity Exchange Authority |
| EDA | Exploratory Data Analysis |
| EIAR | Ethiopian Institute of Agricultural Research |
| FAMM | Factor Analytic Multiplicative Mixed |
| GDP | Gross Domestic Product |
| GEI | Genotype by environment interaction |
| GLM | General Linear Model |
| ICO | International Coffee Organization |
| IPCA | Interaction Principal Component Analysis |
| JARC | Jimma Agricultural Research Center |
| LMM | Linear Mixed Model |
| LRT | Likelihood Ratio Test |
| MANOVA | Multivariate Analysis of Variance |
| MIVQUE | Minimum Variance Quadratic Unbiased Estimator |
| MLE | Maximum Likelihood Estimator |
| MoARD | Ministry of Agriculture and Rural Development |
| MSE | Mean Square Error |
| PCA | Principal Component Analysis |
| RCBD | Randomized Complete Block design |
| REML | Restricted Maximum Likelihood Estimation |

# TABLE OF CONTENTS

## LIST OF TABLES

# LIST OF FIGURES

# 1. NTRODUCTION

## 1.1. Background

Arabica coffee (*Coffea arabica* L.*)* belongs to the genus Coffea in the Rubiaceae family and is a self-fertile allotetraploid species that is mostly grown in the tropical and subtropical regions(Berhanu *et al*, 2015). Of the 124 species in the genus Coffea, Arabica coffee (*Coffea arabica* L.*)* and Robusta coffee (*Coffea canephora* P.*)* are the two most important commercial species (Gichuru, *et al*, 2008; Davis *et al*, 2011, as cited in Lemi, 2016).

Economically, coffee is the second most exported commodity after oil, and employs over 100 million people worldwide (Gray *et al.*, 2013). According to ECEA (2008), coffee is exported in its raw, roasted, or soluble product forms to more than 165 countries worldwide by more than 121 countries, and about 17 of these countries get 25 percent of their foreign exchange earnings from coffee. Among 124 species of coffee (genus *Coffea*) (Davis *et al*, 2011), Arabica coffee and Robusta are the two botanical varieties which originate from Africa and consumed widely as non-alcoholic stimulant beverage in the world (Nuhu, 2014). Arabica coffee is the fine-flavored, aromatic type makes up 60-65% of total production and usually fetches the highest prices, whereas Robusta is easier to produce and more resistant to disease (Hilten, 2002, Chauhan, 2015).

In Africa, coffee is one of the most important commodities generating substantial income to rural communities, contributing to the fight against extreme poverty and key to achieving sustainable development. It accounts for the primary source of income for more than 10 million households in 25 African coffee-growing countries. Some of these countries depend on coffee as a primary source of income for their rural population and an important source of export revenues. It is a vital contributor to foreign exchange earnings in addition to accounting for a significant proportion of taxable income and Gross Domestic Product (ICO, 2015)

Ethiopia is the birth place of Arabica coffee that originates in the southwest area of the country (Bart *et al*, 2014), a fact of which Ethiopians is understandably proud (Kassahun *et al.*, 2008). It is from this part of the country that coffee spread to the rest of the world and constituted the ancestor of the present day coffee plantations across the globe. Many researchers consolidate the

idea that Ethiopia is the primary center of origin and genetic diversity of Arabica coffee with considerable heterogeneity (Kassahun *et al.*, 2008; Taye, 2011). As the country of origin for crop, Ethiopia produces premium quality Arabica coffee and is the leading producer in Africa, and the 5th in the world, following Brazil, Vietnam, Colombia and Indonesia (ICO, 2015). In the country, at least 15 million people directly or indirectly rely on coffee for their livelihood (Ministry of Trade, 2012; Gray *et al.* 2013). Ethiopia is also the second largest exporter of organic coffee by volume after Peru (Kodama, 2009). Generally, coffee is most important and backbone of Ethiopian economy, which accounts for an average about 5% of GDP, 10% of the total agriculture production and 60% of export earnings (Chauhan *et al*., 2015).

The land covered with coffee in Ethiopia is very large and is estimated to range from 400,000 to 650,000 ha.( Jefuka *et al*., 2012, as cited in Yonas *et al*, 2014b), and total coffee production is about 200,000 tones of clean coffee per year (Chauhan *et al*., 2015). Despite the role coffee plays in the national economy and in spite of the country is origin of Arabica coffee, the national average coffee yield level is low as compare to major coffee growing regions in the world (Taye *et al,* 2011b). The factors attributed to such low productivity include lack of resistant variety to various diseases and insect pests and poor agronomy practice (Workafes *et al*., 2000). Lack of suitable varieties that exhibit stable performance across wide ranges of environments is also one of the major factors attributed to low productivity of Arabica coffee (Yonas and Bayetta, 2008).

Arabica coffee grows in Ethiopia in several places at various altitudes ranging from 550 -2750 meters above sea level (Quintin *et al*., 2013). The major producing areas are concentrating in the southwestern part of Ethiopia where Arabica coffee originated and diverse (Kassahun, 2008; Taye *et al*, 2011a). Arabica coffee grows under very diverse environments including annual rainfall (1000 – 2000 mm), temperature (minimum 8 – 15 $^0$C, maximum 24 – 31 $^0$C) and soil type, and this has a lot contribution to the high genetic diversity and as though high yield variability within Arabica coffee in the country (Mesfin and Bayetta, 1987). Yonas (2005) pointed out that there is strong variation within southwestern region of Ethiopia due to climatic and edaphic variations along altitudinal gradient. Environment and genotype have roles in determining the yield of Arabica coffee and they are also important factors for breeding purpose

(Tesfaye *et al.*, 2008; Alemseged and Tesfaye, 2012). Since Ethiopia has both wide genetic diversity and diverse environment for growing Arabica coffee, conducting multi-location trial

over years is important to assess GEI and identify stable genotype which can increase productivity of Arabica coffee in the country (Yonas, 2014a).

This study was greatly intended to model multi environment coffee yield data using cluster longitudinal data-modeling framework. Longitudinal data is a term refers to measurements made repeatedly over time to study how the subjects evolve over time. Observations from cross sectional data are uncorrelated but in longitudinal study the measurements made for subjects over a period of time are correlated. Longitudinal studies play a prominent role in almost all endeavors. They are indispensable to the study of change in an outcome over time. By measuring study participants repeatedly through time, longitudinal studies allow the direct study of temporal changes within individuals and the factors that influence change (Fitzmaurice *et al.*, 2009). Laird and Ware (1982) proposed a flexible class of linear mixed-effects models for longitudinal data. These models could handle the complications of mistimed and incomplete measurements in a very natural way.

Correspondingly, the study also aimed to assess the effect GEI in coffee yield to investigate performance coffee genotypes across environment. In previous studies, a number of parametric statistical procedures have been elaborated over the years to analyze genotype by environment interaction and yield stability over environment. These statistical methods broadly categorized in to two classes, univariate and multivariate models. Univariate models encompass a range of models, such as combined ANOVA, regression slope, deviation from regression, environmental variance, and Kang's yield-stability (Eberhart & Russell, 1966; Freeman, 1973; Chakroun *et al*., 1990). Multivariate models are more powerful and flexible to investigate GEI, and they have gotten special attention in theory and application (Zobele *et al*, 1988; Girma *et al*, 2000; Girma and Dan, 2014). These models are linear-bilinear models such as, AMMI, Site Regression (SREG), Genotypic Regression (GREG), Completely Multiplicative Model (COMM) and Factor Analytic (FA).

The AMMI model combines the analysis of variance for the genotype and environment main effects with principal component analysis of the genotype by environment interaction. It has proven useful and widely used for understanding complex GEI (van Eeuwijk, 1995; Girma, 2000 Zelalem, 2011; Dejene, 2016). The results can be graphed in a useful bi-plot that shows both main and interaction effects for both genotypes and environments.

### 1.2. Statement of the Problem

In Ethiopia, agriculture is the backbone of the economy, employs about 85% of the population and account for around 90% of foreign exchange and 40% of the growth domestic product (Chauhan *et al.*, 2015). Even thought Ethiopia is the center and origin of several crops (IBC, 2008), the increase in human population and the decrease in arable land cause the country fail to increase agricultural production output. With this scenario, therefore, the best strategy of increasing production of agricultural output is by increasing productivity per unit area using improved and reliable production technologies which are achieved by proper statistical design and model (Girma, 2010). Accordingly, plant breeders conduct multi-location trials over years to evaluate yield performance of genotypes across environment and to assess the GEI interaction, aiming to release quality, stable, high yielder, and disease resistant improved varieties. Such improved variety can be obtained after rigorous breeding and selection procedures that involve testing of a large collection of genotypes across diverse environments and use of efficient statistical methods (Asfaw *et al*., 2011, as cited in Degene, 2016).

Data collected in multi-location trials are intrinsically complex, having three fundamental aspects: structural patterns, nonstructural noise and relationships among genotypes, environments, and genotypes and environments considered jointly (Crossa, 1990). For the analysis of such data especially for GEI interaction, various statistical methodologies have been extensively reviewed and documented (Zelalem, 2011; Hasan *et al*., 2011; Degene, 2016). Among these statistical methodologies, the most commonly used statistical techniques for analyzing multi-environment trial (MET) data are the combined analysis of variance (ANOVA) and linear regression techniques. However, they are open to criticism due to the fact that they do not discern patterns of the underlying genotype by environment interaction, and the assumptions of normality, independency and constant variance may not be always satisfied (de Resende,

2007). Another statistical method that commonly used in agricultural researches is the additive main effect and multiplicative interaction effect (AMMI) model. But, current studies show that the linear mixed model is more powerful than AMMI model because of its flexibility in the assumption of ANOVA and missing data (de Resende, 2007; Crossa *et al*., 2010, as cited in Degene, 2016).

However, such extensive studies have been well reviewed and documented on annual crops but not that much for perennial crops like coffee by differ the distinction between them. Field trials with perennial crops give rise longitudinal measurements taken on the same plot on several occasions (Piepho and Eckl, 2014). So, it is important to account for correlation among repeated measurements in such trials. Similarly, time effects need to be taken into account to avoid overestimation in genetic parameters and thereby estimate genetic trend (longitudinal evolution) (Laidig *et al*., 2014).

Coffee is a perennial crop with more than 124 species of which Arabica coffee is economical important (Gichuru *et al*, 2008). Like annual crops, coffee breeders generate multi-location trial data over year to evaluate the yield performance of coffee genotypes across location over year. Again, the statistical methods which used to analyze such data are also open to criticism not only due to the correlation among measurements but also the biennial property of coffee (Rodriguez *et al.*, 2013). Biennial is a phenomenon that occurred in two year interval which results alternation of high and low yield along with consecutive years, and it is more pronounced in the species Arabica coffee. (Taye *et al.*, 2001; Tesfaye *et al.*, 2002; Bernardes *et al*, 2012; Rodriguez *et al.*, 2013)

In Ethiopia, due to the wealth of coffee ecology and the dominant role of Arabica coffee in the national economy, the country is emerged with an opening opportunity to carry out coffee research aiming to increase coffee productivity with improved technologies (Taye *et al*., 2001; Bayetta *et al.*, 2008). Accordingly, in the conventional linear model setting, various studies have been conducted to analyze the effect of genotype, environment, and to asses GEI interaction and yield stability of Arabica coffee regardless of its longitudinal (repeated since perennial) and biennial property. Thus, no information is available on the correlation among measurements of

coffee, longitudinal time effect (genetic trend or evolution of coffee yield over time), and biennial effect. By using linear mixed model, therefore, handling these open criticisms is a great deal of interest in this study.

In general, the motivation behind this study is intended to address the following research questions in two major approaches

- Longitudinal Analysis

✓ How does the yield of coffee change over time/year?

✓ Is there significant biennial effect on the yield of coffee?

✓ Is there significant correlation among measurements of coffee yield over year?

- Genotype by Environment Interaction Analysis(GEI)

✓ Is there significant coffee genotype by Environment interaction?

✓ How do coffee genotypes perform across environment?

## 1.3. Objective of the Study

### 1.3.1. General objective

To model Arabica coffee bean yield longitudinally by using Linear Mixed Model (LMM), and to assess its Genotype by Environment Interaction (GEI)

### 1.3.2. Specific objectives

- To assess the evolution of coffee bean yield over time
- To investigate the effect of biennial in coffee bean yield
- To model the variance and correlation structure among repeated measurements of coffee bean yield
- To assess effect of genotype by environment interaction in coffee yield

## 1.4. Significance of the Study

The result of this study will provide information of how the yield of coffee changes over time, the effect of biennial in Arabica CBY, and the relation between coffee genotype and environment. Specifically,

- ✓ To provide more valid and reliable information about the factors which affect coffee yield
- ✓ To provide information for coffee researchers on longitudinal data analysis approach for multi-location trial over year.
- ✓ Use as a stepping stone for further studies related to multi-location trial over year for perennial crops like coffee.

# 2. LITERATURE REVIEW

## 2.1. Origin and Genetic Diversity Coffee

The coffee plant is indigenous to Africa, and it was in Ethiopia that the habit of drinking coffee first developed (ICO, 2015). Ethiopia is widely regarded as the birth place of coffee (Amamo, 2014). The name 'coffee' is believed to originate from the name of the province Kaffa in the southwestern part of the country, where according to legend; a goat herder discovered coffee beans during the sixth century A.D. (Gomez-Ruiz *et al.*, 2007). Coffee is under-story plant of the ever-green afromontan rain-forest of the southern western part of the country and it originated in the southwestern part of Ethiopian highland where it was first discovered (Getachew *et al.*, 2013). It is from this part of the country that coffee spread to the rest of the world and constituted the ancestor of the present day coffee plantations across the globe (Kassahun *et al.*, 2008).

On the perspective of coffee genetic diversity, researchers have seen that the country possesses a diverse genetic base for Arabica coffee with considerable heterogeneity (Yonas and Bayata, 2008; Bayetta, 2011). Many native researchers consolidate the idea that Ethiopia is the primary center of origin and genetic diversity of coffee (Arabica coffee) with considerable heterogeneity. Tesfaye (2006) reported that there was high genetic variability within and between different wild populations in Ethiopia. He further noted that these populations of Arabica coffee in the montane rainforests are the most important, and are genetically distinct and more diverse when compared to the cultivated varieties grown in Ethiopia and around the world. According to Anthony *et al*. (2002) and Denich *et al*. (2006), this was also confirmed by the fact that within small area, the wild coffee plants of Ethiopia have relatively high genetic variability as compared to the wild coffee of other countries that showed a characteristically low genetic diversity. Moreover, the existence of high genetic diversity of coffee plants is due to Ethiopia's suitable altitude, ample rain fall, optimum temperature and planting materials. Generally, Arabica coffee from southwest of Ethiopia showed a relatively high genetic diversity (Anthony *et al*., 2001; Anthony *et al*., 2002; Chaparro *et al*., 2004).

## 2.2. Worldwide Production and Economy of Coffee

Coffee is one of the most important traded commodities in the world. The coffee sector's trade structure and performance have large development and poverty implications, given the high concentration of production by small-holders in poor developing countries (ICO, 2015).

Coffee's global value chains are quickly transforming because of shifts in demands and an increasing emphasis on product differentiation in importing countries (Ponte, 2002; Daviron and Ponte, 2005). There is a growing willingness-to-pay for premium; high quality coffee by rich consumers and the demand for specialty and certified coffee is on the rise. These changes have important implications for a number of the poorest developing countries, as most coffee production takes place in these countries, even though most coffee consumption is in developed countries (Ponte, 2002; Pendergrast, 2010).

Coffee is cultivated by over 4 million primarily smallholder farming households (CSA, 2013) and with those employed in ancillary activities to coffee production, even more households are dependent on coffee for part of their livelihoods. Ethiopia is also unique in Africa in so far as it has a strong domestic coffee consumption culture, which frequently accounts for over half of production (ICO, 2015).

The two coffee species currently used for commercial purposes are Arabica Coffee and Robusta Coffee. Ethiopia only produces Arabica coffee, which is widely believed to have originated there. Arabica coffee still grow in wild in the forests of the south-western part of the country, which remains an important source of genetic resources for the world coffee industry (Gole *et al*, 2002).

### 2.3. Importance of Coffee in the Ethiopian Economy

Agriculture is the main sector in the national economy of Ethiopia, employs about 85% of the population. It account for around 90% of foreign exchange and 40% of GDP. The coffee industry dominates agriculture sector in its contribution to the national economy in general and export sector in particular (Birhe, 2010). This implies that the sector is the back bone of Ethiopian Economy. Coffee in Ethiopia accounts for 25% of GNP, 40% of total export and 10% of total government revenue (MoARD, 2007). It also accounts for an average 5% of GDP, 10% of the total agriculture production and 60% of export earnings (Girma, 2011). Coffee has always been the Ethiopia's most important cash crop and largest export commodity, which account 90% of exports and 80% of total employment (CSA, 2008). In 2015, Ethiopia produces 6.4 million bags, making it the leading African coffee producer and the five largest producers in the world, and

also exported 3.2 million bags, ranked it the most important African coffee exporter and the eighth largest exporter in the world (ICO, 2015).

Coffee has thus significant impact on the socio-economic life of the people and economic development of the country. It is estimated that more than 15 million people are directly or indirectly engaged in the production, processing and trading of coffee. Coffee plays a significant role in the Ethiopian export earnings (ICO, 2013).

**2.4. Alternate Bearing in Perennial Crops (Biennial)**

Problems of alternate (biennial) bearing in perennial plants (especially in fruit trees) have been reviewed and documented (Singh, 1948; Davis, 1957; Singh, 1971; Jonkers, 1979). The term biennial bearing is used sometimes interchangeably with alternate bearing, and it is the phenomenon observed in most of perennial crops that results not bear regular crop year after year rather heavy yields are followed by extremely light ones and vice-versa (Pearce and Urbane- Urbanc, 1967). Biennial (alternate) bearing also refers to the tendency of an entire tree to produce a greater than average crop one year, and a lower than average crop the following year.

According to Bernardes *et al.* (2012), a coffee plot exhibits high and low production in alternated years, and it is a characteristic called biennial yield**.** Rodrigues *et al.* (2013), also reports that coffee plantations present large spatial and temporal variability of yield, and the variation along the years with high and low productions is known as biennially. The phenomenon biennial is more pronounced in the species coffee Arabica, but it is also present in Robusta coffee, usually less intense due to its mitigation with the pruning practices and alternation of the plagiotropic branches in production (Taye *et al*, 2011; Rodrigues *et al*., 2013; Omondi *et al*., 2016). This biennial alternation of yield is the result of the physiological nature of the coffee plant, which needs to vegetate along a year to sustain the fruit production in the next year (Davis, 1957).

The occurrence of biennially in coffee plants is connected to a source-sink relationship existing between fruit and leaves. Leaves are sources of photosynthesis while the growing tissues act as sinks. As both of the reproductive and vegetative growths occur simultaneously, the plant

needs to balance the partition of photosynthesis for both processes (Barros, 1997). In years of high production, the plant directs the photosynthesis to the formation and growth of fruits, reducing the formation of new vegetative buds. In years of low production, the photosynthesis is directed to the formation of new vegetative buds that will produce new branches. Therefore, the over production of fruits in a year causes a reduction in growth in the current year, exhausting the metabolic reserves for the fruit production. Consequently, the growth is restricted and the emission of new plagiotropic branches is limited, compromising the fruit production of the next season (Picini, 1998). Additionally, this relationship between leaf biomass and coffee yield is influenced by the occurrence of diseases, especially coffee rust (Hemileia vastatrix) (Costa *et al*., 2006). In years with high production, rust infestations are more severe resulting in high leaf fall after harvest and, consequently, it causes yield reduction the following year Therefore, the occurrence of rust in years of high yield accentuates the effect of coffee biennially (Zambolim *et al*., 2002 )

### 2.5. Difference among Clustered, Repeated Measures, Longitudinal, and Clustered longitudinal data

According to (west *et al*., 2014), clustered data is a data sets in which the dependent variable is measured once for each subject (the unit of analysis), and the units of analysis are grouped into or nested within clusters of units. For example, the math scores of students (the units of analysis) nested within classrooms (clusters of units), which are in turn nested within schools (clusters of clusters). This type of data set is defined to be three-level clustered data set. Repeated measures data quite generally as data sets in which the dependent variable is measured more than once on the same unit of analysis across levels of a repeated-measures factor (or factors). The repeated-measures factors, which may be time or other experimental or observational conditions, are often referred to as within-subject factors.

Longitudinal data is data sets in which the dependent variable is measured at several points in time for each unit of analysis. We usually conceptualize longitudinal data as involving at least two repeated measurements made over a relatively long period of time. In some cases, when the dependent variable is measured over time, it may be difficult to classify data sets as either longitudinal or repeated-measures data. In the context of analyzing data using LMMs, this

distinction is not critical. The important feature of both of these types of data is that the dependent variable is measured more than once for each unit of analysis, with the repeated measures likely to be correlated. Clustered longitudinal data sets combine features of both clustered and longitudinal data. More specifically, the units of analysis are nested within clusters, and each unit is measured more than once. Generally clustered, repeated-measures, and longitudinal data are hierarchical data sets, because the observations can be placed into levels of a hierarchy in the data (West *et al*., 2014).

### 2.6. The Concept of Levels in Data Structure

The concept of "levels" of data is based on ideas from the hierarchical linear modeling (HLM) literature (Raudenbush and Bryk, 2002). All data sets appropriate for an analysis using LMMs have at least two levels. Very often data sets in HLM may exist as a two-level or three-level structure, depending on how many levels of data are present. Level 1 denotes observations at the most detailed level of the data. In a clustered data set, Level 1 represents the units of analysis (or subjects) in the study. In a repeated-measures or longitudinal data set, Level 1 represents the repeated measures made on the same unit of analysis. The continuous dependent variable is always measured at Level 1 of the data. Level 2 represents the next level of the hierarchy. In clustered data sets, Level 2 observations represent clusters of units. In repeated-measures and longitudinal data sets, Level 2 represents the units of analysis. Level 3 represents the next level of the hierarchy, and generally refers to clusters of units in clustered longitudinal data sets, or clusters of Level 2 units (clusters of clusters) in three-level clustered data sets. Continuous and categorical variables can be considered at different levels of the data, and these variables refer to the variables as Level 1, Level 2, or Level 3 variables.

### 2.7. Models for Gaussian Longitudinal Data

The analysis of change is a fundamental component of so many research endeavors in almost every discipline. Many of the earliest statistical methods for the analysis of change were based on the analysis of variance (ANOVA) paradigm, as originally developed by R. A. Fisher. Before it was put on a more formal theoretical footing in the seminal work of R. A. Fisher, Airy (1861) laid the foundations for the linear mixed-model in a longitudinal study.

Scheff´e (1956) provides a fascinating discussion of the early contributions to the development of the theory of random-effects models. As such, it can be argued that statistical methods for the analysis of longitudinal data, in common with classical linear regression and the method of least squares have their earliest origins. On the one hand, the univariate repeated-measures ANOVA model provided a natural generalization of Student's (1908) paired *t*-test to handle more than two repeated measurements, A special case of the repeated-measures analysis by MANOVA is a general approach known as profile analysis (Box, 1950; Geisser and Greenhouse, 1958; Greenhouse and Geisser, 1959). It proceeds by constructing a set of derived variables, based on a linear combination of the original sequence of repeated measures, and using relevant subsets of these to address questions about longitudinal change and its relation to between-subject factors.The linear mixed-effects model is the most widely used method for analyzing longitudinal data. Although the early development of mixed-effects models for hierarchical or clustered data can be traced back to the ANOVA paradigm (Scheff´e, 1959) and to the seminal paper by Harville (1977), their usefulness for analyzing longitudinal data, especially in the life sciences, was highlighted by Laird and Ware (1982). Laird and Ware (1982), drawing upon a general class of mixed models introduced earlier by Harville (1977), proposed a flexible class of linear mixed-effects models for longitudinal data. These models could handle the complications of mistimed and incomplete measurements in a very natural way the.

The linear mixed-effects model proposed by Laird and Ware (1982) included the univariate repeated-measures ANOVA and growth curve models for longitudinal data as special cases. In addition, the Laird and Ware (1982) formulation of the model had two desirable features: first, there were fewer restrictions on the design matrices for the fixed and random effects; second, the model parameters could be estimated efficiently via likelihood based methods. To estimate parameters, Jennrich and Schluchter (1986) proposed a variety of alternative algorithms, including Fisher scoring and Newton–Raphson. Currently, maximum likelihood and restricted maximum likelihood estimation, the latter devised to diminish the small-sample bias of maximum likelihood, are the most frequently employed routes for estimation and inference (Verbeke and Molenberghs, 2000; Fitzmaurice, Laird, and Ware, 2004).

## 2.8. The Concept of Random Factor, Random Effect, Fixed Factor and Fixed Effect

Here we need to describe and define concepts of fixed and random effects that are applied in mixed effects model. The distinction between fixed and random factors and their related effects on a dependent variable are critical in the context of LMMs. According to West *et al*. (2014), a fixed factor is a categorical or classification variable, for which the investigator has included all levels (or conditions) that are of interest in the study. Levels of a fixed factor are chosen so that they represent specific conditions, and they can be used to define contrasts (or sets of contrasts) of interest in the research study. A random factor is a classification variable with levels that can be thought of as being randomly sampled from a population of levels being studied. All possible levels of the random factor are not present in the data set, but it is the researcher's intention to make inferences about the entire population of levels.

Thus, West *et al*. (2014) also describe that the classification variables that identify the Level 2 and Level 3 units in both clustered and repeated-measures/longitudinal data sets are often considered to be random factors. Random factors are considered in an analysis so that variation in the dependent variable across levels of the random factors can be assessed, and the results of the data analysis can be generalized to a greater population of levels of the random factor. In contrast to the levels of fixed factors, the levels of random factors do not represent conditions chosen specifically to meet the objectives of the study. However, depending on the goals of the study, the same factor may be considered either as a fixed factor or a random factor.

According to Fitzmaurice (2009), fixed effects, called regression coefficients or fixed-effect parameters, describe the relationships between the dependent variable and predictor variables (i.e., fixed factors or continuous covariates) for an entire population of units of analysis, or for a relatively small number of subpopulations defined by levels of a fixed factor. Fixed effects may describe contrasts or differences between levels of a fixed factor in terms of mean responses for the continuous dependent variable, or they may describe the effect of a continuous covariate on the dependent variable. Fixed effects are assumed to be unknown fixed quantities in an LMM, and we estimate them based on our analysis of the data collected in a given research study. Random effects are random values associated with the levels of a random factor (or factors) in an LMM.

## 2.9. Genotype by Environment Interaction

GEI is a large and complex phenomenon. However, it is an extremely important occurrence that has strong ecological, evolutionary, and commercial implications. Studying GEI not only provides models for quantitative analyses but will lead to improved breeding in both cultivated (agriculture) and natural (conservation) populations (Crossa, 1990; Magari and Kang, 1993; Basford and Cooper, 1998). The phenotype characteristic of an organism is always determined by both the genotype and the environment (Boughey, 1973). However, these two effects are not always additive which indicates that genotype by environment interactions (GEI) is present. The GEI result in inconsistent performances between the genotypes across environments. Significant GEI results from the changes in the magnitude of differences between genotypes in different environments or changes in the relative ranking of the genotypes (Falconer, 1952; Fernandez, 1991)

Genotype by environment interaction (GEI) can be defined as the differential response of varying genotypes under change(s) in the environment (Mather and Caligari, 1976). It is an important aspect in both plant breeding programs and the introduction of new crop, and can occur when specified genotypes are grown across diverse environments (Zobel, 1990). Genotypes are assumed by observing differential effects on their expression. This implies that the most popular method of determining GEI is by studying the resulting phenotypes under the influence of the environment. However, Mather and Caligari (1976) suggest that because variation in a character may result from variation in either genotype or environment, heritable and non-heritable character variation cannot be determined by only inspecting the phenotypes. It is important to know the environment of an organism and its genetic history. Common environmental factors in GEI studies include locations, growing seasons, years, rainfall, the amount of precipitation received in each season, temperature, etc which may have positive or negative impact on genotypes (Dean, 1995).

Genotypic variation originates from differences in the genome of different individuals whereas phenotypic variation occurs when individuals are exposed to different environmental parameters during the development of similar genomes. In phenotypic variation, individuals adapt in response to specific environmental changes. The association between the environment and the

phenotypic expression of a genotype constitute the GEI, and it has profound implications on the evolution of species (Reano, 2010, as cited in Zelalem, 2011). Lande and Shannon (1996) suggest that in constant or unpredictable environments, genetic variance reduces population mean fitness and increases the risk of extinction. In the short-term, genetic variability is often less critical than other determinants of population persistence (Lande, 1988). But over time, it can play the decisive role in allowing a population to persist and adapt in a changing environment (Lande and Shannon, 1996).

**2.10. The Concept of Stability**

The yield stability is one of the most desirable properties of a genotype to be released as a variety for cultivation. Stability is a complex product of genetic yield potential to stress conditions. The yield stability is influenced by several factors, such as environmental factors, agricultural managements and pest pressures (Hu and Buyanovsky, 2003; Berzsenyi and Dang, 2008). Breeding genotypes that are adapted throughout a reasonable large geographical area and that show some degree of stability from year to year is a major problem facing plant breeders. As a result, several methods of measuring and describing genotypic response across environments have been developed a utilized. For this purpose, multi-locational trials, over a number of years are conducted (Luthra *et al*, 1974). The level of performance of any character is a result of the genotype (G) of the cultivar, the environment in which it is grown (E), and the interaction between G and E (GEI).

Genotype x environment interaction (GEI) exists when the responses of two genotypes to different levels of environmental stress are not consistent (Allard and Bradshaw, 1964). GEI greatly affect the phenotype of a variety, so the stability analysis is required to characterize the performance of varieties in different environments, to help plant breeders in selecting varieties. Instability is the result of cultivars response in different environments which usually indicates a high interaction between genetic and environmental factors (Jusuf *et al.*, 2008; Lone *et al.*, 2009). Grain yield depends on genotype, environment and management practices and their interaction with each other. Under the same management conditions, variation in grain yield is principally explained by the effects of genotype and environment (Luquet *et al.*, 2006). Interaction between these two explanatory variables gives insight for identifying genotype

suitable for specific environments. The environmental effect is typically a large contributor to total variation (Blanche *et al.*, 2009). The analyses of genotype x environment has focused on the identification of stable genotype for cultivation.

## 2.11. Related Empirical Literature Review

### 2.11.1. Longitudinal study

Longitudinal studies play a prominent role in various endeavors including agricultural sciences. They are indispensable to the study of change in an outcome over time. By measuring study participants repeatedly through time, longitudinal studies allow the direct study of temporal changes within individuals and the factors that influence change (Fitzmaurize *et al*, 2008). The repeated measurements taken over time give rise to a complex random error variance covariance structure which needs especial concern in the analysis (piepho *et al*, 2014). Thus, multi location trials conducted over year for perennial crops give rise to such data that taken on the same plot on several occasions.

The statistical methods that employed to analyze longitudinal data initially were based on the analysis of variance (ANOVA) paradigm, as originally developed by R. A. Fisher. There has been a remarkable advancement in statistical methodology basing form ANOVA till the repeated measures analysis by MANOVA. However, while these methods can provide a reasonable basis for a longitudinal data analysis in cases where the study design is complete and quite simple, they have many shortcomings that have limited their usefulness in applications (Fitzmaurize *et al*, 2008). Finally, Laird and Ware (1982) proposed a linear mixed model for longitudinal data that could handle issues of unbalanced data, due to either mistimed measurement or missing data, could handle both time-varying and time-invariant covariates, and provided a flexible, yet parsimonious, model for the covariance.

Some recent empirical works have used the mixed effect model to exploit the correlation among repeated measurements of perennial crops that conducted in multi location over year. Thus, Piepho and Eckl(2014) conducted a longitudinal analysis on a perennial ryegrass varieties with annual yields recorded per plot for three consecutive years in southern Germany. Their main object was to account for the correlation among repeated measurements in such trials.

Accordingly, they found a significance serial correlation that can be adequately captured by fitting autoregressive order (AR1) model. In Ivory Coast, Cilas *et al.* (2011) have also conducted a longitudinal analysis on Robusta coffee to estimate the correlation among measurements of coffee yield over year. Regardless of biennial and time effect, they quantified this correlation using *Compound Symmetry* model. In similar way, in Brazil, de Resende (2006) investigated the correlation among yield measurements of tea plant using ARH model. These authors also pointed out that it needs new modeling approach not only for longitudinal correlation but also both special and longitudinal correlations simultaneously.

However, these studies aimed at only the comparison of several models for correlation among measurements of perennial crop, and the effect of time and biennial was not considered. There was no clear published literature relating to longitudinal analysis on yields of Arabica Coffee in the linear mixed model setting including time variant factor biennial to investigate the effect biennial and trend of coffee yield. But in Brazil, Rodriguez *et al.* (2013) investigated only the effect of biennial on the genotypes of Robusta coffee. by calculating the magnitude of biennial (i.e., by subtracting the mean production of the years of low production from the mean of the years of high production based on an even number of years). The result showed high yield variation between years of high and low productions and variation among genotypes on their calculated biennial means.

### 2.11.2. Genotype by Environment Interaction Study (GEI)

Genotype by environment interaction (GEI) is an important phenomenon, and it is a deferential response of genotypes across environments (often location–by–year combinations) (Gauch, 2013, Rodrigues *et al*., 2014)). This phenomenon is an enormous importance to plant breeders, production agronomists, biometricians and other agricultural experts. This is due to the fact that the presence of GEI determines if a genotype is widely adapted for an entire range of environmental conditions or separate genotypes must be selected for different sub environments.

Since GEI is large and complex, and it is an extremely important occurrence that has strong ecological, evolutionary phenomenon, it has got special attention in the theory and application of statistical models (Crossa, 1990; Girma *et al*., 2000). In the beginning, Eberhart & Russell

(1966) and Finlay & Wilkinson (1963) have investigated GEI through joint analysis of variance (ANOVA) and linear regression techniques. Regardless of their limitation in the assumption of conventional ANOVA and linear regression, the method has been widely used for log time (Becker and Leon, 1988, as cited in Degene, 2016). Even thought alternatives have been made by Cruz *et al*. (1989) and Toler & Burrows (1998) for these limitations, it yet again open to another criticism due to the fact that the GEI component has been estimated but not decomposed into structures (paterns)( de Resende, 2006).

Gauch (1988; 1992) attempt to avoid these limitations by describing the technique called AMMI (Additive Main Effects and Multiplicative Interaction Analysis) which was attributed to the work of Mackenzie (1923) and Gollob (1968). AMMI have been popularized in a fixed model context and a number of applications have been developed (Gauch, 1988; 1992; Crosse *et al.,* 1990). AMMI analysis combines, in a model, additive components for main effects (treatments and environments) and multiplicative components for GEI effects. It combines a univariate technique (ANOVA) for the main effects and a multivariate technique (PCA-principal component analysis) for GEI effects. Crossa (1990) suggests that the use of multivariate techniques permits a better use of information than the traditional regression methods.

The proposed models (ANOVA, linear regression, and AMMI) were functioning in the linear fixed model setting. However, families of linear mixed model have being wide used by researchers for GEI study due to the fact that they are flexible and powerful especially in the violation of the classical linear fixed model assumptions and in the area of incomplete multi environment data (Smith *et al*., 2001; Girma *et al*., 2014). According to de Resende (2006), the method of analysis in the mixed model setting encompass the Piepho (1998) factor analytic multiplicative mixed (FAMM) model with random genotype and GEI effects which is conceptually and functionally better than AMMI and Smith *et al*. (2001) which is a general class of FAMM models that encompass the approach of Piepho (1998) and include separate spatial errors for each environment (FAMMS).

On multi environment crop trial data, several recent empirical investigations have been done using those proposed statistical methods to exploit the information contained in the GEI. Girma et *al.* (2014) conducted GEI analysis on the data sets of Intermediate to Late Hybrid Trails (ILHT) conducted in five Eastern and Central Africa (ECA) countries from 2008 to 2011 under different management by fitting and compare linear bilinear models, and they showed that Site Regression (SREG2), Genotypic Regression (GREG2) and Factor Analytic FA(1) are preferred models to identify stable genotype. In Ethiopian context, by using grain yield of 36 field pea genotypes planted in four locations over two years, Girma *et al*. (2000) showed the presence of significance GEI in field pea grain yield , and did AMMI adjustment in modeling the effect of GEI to increase accuracy for yield estimate and classification of genotypes and environments.

Specifically, in Ethiopia, Yonas *et al.* (2014a) conducted stability analysis on 30 genotypes of Arabica Coffee across 8 environments using Eberhart & Russell regression and AMMI method. The investigators reported that the mean squares of genotypes, environments and genotype by environment interaction were highly significant, and the first IPCA axis alone accounted for 36% of the total GEI sum of squares and it was 50% higher in contrast to the interaction accounted by the Eberhart & Russell regression method. The investigators also put their condemnation on the drawback of Eberhart & Russell regression method by their second work Yonas *et al*. (2014b), and they reported that analyzing stability of performance of Arabica coffee variety using Eberhart and Russell stability model which was actually not bad for annual crops leads to invalid results and wrong conclusions. Similarly, Meaza *et al*. (2011) have made GEI analysis on 43 genotypes of Arabica coffee across 8 environments, and it was reported that at 1%, the combined ANOVA showed significant mean squares for genotype, environment, GEI, and four IPCAs. The investigators also reported that the first two principal components explained about 74 percent of the GEI interaction component. Yonas and Tarekegn (2015) and Lemi and Ashenafi (2016)) who reported genetic variation and heritability of various traits in Arabica coffee genotypes also reported that the major factor that influence yield performance of Arabica coffee genotypes is the environment.

# 3. DATA AND METHODS

## 3.1. Data Source

The data for this study came from coffee variety field trials conducted by Jimma Agricultural Research Center (JARC) of Ethiopian Institute of Agricultural Research (EIAR) over several years. The center is located in southwestern Ethiopia and has a national mandate for coffee research in the country, and serves as the center of excellence for coffee research in Ethiopia. It serves for the sustainable production of Arabica coffee in the country by releasing improved disease resistant, high yielder and quality coffee varieties. The field trial was conducted in three locations (Jimma, Agaro and Metu) of southwest Ethiopia. These locations have different soil type and altitudes and could also possibly be differentiated with their mean seasonal rainfall and temperature. Seven year CBY data collected during 2005 to 2011 were used in this study. These observations were obtained from a total of 204 coffee trees with 7 observations per coffee tree (over 7 years period).

### 3.1.1. Experimental design and trial managements

The trials consisted of 17 Arabica coffee genotypes. They were selected for their high potential resistance to CBD, yield and cup quality during a preliminary evaluation. Primarily they were collected from different farmer's field of south-western region of the country along with quite large numbers of coffee accessions. The seedlings were planted on the field when they are approximately ten months old in randomized complete block design with 4 replications. Each plot consisted of ten trees in a single row. The spacing between rows and trees within a row was 2m by 2m, respectively. The plots received uniform application of fertilizer and other cultural practices throughout the period of data collection. All coffee trees were maintained on a single stem pruning system. Yield was recorded in fresh cherry in gram and converted to clean coffee bean yield in kilogram per hectare using a conversion factor. The measurement of coffee tree or plot used in the analysis was the clean average coffee bean yield of 10 coffee trees per plot.

### 3.1.2. Type and structure of variables used for longitudinal study

For longitudinal study, the type of the data set were considered as clustered longitudinal data in which subjects/coffee trees nested in clusters of block. Thus, two ID variables/grouping factors

(Block and Coffee tree) were used in this study. Therefore, the structures of variables included in the longitudinal analysis were as follows.

***Block (Level 3) Variables***

Block ($ID_2$) =block ID number (random factor)

Location = the environment where coffee grown (fixed factor)

***Coffee trees (Level 2) Variables***

Coffee tree ($ID_1$) = coffee tree ID number nested in block (random factor)

Genotype = genetically different types of coffee (fixed factor) (G1 (*Dessu*) =0 is the Reference genotype)

***Time-Varying (Level 1) Variables***

Time = Time points of longitudinal measures (1 = 1 year, 2 = 2 year…..7=7 year)

Biennial = alternating year (0=years at two year interval (even years); 1=the other years)

CBY= yield of coffee tree in kilogram per hectare ($kgha^{-1}$) (response or dependent variable)

### 3.1.3. Data setting for GIE analysis

The Independent variables used for GIE analysis were Block, Genotype and the Environment where the Environment is a specific year-location combination, whereas CBY was the dependent (study) variable. The location-year combinations and the assigned Environment and Genotype code are given in Table 1 and Table 2.

Table 1: Brief summary of Environments and assigned code

| Location(Year) | Environment Code | Location(Year) | Environment Code |
|---|---|---|---|
| Agaro(2005) | E1 | Metu(2009) | E12 |
| Agaro(2006) | E2 | Metu(2010) | E13 |
| Agaro(2007) | E3 | Metu(2011) | E14 |
| Agaro(2008) | E4 | Jimma(2005) | E15 |
| Agaro(2009) | E5 | Jimma(2006) | E16 |
| Agaro(2010) | E6 | Jimma(2007) | E17 |
| Agaro(2011) | E7 | Jimma(2008) | E18 |
| Mutu(2005) | E8 | Jimma(2009) | E19 |
| Metu(2006) | E9 | Jimma(2010) | E20 |
| Metu(2007) | E10 | Jimma(2011) | E21 |
| Metu(2008) | E11 | | |

Table 2: Brief summary of Arabica coffee Genotype and assigned code

| Genotype Name | Genotype Code | Genotype Name | Genotype Code |
|---|---|---|---|
| Dessu(check) | G1 | 39/77 | G10 |
| 744(check) | G2 | 39/82 | G11 |
| 21/81A | G3 | 4/84 | G12 |
| 235/71A | G4 | 43/70 | G13 |
| 29/82 | G5 | 5/81 | G14 |
| 3/77 | G6 | 51/'84 | G15 |
| 32/82 | G7 | 64/84 | G16 |
| 36/82 | G8 | 20/81 | G17 |
| 38/82 | G9 | | |

## 3.2. Methods of Data Analysis for Longitudinal Study

### 3.2.1 Exploratory data analysis

Exploratory data analysis comprises techniques to visualize patterns in the data. Data analysis must begin by making displays that expose patterns relevant to the scientific question. The best methods are capable of uncovering patterns which are unexpected. Most longitudinal studies address the relationship of a response with explanatory variables, often including time. Exploratory data analysis explores the individual profile, the average evolution, the variance function, and the correlation structure of the data. Data exploration is a very helpful tool in the selection of appropriate models. The average evolution describes how the profile of a number of relevant sub-populations (or the population as a whole) evolves over time. The results of such exploration will be useful in order to choose a fixed-effects structure for linear mixed model.

In addition to the average evolution, exploring the evolution of the variance is important to build an appropriate longitudinal model. This could be done by plotting the variances of sub groups with each time points. It helps to check whether the assumption of constant variance is fulfilled. Explanatory data analysis is also very useful to explore the correlation structure between measurements. The correlation structure describes how measurements within a subject correlate. A different way of exploring the correlation structure is using a scatter plot matrix and correlation matrix of the observed data. The decay of correlation with time is studied by considering the evolution of the scatter with increasing distance to the main diagonal. Stationary in this case implies that the scatter plots remain similar within diagonal bands if measurement

occasions are approximately equally spaced. In addition to the scatter plot the histogram on the diagonal also useful to capture the variance structure, including such as skewness. Hence, these possible graphical tools were used in this study.

### 3.2.2. Linear mixed model

Laird and Ware (1982) proposed a flexible class of linear mixed-effects model that could handle the complications of mistimed and incomplete measurements in a very natural way. Suppose we sampled i = 1, 2, …, k independent units each with t = 1, …, $n_i$ repeated measurements. The linear mixed-effects model is given by.

$$\begin{cases} Y_i = X_i\beta + Z_i b_i + \epsilon_i \\ b_i = MVN(0, D), \\ \epsilon_i = MVN(0, R_i) \\ b_i \dots\dots b_k, \epsilon_i \dots\dots \epsilon_k, \text{independent} \end{cases} \quad [1]$$

Where

$Y_i$ is the $ni$-dimensional response vector for subject $i, 1 \leq i \leq k, k$ is the number of subjects, $X_i$ and $Z_i$ are $(n_i\, x\, p)$ and $(n_i\, x\, q)$ dimensional matrix of known covariates, $\beta$ is a $p$-dimensional vector containing the fixed effects, $b_i$ is the q-dimensional vector containing the random effects, and $\epsilon_i$ is $n_i$ –dimensional vector of residual components. Finally, $D$ is a general $(q\ x\ q)$ covariance matrix of random effects and $R_i$ is $(n_i\, X\, n_i)$ covariance matrix random error which depends on $i$ only through its dimension $n_i$, i.e. the set of unknown parameters in $R_i$ will not depend on i. The distributional assumptions of this model imply that

$$Y_i \sim MVN(X_i\beta, V_i) \quad \text{where} \quad V_i = Z_i D Z_i' + R_i)$$

The multilevel model can be written in terms of the LMM model. This point can be best illustrated by constructing a model for our cluster longitudinal CBY dataset. Recall that this data set has three levels of hierarchy: repeated measurements of CBY taken over year on the same coffee tree which is nested within block. The indices account for the clustering in the dataset where j = 1, . . . , m used for blocks, i=1, . . . ,k for coffee tree nested in block and t=1, . . . , n for repeated measurements of CBY taken over time. Though only the balanced case is presented here, note that these models can easily accommodate unbalanced data on any level of the hierarchy. Thus, according to Modur (2010), a linear mixed effect model for cluster longitudinal CBY data set given as

$$\begin{cases} Y_{ji} = X_{ji}\beta + Z^{M}{}_{ji}M_i + Z^{C}{}_{ji}C_{ji} + \epsilon_{ij} \\ M_i \sim MVN(0, D_M), \\ C_{ji} \sim MVN(0, D_C) \\ \epsilon_{ji} = MVN(0, R_{ji}) \end{cases} \quad [2]$$

$X_{ji}$ is a n x p design matrix with covariates defined at different levels. The design matrices for both the block level random effects ($M_i$ ) and coffee tree level random effects ($C_{ji}$ ) are denoted by $Z^{M}{}_{ji}$ and $Z^{C}{}_{ji}$, respectively. The random effects design matrices are formed from a subset of the appropriate columns of $X_{ji}$ . These matrices can contain covariates that vary at lower levels of the hierarchy. The model assumptions here pertain to the sources of variability. The random effects at the same level are correlated within units at that level. Random effects at different levels are assumed to be independent of each other. In other words, all components of the block level random effects vector ($M_i$ ) are allowed to be correlated with each other. This covariance will be captured by the off diagonal components of the covariance matrix $D_M$ . The same applies for the coffee tree level random effects vector, $C_{ji}$. The vectors $M_i$ , $C_{ji}$ , and $\epsilon_{ji}$ , are assumed to be independent of each other.

If we rewrite $Z_{ji} = [Z^{M}{}_{ji}|Z^{C}{}_{ji}]$ and $b_{ji} = (M_i{}^{t}\ C_{ji}{}^{t})^{t}$ then model 2 can be represented as follows:

$$\begin{cases} Y_{ji} = X_{ji}\beta + Z_{ij}b_{ji} + \epsilon_{ji} \\ b_{ji} = (\mathrm{M_i}^{\mathrm{t}}\ \mathrm{C_{ji}}^{\mathrm{t}})^{\mathrm{t}} \sim \mathrm{MVN}(0, \mathrm{D} = \mathrm{D_M} \oplus \mathrm{D_C}) \\ \epsilon_{ji} = MVN(0, R_{ji}) \end{cases} \quad [3]$$

**5.2.3. Variance functions for modeling heteroscedasticity**

Variance functions are used to model the variance structure of the within group errors using covariates. They have been studied in detail in the context of mixed effects models by Davidian and Giltinan (1995) and in the context of the extended linear model by Carroll and Ruppert (1988). Acoording to Davidian and Giltinan (1995), the general variance function model for the within-group errors in the extended linear mixed effects model given as

$$var(\epsilon_{jit}|b_{ij}) = \sigma^2 g^2(\mu_{jit}, v_{jit}, \delta) \quad [4]$$

Where $\mu_{jit}$= E ($y_{jit}$ |$b_{ji}$), $v_{jit}$ is a vector of variance covariates, δ is a vector of variance parameters and g( · ) is the variance function, assumed continuous in δ . For example, if the within-group variability is believed to increase with some power of the absolute value of a covariate $v_{jit}$, we can write the variance model as $var(\epsilon_{jit}|b_{ji}) = \sigma^2(|v_{jit}|^{2\delta})$ . Table 3 shows the most standard variance function classes which are built in R computing statistical package

Table 3: Standard variance function classes

| Name | expression |
|---|---|
| varFixed | fixed variance |
| varIdent different | variances per stratum |
| varPower | power of covariate |
| varExp | exponential of covariate |

### 3.2.4. Correlation functions for modeling dependency

Correlation structures are used to model dependence among observations. In the context of mixed effects models and extended linear models, they are used to model dependence among the within group errors. Historically, correlation structures have been developed for two main classes of data: time-series data and spatial data. The former is generally associated with observations indexed by an integer-valued time variable, while the latter refers primarily to observations indexed by a two-dimensional spatial location vector, taking values in the real plane. The general within group correlation structure for two-level grouping is expressed.

$$cor(\epsilon_{jit}, \epsilon_{jit'}) = h(d(p_{jit}, p_{jit'}), \rho) , \quad [3]$$

Where ρ is a vector of correlation parameters and h (·) is a correlation function taking values between − 1 and 1, assumed continuous in ρ, and such that h (0, ρ) = 1, that is, if two observations have identical position vectors, they are the same observation and therefore have correlation 1. Table 4 shows the most common standard correlation function classes that are built in R computing program.

Table 4: Standard correlation function classes

| Name | Expression |
|---|---|
| corCompSymm | compound symmetry |
| corSymm | General (unstructured) |
| corAR1 | autoregressive of order 1 |
| corAR(p) | Autoregressive of order p (p>1) |
| corExp | exponential |
| corGaus | Gaussian |
| corLin | linear |
| corRatio | rational quadratic |
| corSpher | spherical |

### 3.2.5. Method of parameter estimations

Maximum likelihood estimation and restricted maximum likelihood estimation are the two commonly used methods of estimations in linear mixed model. Maximum likelihood estimation and restricted maximum likelihood estimation both have the same merits of being based on the likelihood principle which leads to useful properties such as consistency, asymptotic normality, and efficiency. ME estimation also provides estimators of the fixed effects, whereas REML estimation, in its self, does not. On other hand, for balanced mixed ANOVA models, the REML estimates for the variance component are identical to classical ANOVA type estimates obtained from solving the equations which set mean squares equal to their expectations. This implies optimal minimum variance properties, and it shows that REML estimates in that context do not rely on any normality assumptions since any moment assumptions are involved (harville, 1977 and Searle, casella and McCulloch,1992). REML corrects for the downward bias in the ML parameters in D and R , and handles strong correlations among the responses more effectively. The differences between ML and REML estimation increase as the number of fixed effects in the model increases. There is also the non-iterative MIVQUE0 method, which performs minimum variance quadratic unbiased estimation of the covariance parameters. However simulation evidence favors REML and ML over MIVQUE0.

First consider parameter estimation in model 1, and then it follows for model 3 in the same way. Let $\alpha$ denote the vector of all variance covariance parameters (usually called variance component) found in $V_i = Z_i D Z_i^{'} + R_i$, that is, $\alpha$ consists of the $q(q+1)/2$ different elements in

D and of all parameters in $R_i$. Finally, let $\theta = (\beta', \alpha')'$ be the s-dimensional vector of all parameters in the mariginal model for $Y_i$ and let $\theta = \theta_\beta X \theta_\alpha$ denote the parameters for $\theta$, with $\theta_\beta \; and \;\; \theta_\alpha$ the parameters space for the fixed effects and for the variance components respectively. The classical approach to inference is based on estimations obtained from maximizing the marginal likelihood function

$$L_{ML}(\theta) = \prod_{i=1}^{n} \left\{ (2\pi)^{-\frac{n_i}{2}} |V_i(\alpha)|^{-\frac{1}{2}} \; x \; exp(-1/2(Y_i - X_i\beta)' V_i^{-1}(\alpha)(Y_i - X_i\beta)) \right\} \qquad [4]$$

With respect to $\theta$. Let as first assume $\alpha$ to be known. The maximum likelihood estimator (MLE) of $\beta$, obtained from maximizing 4, conditional on $\alpha$, is then given by(Laird and Ware 1982).

$$\hat{\beta}(\alpha) = \left( \sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \sum_{i=1}^{N} X_i' W_i X_i, \text{ where } W_i \text{ equals } V_i^{-1} \qquad [5]$$

When $\alpha$ is not known, but an estimator $\hat{\alpha}$ is available, we can set $\widehat{V_i} = V_i(\hat{\alpha}) = \widehat{w_i}^{-1}$, and estimate $\beta$ by using the expression 4 in which $w_i$ is replace by $\widehat{w_i}$.

The Maximum likelihood estimator (MLE) of $\alpha$ is obtained by maximizing 4 with respect to $\alpha$, after $\beta$ is replaced by expression 5

For variance estimation in normal distribution, Consider a sample of N obviations $\boldsymbol{Y_i} \ldots \ldots, \boldsymbol{Y_N}$ from$\boldsymbol{N(\mu, \sigma^2)}$, for known $\boldsymbol{\mu}$, MLE of $\boldsymbol{\sigma^2}$ equals $\boldsymbol{\widehat{\sigma}^2 = \sum_{i=1}^{n} (Y_i - \mu)^2 / N}$. Thus $\boldsymbol{\widehat{\sigma}^2}$ unbiased estimator for $\boldsymbol{\sigma^2}$. When $\boldsymbol{\mu}$ is not known, the MLE of $\boldsymbol{\sigma^2}$ equals $\boldsymbol{\widehat{\sigma}^2 = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 / N}$. In this case $\boldsymbol{\widehat{\sigma}^2}$ is biased for $\boldsymbol{\sigma^2}$ since $\boldsymbol{E(\widehat{\sigma}^2) = \frac{N-1}{N} \sigma^2}$. The biased expression tells us to derive an unbiased estimate

$$S^2 = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 / (N - 1).$$

Apparently, having to estimate μ introduces bias in MLE of $\sigma^2$. To estimate $\sigma^2$ without estimating μ , we transform Y such that μ vanishes from the likelihood.

$$U = \begin{bmatrix} Y_1 - Y_2 \\ Y_2 - Y_3 \\ \vdots \\ Y_{N-2} - Y_{N-1} \\ Y_{N-1} - Y_N \end{bmatrix} = A'Y \sim N(0, \sigma^2 A'A) \qquad [6]$$

The MLE of $\sigma^2$, based on U, equals $S^2 = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 / (N - 1)$. A defines a set of $N - 1$ linearly independent error contrasts and $S^2$ is called REM estimator of $\sigma^2$, and $S^2$ is independent of A.

Also for the estimation of residual variance in linear regression model, consider a sample of observations $\boldsymbol{Y_i} \ldots \ldots, \boldsymbol{Y_N}$ from a linear regression model:

$$U = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} \sim N(X\beta, \sigma^2 I) \qquad [7]$$

The maximum likelihood estimator of $\sigma^2$ is $\hat{\sigma}^2 = \left(Y - X\hat{\beta}\right)'(Y - X\hat{\beta})/N$. Thus, $\hat{\sigma}^2$ is biased for $\hat{\sigma}^2$ since$E\left(\hat{\sigma}^2\right) = \frac{N-p}{N}\sigma^2$. The bias expression tells us how to derive an unbiased estimator. MSE=$\left(Y - X\hat{\beta}\right)'(Y - X\hat{\beta})/(N - p)$ can also be obtained from transforming the data orthogonal to columns of the design matrixX.

$$U = A'Y \sim N(0, \sigma^2 A'A) \qquad [8]$$

The MLE of $\sigma^2$, based on U,now equals the mean square error (MSE). The MSE is again called REME estimate of $\sigma^2$

Again, for the estimation of REML for the Linear Mixed Model, we first combine all models $\boldsymbol{Y_i} \sim \boldsymbol{N(X_i B, V_i)}$ into one model $\boldsymbol{Y} \sim \boldsymbol{N(XB, V)}$ in which

$$Y = \begin{bmatrix} Y_i \\ \vdots \\ Y_n \end{bmatrix}, \qquad X = \begin{bmatrix} X_i \\ \vdots \\ X_N \end{bmatrix}, \qquad V(\alpha) = \begin{bmatrix} V_i & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_i \end{bmatrix}.$$

Again the data are transformed orthogonal to X. $U = A'Y = N(0, AV(\alpha)A$. Thus the MLE based on U is called the RELM estimator, and is denoted by $\hat{\alpha}_{REML}$. The resulting estimator $\hat{\beta}(\hat{\alpha}_{REML})$ for $\beta$ will be denoted by $\hat{\beta}_{REML}$. $\hat{\alpha}_{REML}$ and $\hat{\beta}_{REML}$ can also be obtained from maximizing $L_{REML}(\theta) = \left|\sum_{i=1}^{N} X_i W_I(\alpha) X_i\right|^{-\frac{1}{2}} L_{ME}(\theta)$ with respect to $\theta$, i.e. with respect to $\alpha$ and $\beta$ simultaneously.

So far, we have seen methods of parameter estimation in linear mixed model. Usually, one is primarily interested in drawing inference on the parameter in the model in order to generalize results obtained from a specific sample to a general population from which the sample was taken. Commonly used statistical tests to make inference for fixed effect in LMM are T-test, F-test, Wald test, and LR test. Similarly, Wald and LR tests are used to make inference for variance components.

### 3.2.6 Model selection

LRTs are a class of tests that are based on comparing the values of likelihood functions for two models (i.e., the nested and reference models) defining a hypothesis being tested. LRTs can be employed to test hypotheses about covariance parameters or fixed-effect parameters in the context of LMMs. In general, LRTs require that both the nested (null hypothesis) model and reference model corresponding to a specified hypothesis are fitted to the same subset of the data. The LRT statistic is calculated by subtracting −2 times the log-likelihood for the reference model from that for the nested model, as shown in the following equation:

$$LRT - 2\log\left(\frac{L_{nested}}{Lreference}\right) = -2\log(L_{nested}) - (-2\log(L_{reference}))\sim {x^2}_{df} \quad [9]$$

In Equation 9, $L_{nested}$ refers to the value of the likelihood function evaluated at the ML or REML estimates of the parameters in the nested model, and $L_{reference}$ refers to the value of the likelihood function in the reference model. Likelihood theory states that under mild regularity conditions the LRT statistic asymptotically follows a $x^2$ distribution, in which the number of degrees of freedom, *df*, is obtained by subtracting the number of parameters in the nested model from the number of parameters in the reference model. Using the result in Equation 1, hypotheses about the parameters in LMMs can be tested. The significance of the likelihood ratio test statistic can be determined by referring it to a $x^2$ distribution with the appropriate degrees of freedom. If the LRT statistic is sufficiently large, there is evidence against the null hypothesis model and in favor of the reference model. If the likelihood values of the two models are very close, and the resulting LRT statistic is small, we have evidence in favor of the nested (null hypothesis) model.

The likelihood ratio tests that we use to test linear hypotheses about fixed-effect parameters in an LMM are based on ML estimation; using REML estimation is not appropriate in this context (Morrell, 1998; Verbeke and Molenberghs, 2000). For LRTs of fixed effects, the nested and reference models have the same set of covariance parameters but different sets of fixed-effect parameters. The test statistic is calculated by subtracting the –2 ML log-likelihood for the reference model from that for the nested model. The asymptotic null distribution of the test statistic is a $x^2$ with degrees of freedom equal to the difference in the number of fixed-effect parameters between the two models. When testing hypotheses about covariance parameters in linear mixed model, REML estimation should be used for both the reference and nested models.

REML estimation has been shown to reduce the bias inherent in ML estimates of covariance parameters (Morrell, 1998). We assume that the nested and reference models have the same set of fixed-effect parameters, but different sets of covariance parameters. To carry out a REML-based likelihood ratio test for covariance parameters, the −2REML log-likelihood value for the reference model is subtracted from that for the nested model. The null distribution of the test statistic depends on whether the null hypothesis values for the covariance parameters lie on the boundary of the parameter space for the covariance parameters or not.

Another set of tools useful in model selection are referred to as information criteria. The information criteria (sometimes referred to as fit criteria) provide a way to assess the fit of a model based on its optimum log-likelihood value, after applying a penalty for the parameters that are estimated in fitting the model. A key feature of the information criteria is that they provide a way to compare any two models fitted to the same set of observations; i.e., the models do not need to be nested. We use the "smaller is better" form for the information criteria that is, a smaller value of the criterion indicates a "better" fit. The Akaike information criterion (AIC) may be calculated based on the (ML or REML) log-likelihood, $l(\hat{\beta}, \hat{\theta})$, of a fitted model as follows (Akaike, 1973)

$$AIC = -2\ x\ l(\hat{\beta}, \hat{\theta}) + 2p \qquad [10]$$

In Equation 10, $p$ represents the total number of parameters being estimated in the model for both the fixed and random effects. Note that the AIC in effect "penalizes" the fit of a model for the number of parameters being estimated by adding $2p$ to the −2 log-likelihood. Some software procedures calculate the AIC using slightly different formulas, depending on whether ML or REML estimation is being used.

The BIC is also commonly used and may be calculated as follows:

$$BIC = -2l(\hat{\beta}, \hat{\theta}) + p\ x\ ln(n) \qquad [11]$$

The BIC applies a greater penalty for models with more parameters than does the AIC, because we multiply the number of parameters being estimated by the natural logarithm of $n$, where $n$ is the total number of observations used in estimation of the model.

### 3.2.7. Checking model assumptions (diagnostics)

After fitting an LMM, it is important to carry out model diagnostics to check whether distributional assumptions for the residuals are satisfied and whether the fit of the model is sensitive to unusual observations. In this study, two basic assumptions were considered for mixed effects.

Assumption 1 - The within-group errors are independent and identically normally distributed, with mean zero and variance σ2, and they are independent of the random effects.

Assumption 2 - The random effects are normally distributed, with mean zero and covariance matrix D and are independent for different groups.

Diagnostic methods for standard linear models are well established in the statistics literature. In contrast, diagnostics for LMMs are more difficult to perform and interpret, because the model itself is more complex, due to the presence of random effects and different covariance structures (Schabenberger, 2004). The primary quantities used to assess the adequacy of Assumption 1 are the within-group residuals, defined as the difference between the observed response and the within-group fitted value. Conditional on the random effects variance covariance components, the within-group residuals are the BLUPs of the within-group errors. In practice, the within-group residuals are only estimated BLUPs, as the random-effects variance covariance components need to be replaced with their estimates. Nevertheless, they generally provide good surrogates for the within-group errors and can be used to qualitatively assess the validity of Assumption 1. Other quantities used for assessing Assumption 1 graphically include the within-group fitted values, the observed values, and any covariates of interest.

Graphically, Assumption2 can also be assessed by using two types of diagnostic plots.

• qqnorm: normal plot of estimated random effects for checking marginal normality and identifying outliers;

• pairs: scatter plot matrix of the estimated random effects for identifying outliers and checking the assumption of homogeneity of the random effects covariance matrix

## 3.3. Methods of data analysis for GEI and stability study

Various statistical procedures have been proposed to assess GEI and find out the stability of new cultivars. One of the most frequently used stability measures is based on a regression model

(Yates and Cochran 1938). However, it was developed by Finlay (Finlay and Wilkinson 1963) to describe the adaptation of individual varieties to changing environment and while Eberhart (Eberhart and Russell 1966), used regression coefficient values as measures of environmental response and deviations from regression as measures of stability. Several of these statistics have been summarized and compared by Lin (Lin *et al*, 1986) who pointed out that stability statistics fall into four groups depending on whether they are based on the deviation from the average genotype effect or on the genotype by environment term and whether or not they incorporate a regression model on an environment index. Other workers have suggested use of Parameters like Coefficient of Variation, Wricke's ecovalence and AMMI Stability Value as measures of stability. Further, the simultaneous selection for yield and stability in crop performance is also used based on Kang's modified rank-sum method. This yield-stability statistic component is basically based on Shukla's (1972) stability-variance statistic. The Additive Main Effects and Multiplicative Interaction (AMMI) model has found more use recently since it incorporates both the classical additive main effects model for GEI and the multiplicative components into an integrated least square analysis and thus becomes more effective in selection of stable genotypes (Crossa *et al.*, 1991)

### 3.3.1 Combined analysis of variance

Preliminary ANOVAs can be carried out for individual experiments to assess variation among environments for experimental error and, possibly, genotypic variance. Combined ANOVAs for a complete set of experiments or its subsets can be performed with different objectives, such as:

- Verification of the occurrence (i.e. significance) of different effects;
- Estimation and comparison of mean values for levels of fixed factors (in particular, genotype mean values across the region or within sub regions); and
- Estimation of the size of genotypic and genotype-environmental variance components (possibly as a step towards estimation of genetic parameters).

The ANOVA may also represent one step in the analysis of adaptation or in the assessment of yield stability measures. In the analysis of combined experiment of data from several environments, the first requirement is to assess the homogeneity of the error variance at the various environments. If the errors are homogeneous, the analysis can proceed. However, if the

error variances are heterogeneous, the data will be transformed to produce homogenous variance or the locations may be separated into groups within which the variance is homogenous. In multi-environment yield trials of G genotypes (i=1,2,…,g), E environments(j=1,2,…e) and r replicates(l=1,2,…,r) arranged in RCBD, the liner model for the conventional combined analysis variance(ANOVA) is

$$Y_{jir} = \mu + G_i + E_j + GE_{ji} + B_{jr} + \epsilon_{jir} \quad [12]$$

where,

$Y_{jil}$ is the observed yield response of the $i^{th}$ genotype of the $j^{th}$ environment

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

$G_i$ is the effect of $i^{th}$ genotype; thus $\sum_1^g Gi = 0$

$E_j$ is the effect of the $j^{th}$ environment and $\sum_1^e E_i = 0$

$GE_{ji}$ is the interaction effect of the $i^{th}$ genotype in the $j^{th}$ environment.

$B_{jr}$ is the effect of the $i^{th}$ replication in the $j^{th}$ environment, and

$\epsilon_{jir}$ is random error term with mean 0 and variance $\sigma^2_{jir}$ and distributed as NID (0, $\sigma^2_{jir}$)

### 3.3.2. The Additive Main Effect and Multiplicative Interaction effect Model (AMMI)

Gauch(1988,1992) has advocated the use of AMMI for yield trials. Gauch and Zobel (1988) compared the performance of AMMI analysis with the ANOVA approach and regression approach and found that ANOVA fails to detect a significant interaction components and the regression approach accounts only a small portion of the interaction sum of squares only when the patterns fits a specific model. AMMI combines analysis of variance (ANOVA) in to a single model with additive and multiplicative parameters. After removing the replicate effect when combining the data, the observations are portioned in to two sources: Additive main effects for genotypes and environments, and Non additive effects due to genotype-environment interaction. The AMMI model for G genotypes and E environments is given as

$$Y_{ij} = \mu + G_i + E_j + \sum_{n=1}^{n'} \lambda_n \alpha_{in} \gamma_{jn} + \epsilon_{ij}\ ; \quad \epsilon_{ij} \sim N(0, \sigma^2); i = 1,2, \dots g\ ; j = 1,2, \dots e \quad [13]$$

Where $Y_{ji}$ is the mean yield of $i^{th}$ genotype in the $j^{th}$ environment ; $\mu$ the grand mean; $G_i$ is the $i^{th}$ genotype effect ; $E_j$ is the $j^{th}$ environment effect; $\lambda_n$ is eigen value of the PCA axis n; $\alpha_{in}$ and $\lambda_{jn}$ are the $i^{th}$ genotype $j^{th}$ environment PCA scores for PCA axis n; $\varepsilon_{ji}$ is the residual; n' is the number of PCA axes in the model. Ordinarily the number n' is judged on the basis of empirical consideration on F-test of significance Gauch( 1988,1992). The residual combines the PCA

scores from the N-n' discarded axes, where N=min(g-1,e-1). The other constraints in the model 12 are

$$\sum_i \alpha_{in}^2 = \sum_j \gamma_{jn}^2 = 1 \; \forall \; n; \sum_i \alpha_{in}\alpha_{in*} = \sum_j \gamma_{jn}\gamma_{jn*} = 0,$$

$$n \neq n*; and \; \lambda_1 > \lambda_2 > \cdots > \lambda_{3n'} > 0$$

The model in (1) can be reparameterized as

$$Y_{ij} = \mu + G_i + E_j + Z_{ij}$$

Where $Z_{ij} = \sum_{n=1}^{n'} \lambda_n \alpha_{in}\gamma_{jn} + \epsilon_{ij}$

Let the estimate of the interaction in the (i,j)[th] cell of $Z_{ij}$ be $\hat{Z}_{ij} = Y_{ij} - \mu - G_i - E_j$. Using matrix notation, denote $\boldsymbol{Z} = (\hat{Z}_{ij})$ a matrix of order GEI. Now, the estimate of the parameters of the model is

$\hat{\lambda}_n$= the non-zero eigen values of **Z'Z** (in descending order), and

$\hat{\alpha}_{in}$= the principal components of the row sum of squares and cross product matrix **ZZ'**

$\hat{\gamma}_{jn}$= the principal components of the column sum of squares and cross product matrix **Z'Z**

Using these we can write $\hat{Z}_{ij} = \sum_{n=1}^{n'} \hat{\lambda}_n \hat{\alpha}_{in}\hat{\gamma}_{jn}$

### 3.3.3. Graphical plots (Bi-plots and 3-D plots)

The model formulation for AMMI shows its interaction part consists of summed orthogonal products. Because this form the interaction lends itself to graphical display in the form so-called biplots(Gabriel,1971). Let start with AMMI and assume that either two terms suffice for an adequate description of the interaction. For AMMI the interaction consists of the sum of two products: $\alpha_{i1}^*\gamma_{j1}^* + \alpha_{i2}^*\gamma_{j2}^*$ . The genotype scores, $\alpha_{i1}^*$ and $\alpha_{i2}^*$, are now interpreted as coordinates for planar depiction of the genotype, and the environmental scores, $\gamma_{j1}^*$ and $\gamma_{j2}^*$, for a similar depiction of the environment. The score determines the end points of the genotypic and environmental vectors, which depart from the origin. Simple geometric reveals that the interaction between a genotype i and an environment j can be obtained from a projection of either vector onto the other. The reason is that the interaction according to an AMMI model with two product terms of interaction, $\alpha_{i1}^*\gamma_{j1}^* + \alpha_{i2}^*\gamma_{j2}^*$ , is equal to the inner product between vectors $(\alpha_{i1}^*, \alpha_{i2}^*)$ and $(\gamma_{j1}^*, \gamma_{j2}^*)$, or the projection of either vector on to the other, times the length of the

vector on which projection take place. It is easy to read from a bi-plot the relative interaction that genotypes exhibit in a particular environment.

To have a better discussion on the graphical plots IPCAs (bi-plots, three dimensional plots *ets*.) resulted from the AMMI analysis, we must consider the following points (Kempton, 1984; Kroonenberg, 1995, as cited in Rashidi *et al*., 2013):
(i) The center of bi-plot shows the mean of genotypes or environments.
(ii) A long distance of a genotype (or an environment) from the center of bi-plot indicates a large interaction with that genotype (or environment).
(iii) The long length of a genotype on the environmental vector reveals more deviation from the mean and vice versa.
(iv) The angle between the vectors of a genotype and an environments shows that the interaction is positive or negative.

AMMI1 bi-plot is constructed with additive main effects or mean yield along the abscissa and the first IPCA or multiplicative interaction on the ordinate axis. Thus, the interpretation of the bi-plot assay is that if main effects have IPCA score close to zero, it indicates negligible interaction effects and when a genotype and an environment have the same sign on the IPCA axis, their interaction is positive; if different, their interaction is negative. The Bi-plot space of AMMI1 is divided into 4 sections(quadrants) from low yielding environments in quadrants 1 (up left) and 4 (low left) to high yielding environments in quadrants 2 (up right) and 3 (low right). From the bi-plot, if the points for environment are more scattered than the point for genotypes indicating that variability due to environments is higher than that due to genotypes difference, and the reverse is true if genotypes take the situation(Zobel *et al*. 1988). On the bi-plot, the points for the generally adapted genotypes would be at right hand side of grand mean levels (this suggests high mean performance) and close to the line showing IPCA= 0 and (this suggests negligible or no G × E Interaction).

AMMI2 biplot The IPCA 1 versus IPCA 2 biplot (i.e. AMMI 2 biplot) explain the magnitude of interaction of each genotype and environment. The genotypes and environments that are farthest from the origin being more responsive fit the worst. Genotypes and environments that fall into

the same sector interact positively; negatively if they fall into opposite sectors. A genotype showing high positive interaction in an environment obviously has the ability to exploit the agro-ecological or agro-management conditions of the specific environment and is therefore best suited to that environment (Rashidi *et al*., 2013). The interpretation for AMMI3 (3-dimentional plot) follows like AMMI2 interpretation.

# 4. RESULTS AND DISCUSSION

## 4.1. Longitudinal Analysis

### 4.1.1. Baseline information and descriptive statistics of coffee bean yield

The base line (year1) is the time when the age of coffee trees was 5 years after planted on the field. As explained in section 3.1, the data set consists of 204 subjects (coffee trees) with 7 measurements per subject. The data set is complete and balanced since the number of measurements at each time points is equal and there is no missing value in the data set. In the actual coffee bean yield, the minimum and maximum values were 0.0 and 5722.91 $kgha^{-1}$, respectively.

Table 5: Summary of coffee bean yield ($kgha^{-1}$) by location and coffee genotype over years

| | Year1 | Year2 | Year3 | Year4 | Year5 | Year6 | Year7 | Mean |
|---|---|---|---|---|---|---|---|---|
| Location | | | | | | | | |
| Agaro | 491.62 | 1193.04 | 1223.28 | 2595.83 | 973.57 | 2290.41 | 1225.01 | 1427.54 |
| Jimma | 714.47 | 1366.29 | 2103.67 | 2567.67 | 2626.28 | 2737.18 | 951.70 | 1866.75 |
| Metu | 909.08 | 1454.78 | 1034.08 | 1984.62 | 487.09 | 1710.08 | 1171.17 | 1250.13 |
| Genotype | | | | | | | | |
| G1 | 1000.31 | 1688.69 | 1584.85 | 2763.12 | 1765.86 | 2660.55 | 1435.07 | 1842.64 |
| G2 | 722.90 | 1561.16 | 2107.42 | 2734.05 | 2039.06 | 2706.10 | 1897.70 | 1966.91 |
| G3 | 647.98 | 1407.96 | 1013.64 | 2153.02 | 893.29 | 2362.30 | 373.15 | 1264.48 |
| G4 | 846.58 | 1475.75 | 1390.59 | 2301.97 | 1698.21 | 2189.88 | 895.92 | 1542.70 |
| G5 | 584.95 | 1372.53 | 1284.12 | 2201.45 | 1617.93 | 2035.21 | 1259.86 | 1479.44 |
| G6 | 484.59 | 1073.91 | 962.60 | 3788.05 | 657.74 | 2769.74 | 1078.47 | 1545.01 |
| G7 | 600.41 | 1016.29 | 1071.16 | 1780.97 | 1066.07 | 1583.42 | 706.95 | 1117.90 |
| G8 | 707.75 | 1625.22 | 1756.92 | 3292.48 | 1439.22 | 2850.34 | 1124.68 | 1828.09 |
| G9 | 562.74 | 1591.42 | 2068.46 | 2442.14 | 1741.19 | 2239.94 | 1500.26 | 1735.17 |
| G10 | 813.12 | 1344.07 | 1623.98 | 2939.42 | 1521.42 | 2589.03 | 1313.52 | 1734.93 |
| G11 | 510.65 | 998.25 | 1039.32 | 1546.32 | 905.28 | 1882.27 | 654.03 | 1076.59 |
| G12 | 679.64 | 1278.28 | 1551.99 | 2534.07 | 1600.42 | 1965.22 | 1137.05 | 1535.24 |
| G13 | 1089.73 | 1508.30 | 2037.63 | 2436.40 | 1973.64 | 2552.42 | 1853.50 | 1921.66 |
| G14 | 643.88 | 1071.92 | 1196.90 | 1564.27 | 682.27 | 1941.00 | 705.81 | 1115.15 |
| G15 | 629.15 | 1150.08 | 1366.18 | 1798.10 | 1384.31 | 1767.06 | 881.26 | 1282.31 |
| G16 | 794.28 | 1197.00 | 1686.53 | 1802.03 | 1455.91 | 1939.12 | 1154.33 | 1432.74 |
| G17 | 667.37 | 1385.80 | 970.25 | 2428.11 | 717.51 | 2146.56 | 999.76 | 1330.77 |
| Mean | 705.06 | 1338.04 | 1453.68 | 2382.70 | 1362.31 | 2245.89 | 1115.96 | 1514.81 |
| SD | 373.09 | 501.79 | 834.16 | 982.75 | 1310.45 | 901.61 | 867.16 | 1033.34 |

SD=standard deviation

Table 5 shows that the least and the highest overall mean value of coffee bean yield (705.06 $kgha^{-1}$ and 2382.70 $kgha^{-1}$ respectively) were observed at the base line and year four, respectively. It also presents that there is yield difference among locations and coffee genotypes. The highest mean value of coffee bean yield was recorded at Jimma (658.06 kg $ha^{-1}$). On the basis of combined mean values across location over years, genotypes: G2 (check), G13, and G1 (check) exhibited the top three mean values of coffee bean yield. Starting from the base line, both the mean and variance of the observed coffee bean yield have increased up to the fourth year, and then fluctuates through seventh year (Table 5).

### 4.1.2. Exploring the individual profile of coffee bean yield

Individual profile plots in Figure 1a show that there is variability within and between coffee trees. From Figure 1a, the variability between coffee trees at the base line is clearly observed and evident to include random intercepts in a linear mixed model. In Figure 1b, the coffee bean yield values for almost all coffee trees within a given block tend to follow the same trend over time. But for the levels of block, the trend is different over time. Thus, block b6 and b12 are evident for different trends in coffee bean yield values over time. These patterns suggest that an appropriate model for the data might include random block-specific intercepts and slopes.

On the other hand, besides fixed factors, the coffee bean yield values which tend to be vary from block to block at base line and over time suggesting that a model should also include random block-specific intercepts and slopes of time variant variable. Similarly, coffee bean yield values which tend to be different by coffee tree at base line and over time suggesting that random coffee tree-specific intercepts and random slopes of time variant variables should be included in the model selection. The growth trend in the individual profile plot shows that the coffee bean yield is in somehow increasing and then tends to decline with zigzag trajectory over time. The factor that causes the growth trend to follow a zigzag (rises and down) trajectory was clearly investigated in section 4.1.3 by exploring the mean profile of coffee yield over time. Of course at this point it is not yet possible to decide the trend (i.e. linear, quadratic, cubic and others), but by adjusting this zigzag trajectory, it is expected that the possible trend could be determined.

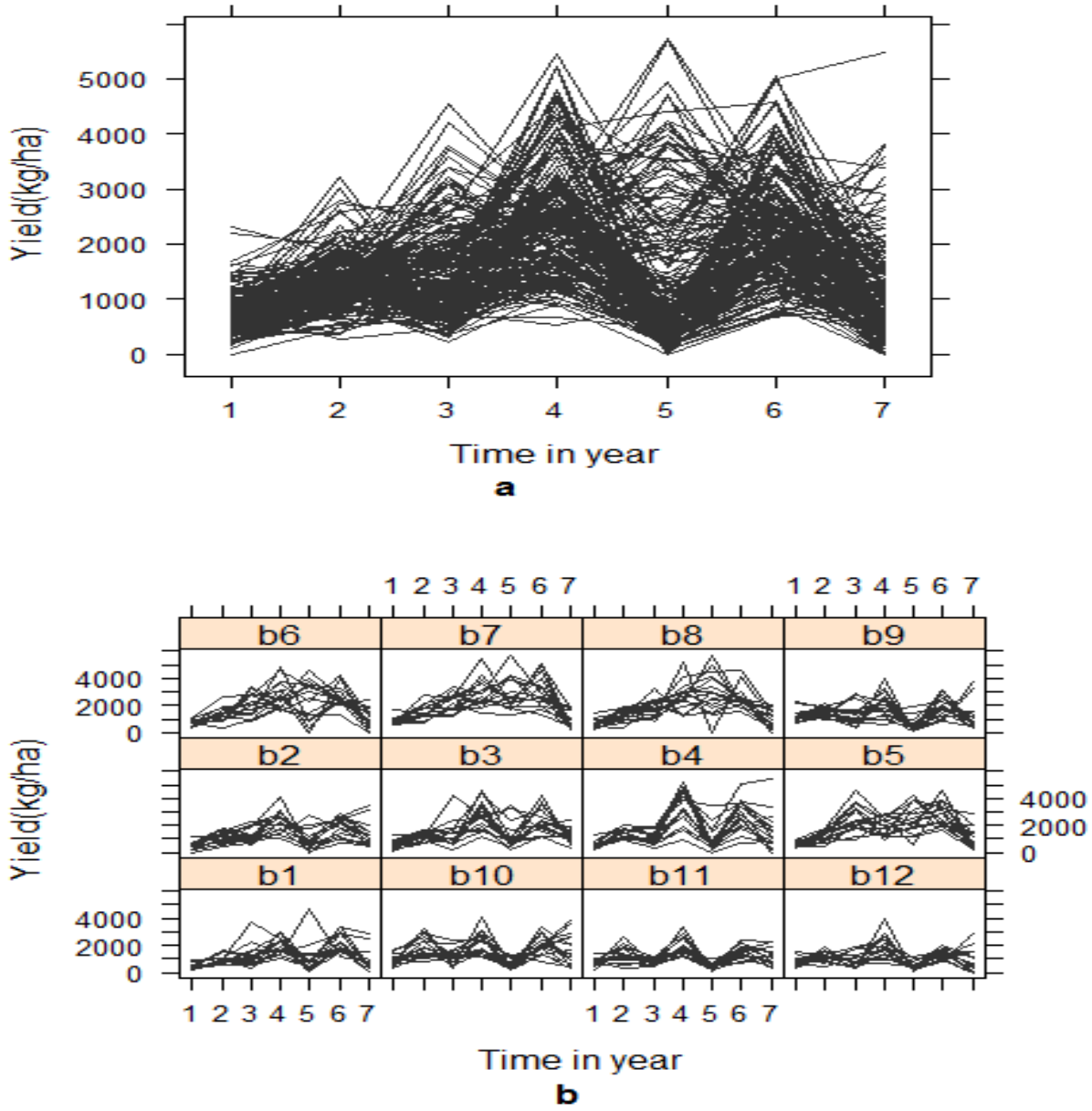



Figure 1: Individual profile plots of CBY by coffee tree (a) and coffee tree nested in block (b)

### 4.1.3. Exploring the mean structure of coffee bean yield

Besides plotting the yield of coffee tree over time, it is also useful to include graph for different subgroups to illustrate the relationship between coffee bean yield and explanatory variables over time. The results of this exploration are useful in order to choose a fixed- effects structure for the linear mixed model. The mean profile per location and genotype arm are plotted in Figure 2. The mean profile plot by location in Figure 2a shows that there is location by time interaction

effect, and thus the average evolution of coffee bean yield in Jimma is quite different from that of Agaro and Metu. But the trends for Agaro and Metu are almost similar with the falling and the rising trajectory. The mean profile plot in Figure 2b shows that there is an average evolution of coffee bean yield for each coffee genotype over time. Figure 2b also shows that there is a mean difference between coffee genotypes at each point of time. Accordingly, all fixed factors such as, location, genotype, and genotype by location interaction and location by time interaction may have significant fixed effects and can be included in the mean structure of linear mixed model.

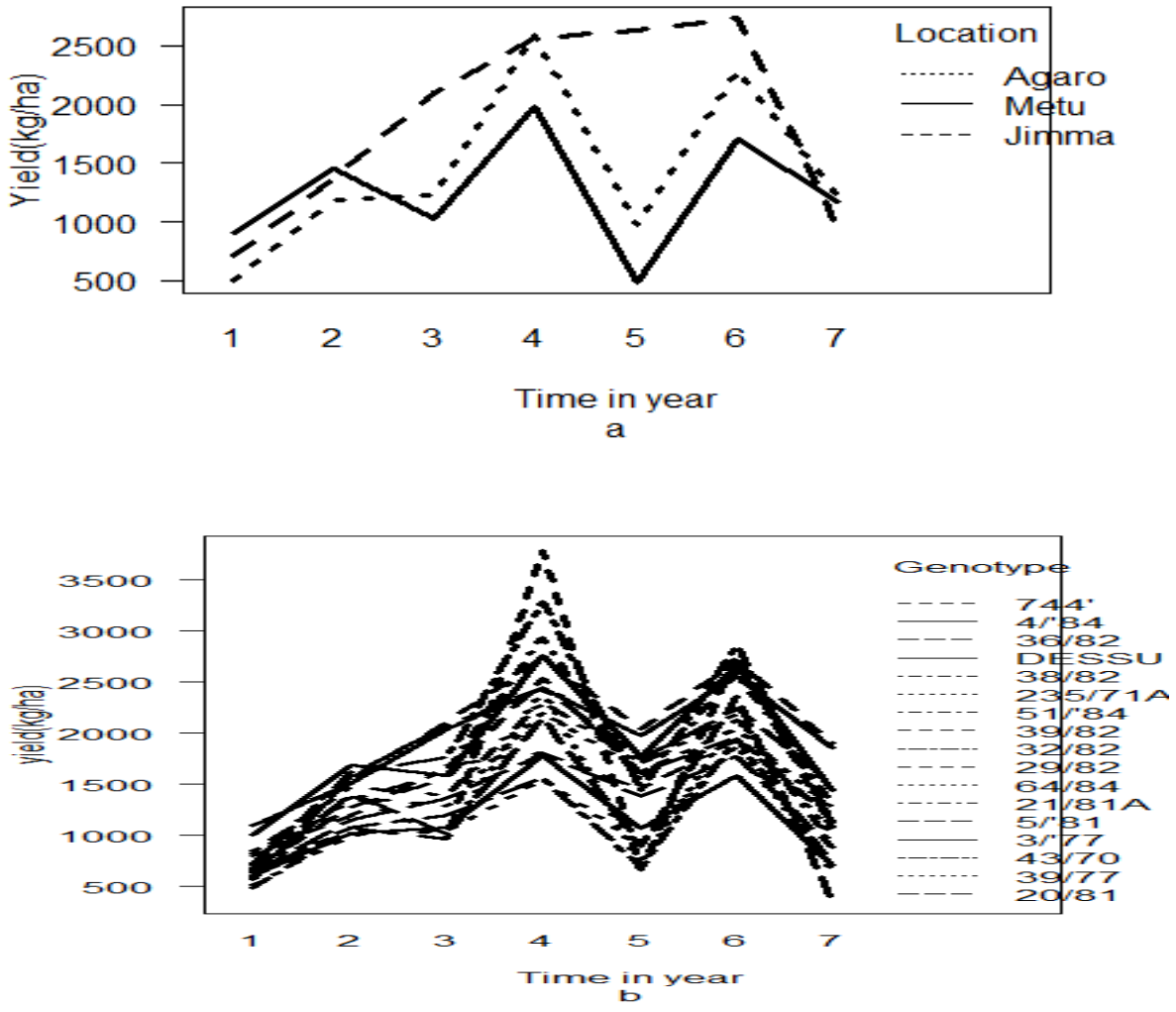



Figure 2: The mean profile plot of coffee yield by location (a) and genotype (b)

In all mean profile plots, we can observe that there is a time variant factor which alters the growth trend with a zigzag (rise and fall) trajectory in regular basis (rise for even years and fall for odd years). Having such unusual shape of growth trajectory, it was clear that a linear model would be inappropriate. Similarly, using a polynomial model would not solve the problem due to two reasons. First, it couldn't be able to show the growth trend regardless of the rising and the falling trajectory. Second, it is difficult to set up parsimonious model. Thus, quadratic model would allow for a change in the rate of change. However, quadratic model could only capture the shape of a growth trajectory with one bend. To capture the shape of growth trajectory which has two bends would require a cubic model. The cubic model allows for a change in a change in a rate of change. As time increases, degree of polynomials also increases and at the same time the number of random effects associated with the slope of each polynomial also increases. This makes the work so complicated in the computation of the variance covariance matrix of random effects in the linear mixed model, and as a result we couldn't have a parsimonious model.

However, it is possible to control the variability due to rise and down alternation in the growth trend by inserting a time variant factor in the model with binary coding scheme for rising and down if the factor for this alternating is known. Thus, the property "*biennia"* in the bearing habit of coffee which occurred in two years interval was evident for this unusual trajectory (rise and down alternation). The general mean profile in Figure 3 clearly shows the rise and fall alternations over time due to biennial effect. From Figure 2a&b), we can also observe that the effect of this factor is different among locations and genotypes. Thus, Figure 2a shows that the effect of biennial is clearly observed in Agaro and Metu but not Jimma. This indicates that coffee trees that grown in Jimma are not affected by biennial relative to Agaro and Metu. Similarly, Figure 2b shows that some genotypes relatively more affected by biennial.

In this study the time variant factor "*biennia"* is coded (0 for raising and 1 for falling) and considered as a factor in the subsequent analysis to control the variability that brings unusual shape of trajectory in the trend of coffee bean yield. As a result, we expect biennial by location and genotype interaction effects in the linear mixed model, and therefore a quadratic evolution is also expected on the long run trajectory (Figure 3). Regardless of the alternations due to biennial in the growth trend, the clear shape of growth trajectory is depicted in Figure.4 by using loess

smoothing method. The plot shows that the mean evolution of coffee yield is a quadratic function of time.

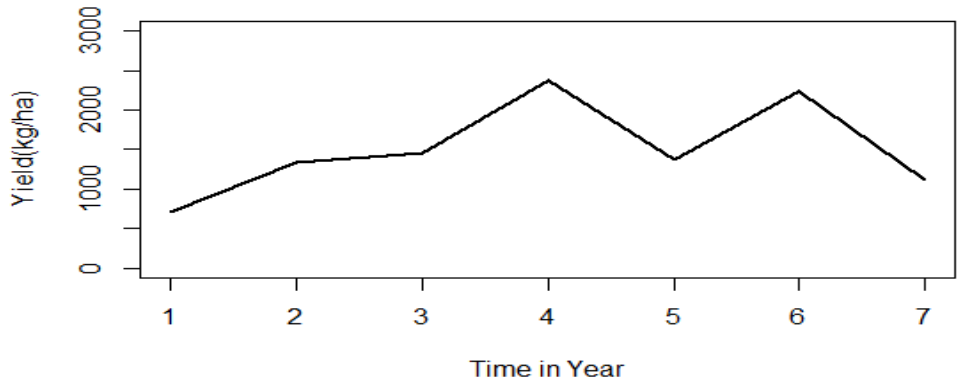



Figure 3: General mean profile plot of CBY

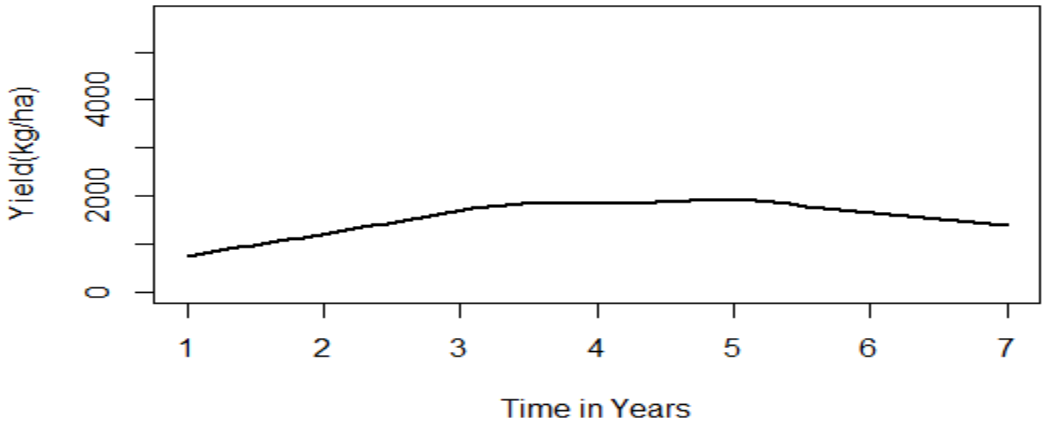



Figure 4: General mean profile plot of CBY by using loess smoothing

### 4.1.4. Exploring the variance and correlation structure in coffee bean yield

To have an appropriate model, investigation of how the base line variance could evolve over time is very important step in the modeling approach. This can be investigated with plots of individual profile, variance function, and by using variance covariance matrix of the observed measurements. As shown in the individual profile plot (Figure 1a&b, there is a considerable within and between subject variability over time. This can also be shown in Table 6 on the

diagonal of variance covariance matrix of the observed data. It also shows that variance of coffee yield increases up to year 5 and then decreases. Graphically, the evolution of variance is also depicted in Figure 5a&b, and shows that the overall variance increases up to year 5$^{th}$ and then tends to decrease afterwards. But from Figure 5), it is expected that the variances of each location differ in magnitude and evolution. Having this, it is clear that different variance models should be fitted and compared to handle heterogeneous variability across location over time.

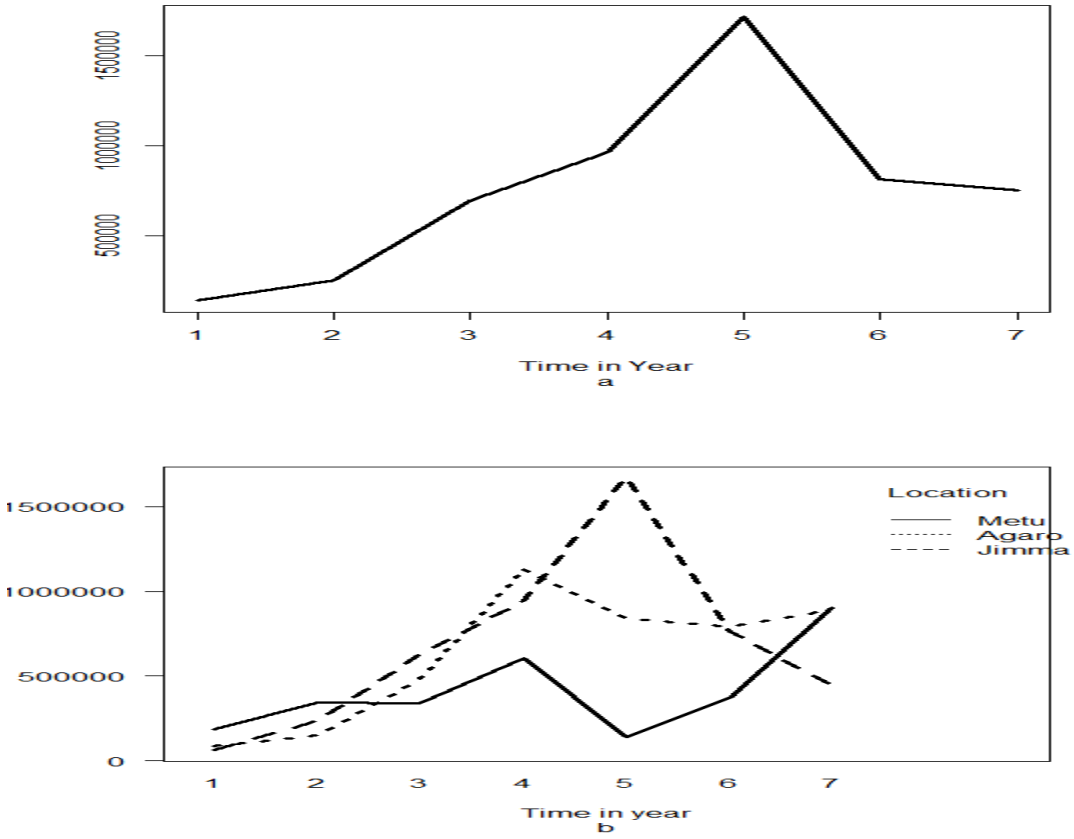



Figure 5: Plot of variance of CBY over year: overall (a) and by location (b)

Observations that are obtained from repeated measurements over time may not be independent in many circumstances and this induces what is known as autocorrelation among observations. Thus, investigating the structure of correlation among observations is an important aspect of exploratory data analysis to set up a parsimonious model. The correlation structure describes how measurements within a subject are correlated. This correlation structure can be studied through the correlation matrix, or scatter plot matrix. The correlation scatter plot matrix in Figure 6 and the correlation structure matrix in Table 6 show that there is a correlation among

measurements of coffee yield over time. From Table 6, the correlations of the observations at time 3 with time 5 and at time 4 with time 6 are evident for the presence of correlation among measurements of coffee yield over year. Except some time points, Table 6 show that there is decaying of correlation over time (for instance, among even years). Both Figure 6 and Table 6 shows that models for correlation structure should be compared by including autoregressive order one and unstructured correlation model.

Table 6: Variance and correlation structure of coffee bean yield over year

| | | **Year1** | **Year2** | **Year3** | **Year4** | **Year5** | **Year6** | **Year7** |
|---|---|---|---|---|---|---|---|---|
| Var-Cov | **Year1** | 139193 | | | | | | |
| | **Year2** | 51952 | 251795 | | | | | |
| | **Year3** | 82443 | 43799 | 695816 | | | | |
| | **Year4** | 6139 | 162626 | 80656 | 965795 | | | |
| | **Year5** | 41812 | 115405 | 714341 | 125243 | 1717291 | | |
| | **Year6** | -4145 | 83181 | 232262 | 478531 | 389146 | 812893 | |
| | **Year7** | 65972 | 101868 | 145480 | 202990 | 115737 | 93022 | 751968 |
| Corr | **Year1** | 1 | | | | | | |
| | **Year2** | 0.28 | 1 | | | | | |
| | **Year3** | 0.26 | 0.10 | 1 | | | | |
| | **Year4** | 0.02 | 0.33 | 0.10 | 1 | | | |
| | **Year5** | 0.09 | 0.18 | 0.65 | 0.10 | 1 | | |
| | **Year6** | -0.01 | 0.18 | 0.31 | 0.54 | 0.33 | 1 | |
| | **Year7** | 0.20 | 0.23 | 0.20 | 0.24 | 0.10 | 0.12 | 1 |

Var-Cov= variance covariance; Corr= correlation

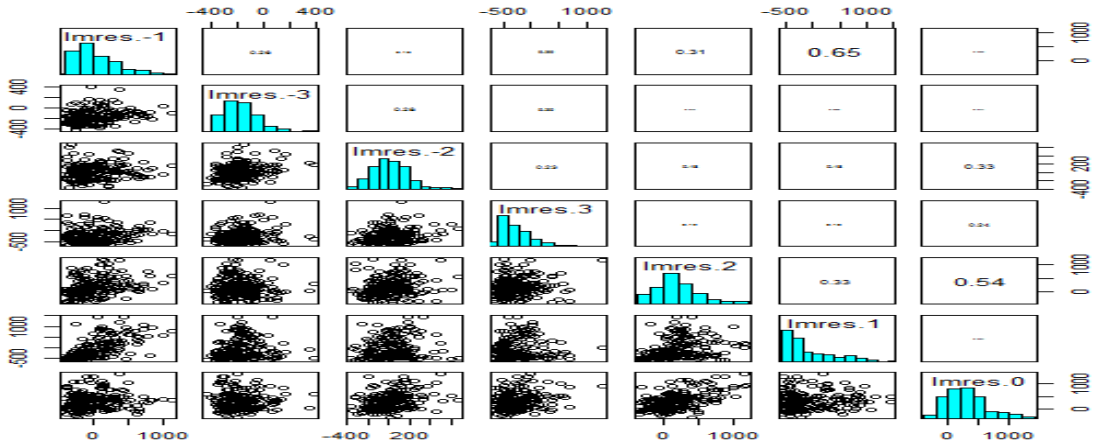



Figure 6: Scatter plot of correlation matrix for coffee yield over year

### 4.1.5 Linear mixed model results

#### 4.1.5.1. Possible structures of fixed and random effects of coffee bean yield

The top-down strategy that were suggested by Verbeke and Molenberghs (2000, Chapter 9) for building an LMM for a given data set was used in this study. The possible variables (factors and interaction terms) that able to show variation in the part exploratory data analysis were candidates in the specification of the initial model. The possible random effects and the shape of trajectory (quadratic evolution) that was explored through exploratory data analysis have been also considered. Hence, the initial model with a potential "loaded" mean structure given by

$$\left.\begin{aligned} Y_{tji} = {} & \beta_0 + \beta_1 L2_j + \beta_2 L3_j + \beta_3 G2_{ji} + \beta_4 G3_{ji} + \cdots + \beta_{18} G17_{ji} + \\ & \beta_{19} L2_j G2_{ji} + \beta_{20} L2_j G3_{ji} + \cdots + \beta_{34} L2_j G17_{ji} + \\ & \beta_{35} L3_j G2_{ji} + \beta_{36} L3_j G3_{ji} + \cdots + \beta_{50} L3_j G17_{ji} + \\ & \beta_{51} B_t + \beta_{52} t_t + \beta_{53} t^2{}_t + \\ & \left(\begin{aligned} & \beta_{54} L2_j + \beta_{55} L3_j + \beta_{56} G2_{ji} + \beta_{57} G3_{ji} + \cdots + \beta_{71} G17_{ji} + \\ & \beta_{72} L2_j G2_{ji} + \beta_{73} L2_j G3_{ji} + \cdots + \beta_{84} L2_j G17_{ji} + \\ & \beta_{85} L3_j G2_{ji} + \beta_{86} L3_j G3_{ji} + \cdots + \beta_{100} L3_j G17_{ji} \end{aligned}\right) T_t + \\ & \left(\begin{aligned} & \beta_{101} L2_j + \beta_{102} L3_j + \beta_{103} G2_{ji} + \beta_{104} G3_{ji} + \cdots + \beta_{118} G17_{ji} + \\ & \beta_{119} L2_j G2_{ji} + \beta_{120} L2_j G3_{ji} + \cdots + \beta_{134} L2G17_{ji} + \\ & \beta_{135} L3_j G2_{ji} + \beta_{137} L3_j G3_{ji} + \cdots + \beta_{151} L3_j G17_{ji} \end{aligned}\right) T_t{}^2 + \\ & \left(\begin{aligned} & \beta_{152} L2_j + \beta_{153} L3_j + \beta_{154} G2_{ji} + \beta_{155} G3_{ji} + \cdots + \beta_{169} G17_{ji} + \\ & \beta_{170} L2_j G2_{ji} + \beta_{171} L2_j G3_{ji} + \cdots + \beta_{195} L2_j G17_{ji} + \\ & \beta_{196} L3_j G2_{ji} + \beta_{197} L3_j G3_{ji} + \cdots + \beta_{211} L3_j G17_{ji} \end{aligned}\right) B_t + \end{aligned}\right\} \text{fixed}$$

$$\left.\begin{aligned} & b_{0j} + b_{1j} T_t + b_{2j} T_t{}^2 + b_{3j} B_t + \\ & b_{0ji} + b_{1ji} T_t + b_{2ji} T_t{}^2 + b_{3j} B_t + \\ & \varepsilon_{tji} \end{aligned}\right\} \text{random} \qquad [4.1]$$

Where

$T_t$ time at which coffee yield measured
$Y_{jit}$ coffee yield for tree i in block j at time t
$L_j$ location of the j block
Gji Genotype of coffee tree i in block j
$B_t$ Biennaly at time t
$\beta_0, \beta_1, \beta_2 \ldots \beta_{211}$ the fixed effect coefficient parametrs
$b_{0j}, b_{1j}, b_{2j}$ and $b_{3j}$ the random effects for block j associated with intercept, linear time slope , quadratic time slope and bienial effect, respectively
$b_{0ji}, b_{1ji}, b_{2ji}$ and $b_{3ji}$ the random effects for tree i in block j associated with intercept, linear time slope , quadratic time slope and biennial effect, respectively
$\varepsilon_{tji}$ random error

***Note: for the sake of interpretation of intercept ($\beta_0$), time was centered on the average***

#### 4.1.5.2. Selection of the fixed effects structure for coffee bean yield

The selection of fixed effects have been done in the conventional linear model setting by using ML estimation method, and AIC and BIC values without considering the structure of random effects. All terms in the fixed effect structure of initial linear mixed model (Model 4.1) were fitted first so as to identify significant fixed effects for coffee yield over time. The fitted model then reduced by removing none significance terms starting from high order interaction terms by using AIC and BIC values. From the outputs in Table 7, we can observe that all terms except the last four interaction terms (Genotype*Time, Genotype*Time$^2$, Location*Genotype*Time and Location*Genotype*Time$^2$) are statistically significant. Thus, none significant terms should be removed from the model starting with the most none significant one of which is the interaction term (location*Genotype*Time) with p-value of 0.6767. The model was then refitted after removing each none significant interaction terms one by one and finally the AIC and BIC values dropped from 22979.99 to 22914.77 and from 24059.12 to 23488.55, respectively, indicating a better fit.

Table 7: Fixed effects structure with all covariates and interaction terms with the corresponding p-values from the overall F test

| Effects | DF | F-value | p-value |
|---|---|---|---|
| Intercept | 1 | 5621.626 | <.0001 |
| Time | 1 | 109.296 | <.0001 |
| Time$^2$ | 1 | 144.651 | <.0001 |
| Biennial | 1 | 412.948 | <.0001 |
| Location | 2 | 82.29 | <.0001 |
| Genotype | 16 | 12.092 | <.0001 |
| Location*Genotype | 32 | 2.163 | 0.0002 |
| Location*Biennial | 2 | 8.989 | 0.0001 |
| Location*Time | 2 | 15.253 | <.0001 |
| Location*Time$^2$ | 2 | 87.656 | <.0001 |
| Genotype*Biennial | 16 | 4.583 | <.0001 |
| Genotype*Time | 16 | 1.293 | 0.1929 |
| Genotype*Time$^2$ | 16 | 1.025 | 0.4262 |
| Location*Genotype*Biennial | 32 | 2.385 | <.0001 |
| Location*Genotype*Time | 32 | 0.87 | 0.6767 |
| Location*Genotype*Time$^2$ | 32 | 1.015 | 0.4446 |

DF=degree of freedom

#### 4.1.5.3. Selection of the random effects structure for coffee bean yield

After selecting the structure of fixed effects by using ML estimation method, and AIC and BIC values in the conventional linear model setting, the next work was the selection of the structure of random effects. Given the selected fixed effects structure, starting from a simple linear regression model (no random effects), all random effects associated with intercepts and slopes for block and coffee tree nested in block was subjected in the top-down selection strategy by using REML estimation method, and AIC and BIC values. The random effects associated with intercept, biennial, and linear and quadratic slopes of time were selected first for block and then for coffee tree given the selected random effects of block. The inclusions of random effects in the model were done by keeping previously included random effects there (Table 8).

Table 8: Selection of random effects to be included in the linear mixed model

| No | Random Effects | AIC | BIC | logLik | Test | | | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|---|
| | | | | For block | | | | | |
| 1 | No random effect | 21618.4 | 22183.61 | -10700.2 | | | | | |
| 2 | intercept | 21577.48 | 22147.87 | -10678.74 | 1 | vs | 2 | 42.92 | <.0001 |
| 3 | Biennial | 21567.84 | 22148.6 | -10671.92 | 2 | vs | 3 | 13.64 | 0.0011 |
| 4 | Linear slope | 21571.88 | 22168.19 | -10670.94 | 3 | vs | 4 | 1.96 | 0.5798 |
| 5 | Quadratic slope | 21578.62 | 22195.69 | -10670.31 | 4 | vs | 5 | 1.25 | 0.8696 |
| | | | | For coffee tree nested in block | | | | | |
| 1 | No random effect | 21567.84 | 22148.6 | | | | | | |
| 2 | intercept | 21569.84 | 22155.79 | -10671.92 | 1 | vs | 2 | 0.00 | 0.9986 |
| 3 | Biennial | 21573.74 | 22170.06 | -10671.87 | 2 | vs | 3 | 0.10 | 0.9524 |
| 4 | Linear slope | 21574.75 | 22186.63 | -10669.38 | 3 | vs | 4 | 4.99 | 0.1725 |
| 5 | Quadratic slope | 21582.74 | 22215.36 | -10669.37 | 4 | vs | 5 | 0.01 | 1.0000 |

AIC= Akaike Information Criterion BIC= Bayesian Information Criteria; logLik= log likelihood; L.Ratio= likelihood ratio

The choice was made with AIC and BIC values for which smaller value is considered as better. Table 8 shows summary measures; Akaki information criteria (AIC), Bayesian information criteria and likelihood ratio test for the models with different random effects of block and coffee tree nested in block. It indicates that the model is improved when random effects of block associated with intercept and biennial are included in the model (AIC=21567.84 and BIC = 22148.6). But the AIC and BIC values were no more dropped when the random effects of coffee

tree nested in block associated with intercept biennial, and linear and quadratic slopes of time are included in the model. Hence, the final reduced model given by

$$
\left.\begin{aligned}
Y_{jit} = \beta_0 + \beta_1 L2_j + \beta_2 L3_j + \beta_3 G2_{ji} + \beta_4 G3_{ji} + \cdots + \beta_{18} G17_{ji} + \\
\beta_{19} L2_j G2_{ji} + \beta_{20} L2_j G3_{ji} + \cdots + \beta_{34} L2_j G17_{ji} + \\
\beta_{35} L3_j G2_{ji} + \beta_{36} L3_j G3_{ji} + \cdots + \beta_{50} L3_j G17_{ji} + \\
\beta_{51} B_t + \beta_{52} t_t + \beta_{53} {t^2}_t + \\
(\beta_{54} L2_j + \beta_{55} L3_j) T_t + \\
(\beta_{56} L2_j + \beta_{57} L3_j) {T_t}^2 + \\
\left(\begin{aligned}
&\beta_{58} L2_j + \beta_{59} L3_j + \beta_{60} G2_{ji} + \beta_{61} G3_{ji} + \cdots + \beta_{75} G17_{ji} + \\
&\beta_{76} L2_j G2_{ji} + \beta_{77} L2_j G3_{ji} + \cdots + \beta_{91} L2_j G17_{ji} + \\
&\beta_{92} L3_j G2_{ji} + \beta_{93} L3_j G3_{ji} + \cdots + \beta_{107} L3_j G17_{ji}
\end{aligned}\right) B_t +
\end{aligned}\right\} \text{fixed}
$$

$$
\left.\begin{aligned}
b_{0j} + b_{3j} B_t + \\
\varepsilon_{jit}
\end{aligned}\right\} \text{random}
\qquad [4.2]
$$

where,

$T_t$ time at which coffee yield measured

$Y_{jit}$ coffee yield for tree i in block j at time t

$L_j$ location of the j block

Gji Genotype of coffee tree i in block j

$B_t$ Biennaly at time t

$\beta_0, \beta_1, \beta_2 \ldots \beta_{107}$ the fixed effect coefficient parametrs

$b_{0j}$ and $b_{3j}$ the random effects for block j associated with intercept and bienial effect, respectively

$\varepsilon_{jit}$ random error

***Note: for the sake of interpretation of intercept ($\beta_0$), time was centered on the average***

#### 4.1.5.4. Selection of the variance and correlation functions for coffee bean yield

In the cluster longitudinal data setting, we have only fitted a model with fixed and random effects structure that constitutes the major variability in the data. The model was fitted such that the random effects of block associated with intercept and biennial effect were included to account for the variability between blocks. However, in longitudinal study, selecting best model is not selecting a model with only the fixed and random effect structure but also variance and correlation structures. Thus, the part of exploratory data analysis in section 4.1.4 was evident that the model should be refitted and compared again to account for the variance and correlation structures across location over year.

In section 4.1.4, it was shown that yield variability differs in magnitude and evolution across location over years. This variability is expected to be constant to compare levels of fixed factors corresponding to location, genotype and interaction terms. In this study different variance functions like varPower, varFixed, varIdent and varExp were used and compared to model the variance structure within group using covariates location (l) and time (t) (Table 9). Based on the AIC and BIC value, the two variance functions (varIdent(t) and varIdent(t,l)) were preferred variance functions compared to the others . However, varIdent(t) has small AIC value compared to varIdent(t,l)) but the reverse is true on BIC value (Table 9). This is due to the fact that the AIC performs poorly if there are too many parameters in the model (Sugiura, 1978, as cited in Girma, 2010). Thus, in addition to fixed parameters there are 7 parameters if varIdent(t) functions is used, but 21(7x3) parameters if varIdent(t,l)) is used. For this reason, the selection is made based on the BIC value since it applies a greater penalty for models with more parameters than does the AIC. Therefore, the heterogeneous variance function (varIdent(t)) can model different variances over year by using covariates (time) and found to be preferable variance function compared to others (AIC=21347.25, BIC=22013.22).

In addition to variance functions, different correlation functions were used to model the dependency among measurements coming from the same coffee tree. Table (9) presents the common correlation functions (compound symmetry (corCompSymm), autoregressive of order 1, 2, 3, and 4 (corAR1, corAR2, corAR3 and corAR4), exponential (corExp), Gaussian (corGaus), and unstructured (corSymm(UN))) which were compared to model the correlation structure among measurements of coffee bean yield over time. Based on the AIC and BIC values, the fitted model with unstructured correlation function (corSymm(UN)) and autoregressive of order 3 (corAR3) found to be a better fit compared to others. Since many parameters in the unstructured correlation function, the selection was made on the BIC value likewise the variance function. Therefore, autoregressive of order 3 (corAR3), found to be a better fit based on the BIC value (AIC=21358.79, BIC=21986.22).

Table 9: Comparison of different models for variance and correlation structure

| | AIC | BIC | logLik | Test | | | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| **For variance** | | | | | | | | |
| varConstant | 21567.84 | 22148.6 | -10671.92 | | | | | |
| varFixed(t) | 21435.56 | 22016.32 | -10605.78 | | | | | |
| varPower(t) | 21435.68 | 22021.63 | -10604.84 | 2 | vs | 3 | 1.88 | 0.171 |
| varPower(t,l) | 21425.78 | 22022.10 | -10597.89 | 3 | vs | 4 | 13.90 | 0.001 |
| varExp(t) | 21477.00 | 22062.95 | -10625.50 | 4 | vs | 5 | 55.23 | <.001 |
| varExp(t,l) | 21470.03 | 22066.34 | -10620.01 | 5 | vs | 6 | 10.98 | 0.004 |
| varIdent(l) | 21564.01 | 22155.15 | -10668.01 | 6 | vs | 7 | 95.99 | <.001 |
| varIdent(t) | 21401.35 | 22013.22 | -10582.67 | 7 | vs | 8 | 170.67 | <.001 |
| varIdent(t,l) | 21347.25 | 22031.72 | -10541.62 | 8 | vs | 9 | 82.10 | <.001 |
| **For correlation** | | | | | | | | |
| No correlation | 21401.35 | 22013.22 | -10582.67 | | | | | |
| corSymm(UN) | 21325.07 | 22045.84 | -10523.54 | 1 | vs | 2 | 118.27 | <.001 |
| corAR(1) | 21378.10 | 21995.17 | -10570.05 | 2 | vs | 3 | 93.03 | <.001 |
| corAR(2) | 21368.10 | 21990.34 | -10564.05 | 3 | vs | 4 | 12.01 | 0.001 |
| corAR(3) | 21358.79 | 21986.22 | -10558.40 | 4 | vs | 5 | 11.30 | 0.001 |
| corAR(4) | 21360.79 | 21993.41 | -10558.40 | 5 | vs | 6 | 0.00 | 0.990 |
| corCompSymm | 21403.14 | 22020.21 | -10582.57 | 6 | vs | 7 | 48.35 | <.001 |
| corExp | 21403.35 | 22020.41 | -10582.67 | | | | | |
| corGaus | 21403.35 | 22020.41 | -10582.67 | | | | | |

varConstant=constant variance; varFixed(t)=fixed variance with a function of time; varPower(t) variances with power function of time; varPower(t,l)= variances with power function of time and location; varExp(t,l)=variance with exponential function of time and location; varIdent(l)= heterogeneous variance across location over year; corSymm(UN)=unstructured correlation function; corAR =autoregressive correlation; corCompSymm=compound symmetry correlation;

#### 4.1.5.5. Results of the final fitted linear mixed model

The output of the final fitted linear mixed model is summarized in two tables (Table 10 and Table 11). These tables present the parameter estimates with their corresponding 95% confidence interval and p-value for the effect of main and interaction terms ( in both Table 10 and Table 11) , and the parameter estimate of random effects with 95% CI (Table11).

Table 11 presents significant parameter estimates for the intercept, linear and quadratic time, Agaro by quadratic time interaction effect (p-values < 0.001), and Metu by linear and quadratic time interaction effects (p-values < 0.001). Thus, the estimated parameter for intercept was

2623.77 (*with 95% CI: 2191.56, 3055.97 and P-value < 0.001*), represents an estimate of the average coffee bean yield at the average time of the study for coffee genotype G1 that grown in Jimma in the presence of biennially. The significant linear effect of time for coffee bean yield was positive (*estimate =158.92 and 95% CI: 132.04, 185.80*), revealing that there is initially increasing linear rate of growth in Jimma. However, the significant quadratic effect of time for coffee bean yield was found to be negative (*estimate =-151.51, and 95% CI: -167.43, -135.59*), suggesting that there is gradually fast decreasing linear rate of growth in later years in Jimma.

The parameter estimate for the effect of Agaro by quadratic time interaction was 85.47 (*with 95%CI: 62.96, 107.98, p-values < 0.001*), suggesting that the quadratic effects of time in Agaro is 85.47 $kgha^{1}$ greater than that of Jimma. Likewise, the parameter estimates for the effects of Metu by linear and quadratic time interaction respectively were -127.84 and 146.52 (*with 95% CI: -165.86, -89.82 and 124.01, 169.030,* respectively), suggesting that the linear effect of time in location Metu is -127.84 $kgha^{-1}$ lower than that of Jimma, but the quadratic effect of time is 188.91 $kgha^{-1}$ greater than that of Jimma.

Additionally, Table 11 shows there is significant effect of Metu and Agaro by Biennial interaction. Thus, the parameter estimate for the effect of Agaro by Biennial interaction at the average time of the study was -879.54 (*with 95% CI: -1565.75, -193.32 and p-value=0.012*) indicates that the average coffee bean yield of Agaro at the average time of the study is 879.54 $kgha^{1}$ lower than that of Jimma for coffee genotype G1 and in the absence biennially.

Table 10: Parameter estimates and their corresponding 95% CI and p-value for fixed effects from the final fitted LMM

| | Jimma | | | | Agaro | | | | Metu | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Genotype | Estimate | 95%CI | | p-value | Estimate | 95%CI | | p-value | Estimate | 95%CI | | p-value |
| In the presence of biennially | | | | | | | | | | | | |
| G2 | 169.17 | (-354.43 | 692.77) | 0.526 | -501.24 | (-1241.73 | 239.24) | 0.184 | -144.79 | (-885.27 | 595.69) | 0.701 |
| G3 | -92.15 | (-615.75 | 431.45) | 0.73 | -331.71 | (-1072.20 | 408.77) | 0.38 | -263.02 | (-1003.51 | 477.46) | 0.486 |
| G4 | -512.63 | (-1036.23 | 10.97) | 0.055 | -110.16 | (-850.64 | 630.32) | 0.77 | 446.7 | (-293.79 | 1187.18) | 0.237 |
| G5 | -239.04 | (-762.64 | 284.56) | 0.371 | -417.74 | (-1158.22 | 322.75) | 0.269 | 187.7 | (-552.78 | 928.18) | 0.619 |
| G6 | -840.21 | (-1363.81 | -316.61) | 0.002 | -315.65 | (-1056.14 | 424.83) | 0.403 | 471.01 | (-269.47 | 1211.49) | 0.212 |
| G7 | 786.67 | (263.06 | 1310.27) | 0.003 | -1093.77 | (-1834.25 | -353.28) | 0.004 | -1057.55 | (-1798.04 | -317.07) | 0.005 |
| G8 | 61.05 | (-462.55 | 584.65) | 0.819 | -791.4 | (-1531.88 | -50.91) | 0.036 | 184.95 | (-555.53 | 925.43) | 0.624 |
| G9 | 121.05 | (-402.56 | 644.65) | 0.65 | -387.09 | (-1127.57 | 353.39) | 0.305 | -657.84 | (-1398.32 | 82.64) | 0.082 |
| G10 | -671.76 | (-1195.36 | -148.16) | 0.012 | -384.01 | (-1124.49 | 356.48) | 0.309 | 130.87 | (-609.61 | 871.36) | 0.729 |
| G11 | -214.11 | (-737.71 | 309.49) | 0.423 | -401.09 | (-1141.57 | 339.40) | 0.288 | -274.42 | (-1014.91 | 466.06) | 0.467 |
| G12 | 38.19 | (-485.41 | 561.79) | 0.886 | -399.01 | (-1139.49 | 341.48) | 0.291 | -199.73 | (-940.21 | 540.76) | 0.597 |
| G13 | -409.24 | (-932.84 | 114.36) | 0.125 | -760.31 | (-1500.79 | -19.82) | 0.044 | -72.27 | (-812.76 | 668.21) | 0.848 |
| G14 | -502 | (-1025.60 | 21.60) | 0.06 | -716.27 | (-1456.75 | 24.21) | 0.058 | 293.61 | (-446.87 | 1034.09) | 0.437 |
| G15 | -484.75 | (-1008.36 | 38.85) | 0.07 | -63.39 | (-803.87 | 677.09) | 0.867 | -177.83 | (-918.31 | 562.65) | 0.638 |
| G16 | -12.56 | (-536.16 | 511.04) | 0.963 | -658.05 | (-1398.54 | 82.43) | 0.082 | -386.35 | (-1126.83 | 354.14) | 0.306 |
| G17 | 127.53 | (-396.07 | 651.13) | 0.633 | -965.96 | (-1706.45 | -225.48) | 0.011 | -374.2 | (-1114.68 | 366.29) | 0.322 |
| In the absence of biennially | | | | | | | | | | | | |
| G2 | -230.89 | (-877.91 | 416.13) | 0.484 | 869.29 | (-45.73 | 1784.31) | 0.063 | -527.54 | (-1442.56 | 387.48) | 0.258 |
| G3 | 163.44 | (-483.58 | 810.46) | 0.62 | 295.51 | (-619.51 | 1210.53) | 0.527 | -415.4 | (-1330.42 | 499.62) | 0.373 |
| G4 | 481.63 | (-165.39 | 1128.65) | 0.144 | -33.87 | (-948.89 | 881.15) | 0.942 | -1330.06 | (-2245.09 | -415.04) | 0.004 |
| G5 | -188.79 | (-835.81 | 458.23) | 0.567 | 395.15 | (-519.87 | 1310.17) | 0.397 | -650.98 | (-1566.00 | 264.04) | 0.163 |
| G6 | 765.69 | (118.68 | 1412.71) | 0.02 | 210.95 | (-704.08 | 1125.97) | 0.651 | -1415.89 | (-2330.91 | -500.86) | 0.002 |
| G7 | -746.74 | (-1393.76 | -99.73) | 0.024 | 898.55 | (-16.47 | 1813.58) | 0.054 | 285.1 | (-629.92 | 1200.12) | 0.541 |
| G8 | 169.47 | (-477.55 | 816.49) | 0.608 | 293.34 | (-621.69 | 1208.36) | 0.53 | -1335.93 | (-2250.95 | -420.91) | 0.004 |
| G9 | -24.71 | (-671.73 | 622.31) | 0.94 | 115.99 | (-799.03 | 1031.02) | 0.804 | 76.85 | (-838.17 | 991.88) | 0.869 |
| G10 | 425.15 | (-221.87 | 1072.17) | 0.198 | 447.29 | (-467.73 | 1362.31) | 0.338 | -975.11 | (-1890.13 | -60.09) | 0.037 |
| G11 | 239.37 | (-407.65 | 886.38) | 0.468 | -32.82 | (-947.84 | 882.21) | 0.944 | -257.23 | (-1172.25 | 657.79) | 0.581 |
| G12 | 22.3 | (-624.72 | 669.32) | 0.946 | 851 | (-64.02 | 1766.03) | 0.068 | -6.99 | (-922.01 | 908.03) | 0.988 |
| G13 | 192.49 | (-454.53 | 839.51) | 0.56 | 711.33 | (-203.69 | 1626.35) | 0.128 | -389.62 | (-1304.65 | 525.40) | 0.404 |
| G14 | 382.72 | (-264.30 | 1029.73) | 0.246 | 612.59 | (-302.43 | 1527.62) | 0.189 | -898.78 | (-1813.80 | 16.24) | 0.054 |
| G15 | 728.99 | (81.97 | 1376.01) | 0.027 | -151.41 | (-1066.43 | 763.61) | 0.746 | -868.63 | (-1783.66 | 46.39) | 0.063 |
| G16 | -91.88 | (-738.89 | 555.14) | 0.781 | 586.29 | (-328.73 | 1501.31) | 0.209 | -367.66 | (-1282.68 | 547.37) | 0.431 |
| G17 | -158.58 | (-805.60 | 488.44) | 0.631 | 719.79 | (-195.24 | 1634.81) | 0.123 | -539.92 | (-1454.94 | 375.10) | 0.247 |

Table 11: Parameter estimates and their corresponding 95% CI for both random effects and the remaining fixed effects (which are not presented in table 10) from the final fitted LMM

| Fixed effect | Estimate | 95% CI | | P-value |
|---|---|---|---|---|
| Intercept | 2623.77 | (2191.56 | 3055.97) | <0.001 |
| Time | 158.92 | (132.04 | 185.80) | <0.001 |
| $Time^2$ | -151.51 | (-167.43 | -135.59) | <0.001 |
| Biennial | -103.42 | (-588.65 | 381.80) | 0.676 |
| Agaro | -32.82 | (-737.65 | 672.00) | 0.918 |
| Metu | -745.35 | (-1450.18 | -40.52) | 0.040 |
| Agaro*Biennial | -879.54 | (-1565.75 | -193.32) | 0.012 |
| Metu*Biennial | -46.78 | (-732.99 | 639.43) | 0.894 |
| Agaro*Time | -0.87 | (-38.89 | 37.14) | 0.964 |
| Metu*Time | -127.84 | (-165.86 | -89.82) | <0.001 |
| Agaro* $Time^2$ | 85.47 | (62.96 | 107.98) | <0.001 |
| Metu*$Time^2$ | 146.52 | (124.01 | 169.030 | <0.001 |
| Parameters estimates of random effects with their corresponding 95% CI | | | | |
| Parameter | Estimate | | 95% CI | |
| $\sigma$ ($b_{0j}$) | 221.81 | | (129.03 | 381.28) |
| $\sigma$ ($b_{3j}$) | 145.24 | | (68.48 | 308.05) |
| corr($b_{0j,}$ $b_{3j}$) | -0.78 | | (-0.96 | -0.13) |
| $\sigma$ ($\epsilon_{tji}$) | 255.03 | | (221.51 | 293.62) |
| $\phi 1$ | -0.16 | | (-0.23 | -0.12) |
| $\phi 2$ | 0.17 | | (0.07 | 0.26) |
| $\phi 3$ | 0.15 | | (0.06 | 0.24) |
| AIC= 21358.79 | BIC= 21986.22 | logLik= -10558.4 | | |

At the average time of the study and in the presence biennially, the parameter estimate for the effect of location Metu was -745.35 (*with 95%CI: -1450.18, -40.52 and p-value=0.040*), indicates that the average coffee bean yield of Metu at the average time of the study and in the presence biennially is 745.35 $kgha^{-1}$ lower than that of Jimma for G1 genotype.

The parameter estimates for the effects of genotype, genotype*biennial, genotype*location, and genotype*location*biennial are presented in Table 10. In the presence of biennially, three coffee genotypes (G6, G7, and G10) that grown in Jimma showed significant effect on coffee bean yield compared to the reference genotype G1. Thus, the parameter estimate for coffee genotype G6 that grown in Jimma in the presence of biennially was -840.21 (*with 95% CI: -1363.81, -316.61 and P-value < 0.002*), suggests that the average coffee bean yield of coffee genotype G6

that grown in Jimma in the presence of biennially is 840.21 kgha$^{-1}$ lower than that of G1 genotype. The parameter estimate for the effect of coffee genotype G7 that grown in Jimma in the presence of biennially was 786.67 (*with 95% CI: 263.06, 1310.27 and p-value <0.003*). This indicates that the average coffee bean yield of coffee genotype G7 that grown in Jimma in the presence of biennially is 786.67 kgha$^{-1}$ greater than that of coffee genotype G1. Again, the parameter estimate for the effect of coffee genotype G10 that grown in Jimma in the presence of biennially was -671.76 (*with 95% CI: -1195.36, -148.16 and p-value < 0.012*). This suggests that the average coffee bean yield of genotype G10 that grown in Jimma in the presence of biennially is 671.76 kgha$^{-1}$ lower than that of genotype G1.

Also, in the absence of biennially, three coffee genotypes (G6, G7, and G15) that are grown in Jimma showed significant effect on coffee bean yield compared to the reference genotype (G1). Thus, the parameter estimate for coffee genotype G6 that grown in Jimma in the absence of biennial was 765.69 (with *95% CI: 118.68, 1412.71and P-value < 0.02*). This suggests that the average coffee bean yield of genotype G6 that grown in Jimma in the absence of biennially is 765.69 kg ha$^{-1}$ greater than that of genotype G1. And, the parameter estimate for the effect of coffee genotype G7 that is grown in Jimma in the absence of biennially was -746.74 (*with 95% CI:* -1393.76*,* -99.73 *and p-value <0.024*). This indicates that the average coffee bean yield of genotype G7 that grown in Jimma in the absence of biennially is 746.74 kgha$^{-1}$ lower than that of genotype G1. Again, the parameter estimate for the effect of coffee genotype G15 that grown in Jimma in the absence of biennially was 728.99 (*with 95% CI: 81.97, 1376.01 and p-value < 0.027*). This indicates that the average coffee yield of coffee genotype G15 that grown in Jimma in the absence of biennially is 728.99 kgha$^{-1}$ greater than that of genotype G1.

Besides main effects, there were also significant locations by genotype interaction at the midle of the time (Table 10). Accordingly, four genotypes in Agaro and one genotype in Metu were showed significance mean difference compared to the reference category G1 in the presence of biennially, and. From Table 10, we can observe that the effect of these genotypes were negative, suggesting that they have lower mean than the reference category G1. Thus, the parameter estimate for coffee genotype G7 in Agaro was -542.11 (*with 95% CI: -915.27, -168.94 and P-value < 0.005*) suggesting that the average coffee yield of genotype G7 in Agaro is 542.11 lower

than that of genotype G1. Following this, similar interpretation follows for the remaining significance parameter estimates in Agaro. Similarly, the parameter estimate for coffee genotype G7 in Metu was -1057.55 (*with 95% CI: -1798.04, -317.07 P-value < 0.005*) suggesting that the average coffee yield of genotype G7 in Metu is 1057.55 lower than that of genotype G1. Moreover, four genotypes in Metu were showed significance mean difference compared to the reference category G1 in the absence of biennially, and similar interpretation follows for these significant parameter estimates as usual.

### 4.1.6 Model diagnostics for the final fitted linear mixed model

#### 4.1.6 .1. Assessing assumptions of the within-group error (Assumption1)

The primary quantities used to assess the adequacy of Assumption1 are the within-group residuals, defined as the difference between the observed response and the within-group fitted value. Other quantities used for assessing Assumption 1 graphically include the within-group fitted values, the observed values, and any covariates of interest. The first residual plot that was considered in this study is the boxplot of residuals by groups. This plot is useful for verifying that the errors are centered at zero (i.e., $E(\epsilon_{ji}) = 0$), have constant variance across different groups ($Var(\epsilon_{ji}) = \sigma^2$), and are independent of the group.

In appendix, Figure 9a presents box-plots of the residuals by coffee tree (pid) for the final fitted model, and it indicates that the residuals are centered at zero, and the variability is almost the same for coffee tree except some outlying observations to the right. A better feeling for this pattern can also be judged by examining the plot of the standardized residuals versus fitted values in overall or by groups (location, block, time, and genotypes). Figure 9b&12 (in appendix) also present scatter plots of standardized residuals versus fitted values for the final fitted model in overall and by Location ,Block ,Genotype and Time. These figures indicate that the standardized residuals in each group also have about the same variability.

The assumption of normality for the within-group errors can also be assessed with the normal probability plot of the residuals, produced by the qqnorm method. Figure 10 (in appendix) presents normal plot of row (a) and standardize residual (b) for the final fitted linear mixed model, and it indicates that the random error within group follows an approximately normal

distribution. Moreover, Figure 12(in appendix) shows normal plots of residuals for each group (Location (a), Block (b), Genotype(c), and Time (d)). Again, it indicates that the assumption of normality for error term is plausible in each group.

#### 4.1.6.2. Assessing assumptions of the random effects (Assumption2)

Two types of diagnostic plots were used for assessing Assumption2 on the distribution of the random effects. These are the normal plot of estimated random effects for checking marginal normality and identifying outliers (qqnorn), and scatter plot matrix of the estimated random effects for identifying outliers and checking the assumption of homogeneity of the random effects covariance matrix. Figure 13 (in appendix) presents the normality plot of random effects of block associated with intercept and biennial for overall (a) and groups (b and c), and it does not indicate any departures from normality or serious outlying observations. The plots in Figure 14a&b are made on g1 and t1 among the levels of genotypes and time points, respectively to show homogeneity of the random effects covariance matrix. Thus, Figure 14 reveals similar patterns, and it indicates homogeneity of the random effects covariance matrix.

### 4.2. Genotype by Environment Interaction Analysis (GEI)

In the longitudinal study, the analysis was done on the actual data for the purpose of interpretation of parameter estimates and due to the possibility of handling heterogeneity of variance and the correlations among different environments. Fortunately, heterogeneity of variance and the correlations among different environments were adequately modeled, and the assumption of normality found to be believable in the linear mixed model setting.

However, before conducting any analyses of genotype by environment interaction, the data were subjected to data transformation to fix failures of assumptions of normality and homogeneity of error variances among the different environments. The box plots of coffee bean yield measurements over year in appendix (Figure 15a) shows a high degree of skewness and outliers towards high coffee yield measurements. This suggests that the data should be treated with some transformations unless the assumption of normality and constant variance may be seriously despoiled. In this study, the natural logarithm and square root transformation were checked, and the square root transformation found to be plausible transformation for coffee bean

yield measurements (Figure 15b in appendix), so that any analyses of genotype by environment interaction were done on the square root transformation.

### 4.2.1. Combined analysis of variance

After confirming the presence of significant differences among genotypes for coffee yield at the specific environments(Table 13 in appendix), combined analysis of variance was done. The combined analysis of variance in Table 12 shows that there were significant differences among environments ($p<0.001$) and genotypes ($p<0.001$) for coffee bean yield, indicating the presence of variability in genotypes as well as diversity of growing conditions at different locations. The GEI was highly significant ($p<0.001$) reflecting the differential response of genotypes in various environments. The total variation explained was 49.5% for environment, 7.2 % for genotype and 16.2% for GEI. The high percentage of the environment is an indication that the major factor that influence yield performance of coffee genotypes in Ethiopia is the environment. The percentage of variation explained by GEI was relatively large as compared to the variation explained by main effect of genotype.

Table 12: Combined ANOVA for coffee bean yield and the percentage sum of squares of the 17 genotypes tested at 21 environments (three locations over a period of seven years)

| Source | DF | SS | %SS | MS | F-value | p-value |
|---|---|---|---|---|---|---|
| Environments(E) | 20 | 126169 | 49.5 | 6308.5 | 33.6 | <0.001 |
| Block(B(E)) | 63 | 11819 | 4.6 | 187.6 | 3.3 | <0.001 |
| Genotypes(G) | 16 | 18481 | 7.2 | 1155.1 | 20.3 | <0.001 |
| Interactions(GEI) | 320 | 41208 | 16.2 | 128.8 | 2.3 | <0.001 |
| Error | 1008 | 57333 | | 56.9 | | |
| Total | 1427 | 255011 | | 178.7 | | |

### 4.2.2. Additive Main effects and Multiplicative Interaction (AMMI) analysis

The AMMI procedure has been used in order to further investigate the nature of GEI and explore the information contained in it. The result of this procedure was presented in Table 13 with the combined analysis of variance. As mentioned earlier, the environment and genotype main effects are significant, accounting for 49.5% and 7.2% of the total variation in the data set, respectively. It has also been found that 16.2% of total variation was attributed to the genotype by environment interaction.

GEI was further partitioned by principal component analysis. The Gollob F-test that has been used to measure significant of the GEI interaction components, and it shows that the first five IPCAs were significant (P-value<0.01). This indicates that the total information contained in GEI that has 320 degree of freedom can be sufficiently explained using only 155 degree of freedom which captures 80% of the total sum square of GEI, leaving only 20% of sum square of GEI as a noise.

Table 13: Combined analysis of variance (ANOVA) according to the AMMI model and Gollob's tests of interaction PCAs

| Source | DF | SS | MS | Total variation explained (%) | GEI explained (%) | Cumulative (%) |
|---|---|---|---|---|---|---|
| Total | 1427 | 255011 | 178.7 | | | |
| Treatment | 356 | 185858 | 522.1*** | 72.9 | | |
| Environments(E) | 20 | 126169 | 6308.5*** | 49.5 | | |
| Block nested in E | 63 | 11819 | 187.6*** | 4.6 | | |
| Genotypes(G) | 16 | 18481 | 1155.1*** | 7.3 | | |
| Interactions(GEI) | 320 | 41208 | 128.8*** | 16.2 | | |
| IPCA1 | 35 | 12005 | 343*** | 4.7 | 29.1 | 29.1 |
| IPCA2 | 33 | 8232 | 249.5*** | 3.2 | 20.0 | 49.1 |
| IPCA3 | 31 | 6216 | 200.5*** | 2.4 | 15.1 | 64.2 |
| IPCA4 | 29 | 3663 | 126.3*** | 1.4 | 9.0 | 73.1 |
| IPCA5 | 27 | 2852 | 105.6** | 1.1 | 7.0 | 80 |
| IPCA6 | 25 | 2220 | 88.8* | 0.9 | 5.4 | 85.4 |
| IPCA7 | 23 | 2018 | 87.8* | 0.8 | 5.0 | 90.3 |
| IPCA residuals | 117 | 4001 | 34.2 | | | |
| Error | 1008 | 57333 | 56.9 | | | |

***p-value<0.001; **p-value<0.01;*p-value<0.05 IPCA=Interaction Principal Component Axis

At 1%, Table 13 shows that these principal components (PCA1, PCA2, PCA3, PCA4 and PCA5) captured about 29.1%, 20%, 15.1%, 9% and 7% of variation due to GEI sum of squares, respectively. Together they accounted for 80% of GEI sum of squares. However, most of the variation was explained by the first three principle components (PCA1, PCA2 and PCA3) which accounted for cumulative 64.2%. Over all, the contribution of environment, genotype and the first three principal components to the treatment sum square (the sum of sum of squares of genotype, environment and GEI) was around 92%, indicating the reasonableness and parsimoniousness of AMMI model with the first three interaction principal components in partitioning the treatment sum of squares.

Estimates for the genotypic and environmental scores of AMMI-3 (scores of PCA1, PCA2 and PCA3) with their corresponding average coffee bean yield are given in Table 14. The PCA scores of a genotype from AMMI analysis indicate the stability or adaptation of a genotype across environments. The larger the PCA score, either positive or negative, as it is a relative value, the more specifically adapted a genotype is to certain environments. The closer the PCA scores near zero, the more stable or adapted a genotype is over all test environments. Environment scores from AMMI analysis relating to interaction also have meaningful interpretation. Environments with large PCA scores are more discriminating of genotypes, while environments with PCA scores near zero exhibit little interaction across genotypes and low discrimination among genotypes.

Genotype and environment combinations with PCA scores of the same signs produce positive specific interaction effect, whereas combination of opposite signs have negative specific interactions. For example, E3 and G1 have positive specific interaction effect while E2 and G2 have negative specific interaction effect. Environment which have same signs of interaction PCA scores discriminate genotypes similarly, for instance E2 and E8; and Environments with opposite sign of interaction scores discriminate genotypes differently, for example E2 and E3(Table 14).

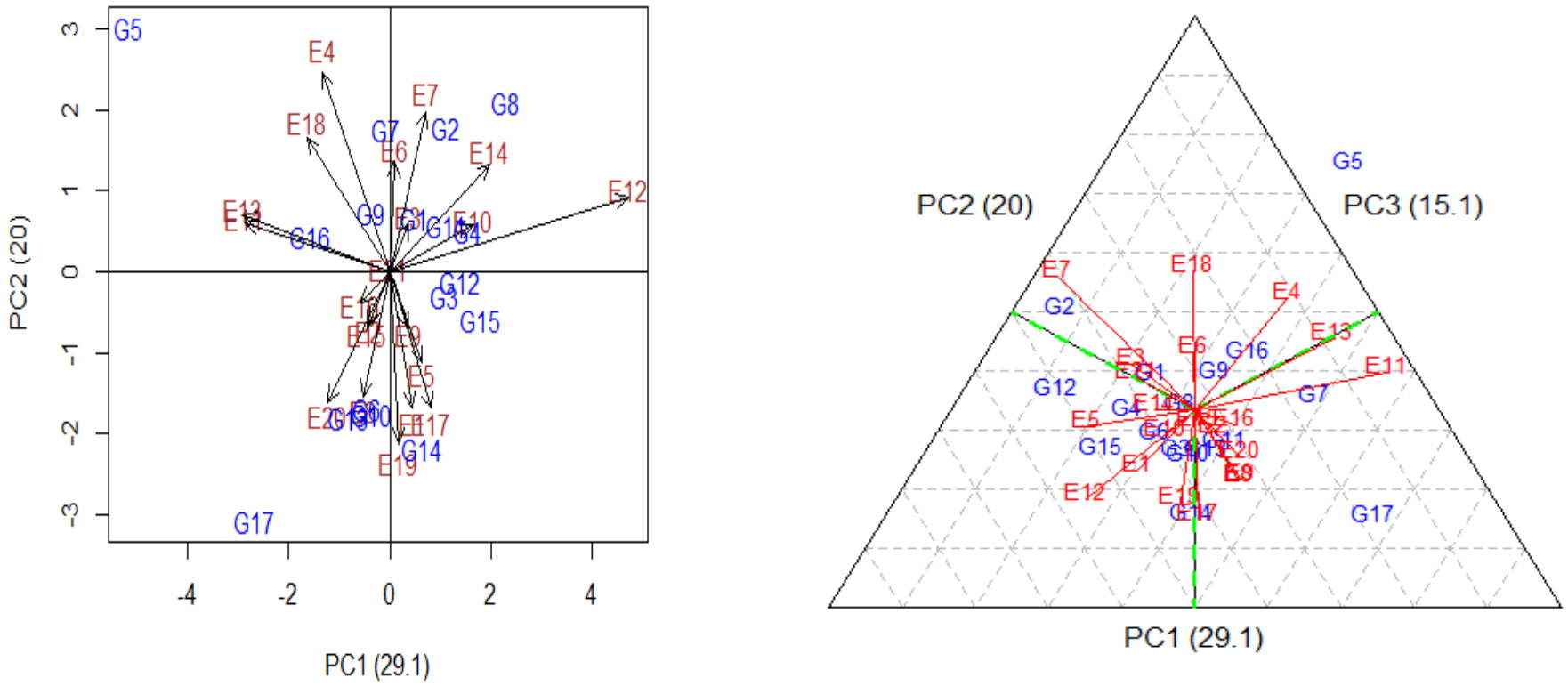
To further explain the GEI and stability, a bi-plot and three dimension plot with IPCAs scores were used. AMMI bi-plot of the first two principle component axes is a powerful way of detecting important score of GEI. This analysis represents stability of the genotypes across environments in terms of principle component analysis. It is used to see generally adapted genotypes that offer stable performance across environments, as well as genotypes that perform well under specific conditions. In this study, the first two principal component axes (PCA1 and PCA2) which capture around 50% of the total GEI sum squares in bi-plot analysis and

Table 14: IPCA1, IPCA2 and PCA3 scores for genotypes and environment with their corresponding estimated mean

| Env | Mean | PCA1 | PCA2 | PC3 | Gen | Mean | PCA1 | PCA2 | PCA3 |
|---|---|---|---|---|---|---|---|---|---|
| E1 | 21.20 | 0.451 | -1.869 | -1.140 | G1 | 41.06 | 0.499 | 0.670 | -0.680 |
| E2 | 34.08 | -0.415 | -0.723 | 0.067 | G2 | 42.23 | 1.129 | 1.774 | -2.482 |
| E3 | 33.83 | 0.367 | 0.671 | -1.336 | G3 | 37.39 | 1.093 | -0.295 | 0.617 |
| E4 | 49.91 | -1.342 | 2.738 | 1.154 | G4 | 36.52 | 1.541 | 0.507 | -0.289 |
| E5 | 28.34 | 0.632 | -1.259 | -2.328 | G5 | 34.58 | -5.174 | 3.019 | -1.004 |
| E6 | 46.95 | 0.093 | 1.516 | 0.024 | G6 | 31.69 | -0.447 | -1.696 | -1.508 |
| E7 | 32.63 | 0.707 | 2.200 | -3.002 | G7 | 39.97 | -0.077 | 1.750 | 3.170 |
| E8 | 26.36 | -0.534 | -1.714 | 0.631 | G8 | 38.85 | 2.300 | 2.087 | 1.885 |
| E9 | 36.35 | 0.365 | -0.778 | 1.585 | G9 | 39.52 | -0.356 | 0.721 | 0.199 |
| E10 | 45.07 | 1.658 | 0.635 | 0.873 | G10 | 31.07 | -0.370 | -1.754 | -0.570 |
| E11 | 49.81 | -2.874 | 0.668 | 2.345 | G11 | 36.90 | 1.115 | 0.573 | 1.898 |
| E12 | 49.10 | 4.725 | 1.021 | 1.785 | G12 | 42.55 | 1.381 | -0.130 | -2.403 |
| E13 | 51.71 | -2.891 | 0.781 | 0.847 | G13 | 31.76 | -0.820 | -1.799 | -0.542 |
| E14 | 28.70 | 1.980 | 1.476 | 0.916 | G14 | 33.79 | 0.644 | -2.182 | 0.535 |
| E15 | 29.33 | -0.456 | -0.775 | -0.351 | G15 | 36.15 | 1.803 | -0.597 | -0.767 |
| E16 | 37.38 | -0.593 | -0.422 | 0.474 | G16 | 34.44 | -1.586 | 0.438 | -0.179 |
| E17 | 31.02 | 0.834 | -1.875 | 0.943 | G17 | 33.09 | -2.674 | -3.086 | 2.122 |
| E18 | 43.71 | -1.645 | 1.837 | -1.693 | | | | | |
| E19 | 20.67 | 0.182 | -2.360 | -0.206 | | | | | |
| E20 | 40.71 | -1.242 | -1.793 | -0.004 | | | | | |
| E21 | 30.94 | -0.001 | 0.025 | -1.584 | | | | | |

Env=environment; Gen=genotype

the 3-dimensional plots (PCA1, PCA2 and PCA3) that explained about 64% of the total GEI sum of squares are presented in Figure 7. On these AMMI plots, genotypes and environment having PCA values close to zero (near the origin) have small interaction effects, whereas those having large positive or negative PCA values (distant from zero) largely contribute to GEI interaction. According to Figure 7b, G5 and G17 are relatively far apart from the origin, indicating strong interaction effects and G1, G3, G4, G5, G6 G8, G9, G10 and G12 appeared close to zero( the center of the axes) , and therefore are relatively more stable. Among the 21 environments, E7, E18, E4, E13, E11 and E12 exhibited larger interactions (i.e. they are relatively far apart from the origin) and were more discriminating of genotypes, whereas the environment E8, E9, E10, E14, E15, E16, E8, E19 and E21 relatively exhibited negligible interaction and low discrimination ( Figure 7a&b).




(a) Bi-plot

(b) 3-dimentional plot

Figure 7: Bi-plot (a) and 3-dimensional plot (b) of interaction principal components analysis (PCA): IPCA1 versus IPCA2 (a) and IPCA1 verses IPCA3 verse IPCA3 (b) for bean yield (kgha$^{-1}$) for17 coffee genotype grown in 21 environments

The bi-plot in Figure 8 presents interaction PCAs score versus mean bean yield of both coffee genotypes and environments. From the bi-plot, environments are distributed from lower yielding environments in quadrants II(top left) and III(bottom left) to the high yielding environments in quadrants I (top right) and IV (bottom right). Thus, The high yielding environments classified according to the AMMI1 model were E12, E10, E6, E4, E16, E20, E18, E11 and E13. The lower yielding environments were E19, E1, E14, E5, E17, E7, E3, E9, E21, E8, E15 and E2. The environments E19 & E1, E12, and E11 &E13 are visible in quadrant I, III, and IV, respectively, and are relatively quite distant from the origin. Accordingly, E11, E12 and E13 were the most favorable season and E19 and E1 were the less favorable seasons among the 21 environments.

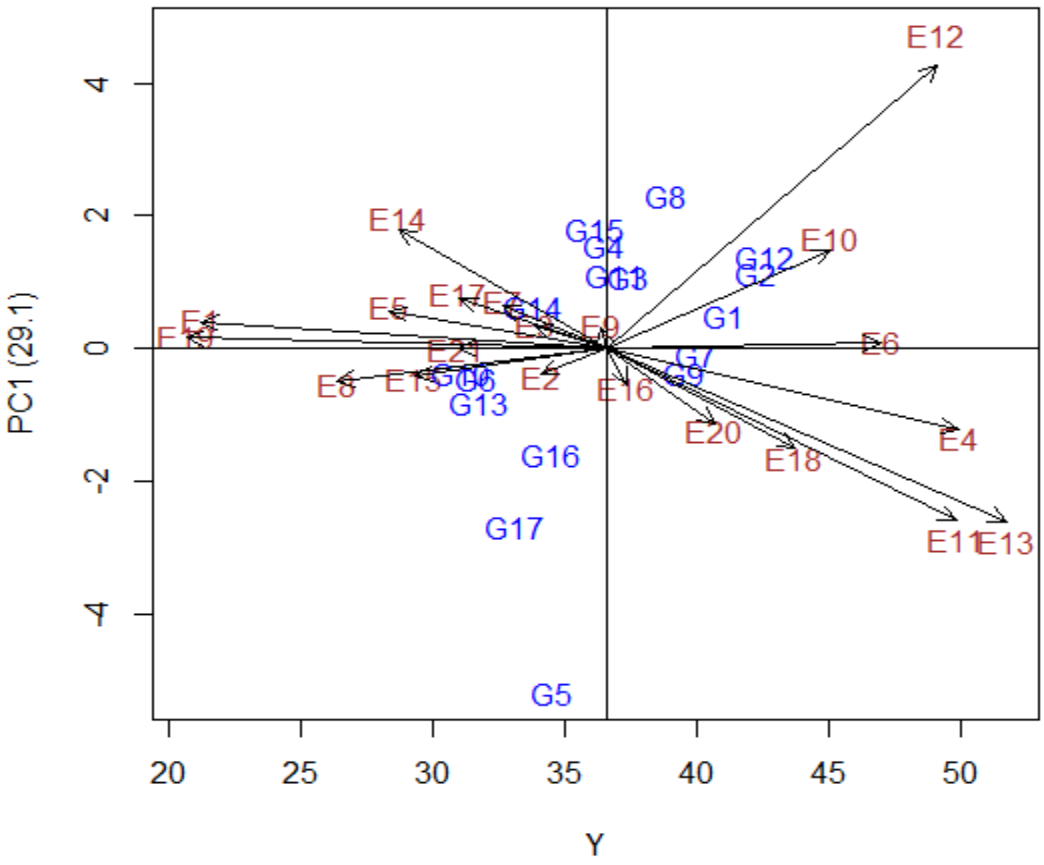



Figure 8: Bi-plot of the first interaction principal component axis (IPCA1) versus means yield for17 coffee genotype grown in 21 environments

Furthermore, the genotypes grouped under favorable environments with above average means were G1, G12, G3, G7, G8, G9 and G12. Among them, G1, G3 were found to be relative more stable. Genotypes grouped under low yielding environments are shown on the left quadrants of the bi-plot. Thus, G5 and G17 were low yielder and the most unstable genotype identified by the AMMI model.

**4.3 General Discussion**

In this study, longitudinal and GIE analysis were done to investigate the biennial effect, the correlation among measurements of coffee yield, the evolution (trend) of coffee yield over time , and the relation between genotype and environment. For longitudinal study, Exploratory data analysis (EDA) and linear mixed model (LMM) was used to analyze clustered longitudinal coffee bean yield data set, in which units of analysis (coffee tree) are nested within clusters (blocks), and repeated measures (coffee bean yield) are collected on the units of analysis over time. Correspondingly, combined ANOVA and AMMI model was used to analyze multi environment data (location-year combination) so as to extract the information contained in GEI.

In the longitudinal study, graphically, the base line covariates by time interaction, mean evolution/trajectory, and between and within coffee tree variability were investigated in the course of exploratory data analysis. Accordingly, it was shown that there is between and within coffee tree variability (Figure 1a), and the variability between clusters (blocks) was also shown Figure 1b. On the mean trend, it was shown that there is up and down trajectory in the evolution of coffee yield over time, and it was indicated that it is evident for the presence of biennial factor on coffee yield. This is more or less similar with the work (Rodriguez *et al.*, 2013) who quantified variability due to biennial in Robusta coffee. This factor was coded and used as indicator variable in LMM for the adjustment of biennial effect. It was also pointed out that the mean evolution could have a quadratic trajectory over time after the adjustment of biennial effect. The base line factors by time interaction (location*time and genotype*time) was investigated, and it was observed that there is location by time interaction (Figure 2). This suggests that there is different CBY growth trend among coffee growing areas over time of year. The variance covariance structure was also explored, and it was shown that there is variance

heterogeneity in magnitude and evolution across location over time, and a correlation among coffee yield measurements that comes from the same coffee tree.

Given the evidences in the graphical investigation and then with possible fixed and random factors , the next work was model selection for the structure of fixed and random effects, and variance covariance components based on AIC and BIC values as well as likelihood ratio test. Without considering the structure of random effects and variance covariance components, the structure of fixed effects was selected first using ML estimation method. Given the selected fixed effects structure, the structure of random effects was also selected using REML estimation method (Table 8). Thus, the random effects of block associated with intercept and biennial found to be a better fit for the random effects structure of *intercept* + *biennial* ($b_{0j} + b_{3j}B_t$).

After selecting fixed and random effect structure, functions of variance and correlation were compared for the variance and correlation structure of random error in LMM. For variance structure, nine variance functions (Table 9) were compared, and heterogeneous variance over time (varIdent(t)) found to be a better fit compared to others . Similarly, nine correlation functions (Table 9) were compared, and the autoregressive order three (AR (3)) found to be better fit. Accordingly, this study showed evidence for the presence of serial correlation among repeated measurements of Arabica coffee bean yield via significant parameter estimates of third-ordered autoregressive model (*$\phi$1= -0.16, $\phi$2=0.17, and $\phi$1=0.15 with 95% CI: (-0.23, -0.12), (0.07, 0.26), and (0.06, 0.24), respectively*). Despite the type of correlation structure, this was similar with the work of Cilas *et al*. (2011) who estimated the Compound Symmetry correlation among measurements of Robusta coffee. bean yield in successive years.

Studies shows that, the phenomenon of biennially is more pronounced in the species Arabica coffee, than Robusta coffee, which results in years with high yield intercalated with years of low yield in production(Taye *et al.*, 2001; Bernardes *et al*, 2012; Rodriguez *et al.*, 2013). This biennial alternation of yield is the result of the physiological nature of the coffee plant, which needs to vegetate along a year to sustain the fruit production in the next year (Rena and Maestri, 1985). There was no clear published literature relating to longitudinal analysis on yields of *Coffee arabica* in the linear mixed model setting including time variant factor biennial. But in

Brazil, Rodriguez *et al.* (2013) investigated the effect of biennial on the genotypes of Robusta coffee by calculating the magnitude of biennial (i.e., by subtracting the mean production of the years of low production from the mean of the years of high production based on an even number of years). The result showed high yield variation between years of high and low productions and variation among genotypes on their calculated biennial means. However, by using linear mixed model, this study revealed that, it is possible to capture the variability due to biennial in terms of fixed and random effect. Thus, the estimated variance of random effect of block associated with intercept and biennial respectively were $\hat{\sigma}^2$ $(b_{0j}) = (221.81)^2$ and $\hat{\sigma}^2$ $(b_{3j}) = 145.24^2$, and which would be benefit from using linear mixed model with time variant factor biennial. This could improve the accuracy and precision of the estimates of genotype contrasts and their standard error.

This thesis also revealed a significant location by linear and quadratic time effect interaction. From Table 11, the estimates of quadratic time effect for Jimma., Agaro and Mutu respectively were -151.51, -151.51+ 85.47=-66.05 and -151.51+146.52=-4, whereas 158.92, 158.92, 158.92-127.84=31.08 for linear time effect. Thus, for each location, the sign of the parameter estimates of linear and quadratic time effect was positive and negative, respectively. This indicates that the coffee bean yield initially increasing and gradually decreasing in linear rate of growth in all location but evolves in different magnitude. Moreover, it was shown that biennial interacts significantly with location and genotype, suggesting that differential response of genotypes and environments in the presence and absence of biennially.

Genotype by Environment Interaction (GEI) analysis was done after square root transformation of the data. The combined analysis of variance revealed that the mean squares of genotypes, environments and genotype by environment interaction were highly significant. The significance of interaction indicates that there is uncertainty in measuring overall performance of genotypes across different environments (Yonas *et al*., 2014b), or reflecting the differential response of genotypes in various environments (Girma *et al*., 2000; Zubair *et al*., 2001, as cited in Zelalem, 2011; Asnake *et al.,* 2013, as cited in Degene, 2016). The proportion of variability attributed to environment was relatively large (Table 12), and it was an indication that the major factor that influence yield performance of coffee genotypes in Ethiopia is the environment. This is in line

with the work of Lemi and Ashenafi (2016) and Yonas and Tarekegn (2015) who reported genetic variation and heritability of various traits in Arabica coffee genotypes. The magnitude of the GEI sum of squares was about 2 times larger than that of genotypes, indicating sizeable differences in genotypic response across environments, and as GEI was significant therefore we can further proceed and calculate phenotypic stability (Rashidi *et al*., 2013).

GEI was further partitioned by principal component analysis (Table 13). The Gollob's test using an approximate F-statistic revealed high significant differences for IPC1, IPC2, IPC3, IPC4 and IPC5 at 1%. The first three interaction principal components (IPC1, IPC2 and IPC3) retained by Gollob's F-test accounted for 64.2% of GEI, indicating the reasonableness and parsimony of AMMI model with the first three interaction principal component axes hereafter called AMMI3, in partitioning the treatment sum of squares effectively ((Gauch and Zobel, 1988; Gauch, 1992). This is also in line with the work of Meaza *et al.* (2011) and Yonas *et al*. (2014a) who reported the possibility of developing stable coffee genotype across environments. But the investigators showed that more than 70% of GEI sum square was explained by the first two interaction principal components. The difference could be due to the nature of the data. The current study also reported that Environments E12, E10, E6, E4, E16, E20, E18, E11 and E13 are found to be high potential environments, where genotypes having high-yield (greater than grand mean). Among 17 genotypes, G1, G2, G3, G7, G8, G9 and G12 are found to have the best performance with G1, G2, G3, G8 and G12 being relatively stable. Among the high-yielding genotypes, G7 and G9 are found to be unstable and particularly adapted to environment E4. E17 and G5 found to be low yielder and highly unstable among 17 genotypes.

# 5. CONCLUSIONS AND RECOMMENDATIONS

## 5.1. Conclusions

For coffee variety field trial conducted over year, the heterogeneous variance function (varIdent(t)) and autoregressive order three (AR3) are better fit , respectively, to the variance and correlation structure among measurements of Arabica CBY. Biennial interacts significantly with location and genotype, suggesting that differential response of genotypes and environments in the presence and absence of biennially. The CBY follows a quadratic trend with positive and negative signs, respectively, to the linear and quadratic time effect, suggesting that the Arabica CBY initially increasing and gradually decreasing in linear rate of growth.

The major factor that influence yield performance of Arabica coffee in Ethiopia is the environment. In particular, GEI highly significant and is about 2 times larger than that of genotypes, implying further proceed of extracting the information contained in GEI to investigating the nature of differential response of genotypes across environments. Among 17 genotypes, G1, G2, G3, G7, G8, G9 and G12 were identified to have the best performance with G1, G2, G3, G8 and G12 being relatively stable across the test environments under investigation using AMMI procedure. Hence, these genotypes can potentially be released for wide adaptation across coffee producing areas that have similar agro-climatic settings.

Generally, the coffee variety trial data set that conducted across location over year with RCBD design can constitute not only multi environment data set but also cluster longitudinal data set in which coffee trees are nested within blocks and repeated measures of CBY are collected on coffee trees over time. Following this, both longitudinal and GEI analysis are important to investigate the longitudinal time effect , the correlation among repeated measure of CBY, the biennial effect, and the relation between genotype and environment

## 5.2. Recommendations

The current study identified location by linear and quadratic time interaction effect which informing CBY initially increasing and gradually decreasing in linear rate of growth. Hence, Future studies should be conducted on the situation associated with the gradually decreasing rate of growth in Arabica coffee bean yield.

The current study also revealed that, the time variant factor (biennial) has significance effect on Arabica coffee bean yield. Hence, future studies that involve estimation of time effect and evolution of Arabica coffee variety should not discard this factor.

Coffee variety field trial conducted over year gives rise to repeated measurements taken on the same plot on several occasions. So, it is important to account for serial correlation among repeated measurements in such trials.

The current study also showed that GEI highly significant ($p<0.001$) accounting for 16.2% of the total sum of squares and is about 2 times larger than that of genotypes, therefore, future studies should do further analysis of GEI using not only the conventional methods but also more reliable statistical methods.

# 6. REFERENCES


Airy, G., 1861. On the algebraical and numerical theory of errors of observations and the combination of observations. Macmillan & Company, London

Alemseged, Y. and Tesfaye, S., 2012. Study on Arabica coffee fruit phenology and bean size in relation to agronomic practice. In 24th International Conference on Coffee Science San José (Costa Rica), 12th-16th November.

Allard, R. and Bradshaw, A., 1964. Implications of genotype-environmental interactions in applied plant breeding. *Crop science*, *4*(5), pp.503-508.

Amamo, A., 2014. Coffee production and marketing in Ethiopia. *Eur J Bus Manag*, *6*(3), pp.109-121.

Ammiraju, J., Jena, K., Reddy, A., Diwaker, T. and Vaidyanath, K., 2000. Identification of a molecular marker linked to a gene conferring resistance to Indian pathotype-I of Xanthomonas oryzae pv oryzae in rice. Plant Cell Biotechnology and Molecular Biology, 1(1/2), pp.63-66.and their metabolites. *Journal of agricultural and food chemistry*, *55*(17), pp.6962-6969.

Anthony, F., Bertrand, B., Quiros, O., Wilches, A., Lashermes, P., Berthaud, J. and Charrier, A., 2001. Genetic diversity of wild coffee (Coffea arabica L.) using molecular markers. *Euphytica*, 118(1), pp53-65.

Anthony, F., Combes, M., Astorga, C., Bertrand, B., Graziosi, G., and Lashermes, P., 2002. The origin of cultivated Coffea arabica L. varieties revealed by AFLP and SSR markers . *Theor. Appl. Genet*. 104(5), pp894- 900.

Asfaw, A., Tesfaye, T., Erenso, D., Taye, T., Feyera, M., Wasihun, L., Alemu, T., Andualem, W. and Chemeda .D., 2011. Genotype-by-Environment Interaction and Yield Stability Analysis in Finger Millet (*Elucine coracana* L. Gaertn) in Ethiopia. *American Journal of Plant Sciences*, *2*(3), pp408-415

Asnake, W., Mwambi, H., Temesgen, Z. and Girma, T., 2013. Additive main effects and multiplicative interactions model (AMMI) and genotype main effect and genotype by environment interaction (GGE) biplot analysis of multi-environmental wheat variety trials. *African Journal of Agricultural Research*, *8*(12), pp.1033-1040.

Babic, M, Anelkovic. V. and Babic, V. 2008. Genotype by environment interaction in maizebreeding– *Genetika*, Vol. 40, No. 3,303 -312.

Balzarini, M., 2002. Applications of Mixed Models in Plant Breeding. Quantitative Genetics, Genomics, and Plant Breeding, p.353.

Barrett, G. and Rosenberg, R. 1981. Stress effects on natural ecosystems. John Wiley & Sons, New York.

Becker, H., Léon J. 1988. Stability analysis in plant breeding. *Plant breeding*, *101*(1), pp.1-23.

Berhanu, T., M. Ali, S. Tesfaye, G. Yehenew and G. Essubalew, 2015. Impact of sun drying methods and layer thickness on the quality of highland Arabica coffee varieties at Limmu; South western Ethiopia. *Basic Research Journal*, pp: 12-20.

Bernardes, T., Moreira, M., Adami, M., Giarolla, A. and Rudorff, B., 2012. Monitoring biennial bearing effect on coffee yield using MODIS remote sensing imagery. *Remote Sensing*, *4*(9), pp.2492-2509.

Birhe, Y., 2010. The legal regime regulating coffee trade in Ethiopia. MSc Thesis in Business Law (L.L.M). Addis Ababa University, faculty of law, school of graduate studies. -11 of 112.

Blanche, S., Utomo, H., Wenefrida, I. and Myers, G., 2009. Genotype × environment interactions of hybrid and varietal rice cultivars for grain yield and milling quality. *Crop science*, *49*(6), pp.2011-2018.

Boughey, A., 1973. Ecology of populations, 2nd edition. Macmillian Publishing Co., Inc., New York.

Box, G., 1950. Problems in the analysis of growth and wear curves. *Biometrics*, *6*(4), pp.362-389.

Brady, T., Kathleen, B., and Andrzej, T., 2000. Linear mixed models: A Practical Guide Using Statistical Software, Taylor & Francis Group, LLC, London.

Chaparro, A., Christancho, M., Cortina, H. and Gaitan, A., 2004 Genetic variability of Coffea arabica L. accessions from Ethiopia evaluated with RAPDs. *Genetic Resources and Crop Evolution*, *51*(3), pp.291-297

Chauhan, R., Hooda, M. and Tanga, A., 2015. Coffee: The Backbone of Ethiopian Economy. *International Journal of Economic Plants., 1*(2), 082-086

Cilas, C., Montagnon, C. and Bar-Hen, A., 2011. Yield stability in clones of *Coffea canephora.*P in the short and medium term: longitudinal data analyses and measures of stability over time. *Tree Genetics & Genomes*, *7*(2), pp.421-429.

Costa, M., Zambolim, L. and Rodrigues, F., 2006. Effect of levels of coffee berry removals on the incidence of rust and on the level of nutrients, carbohydrates and reductor sugar. *Fitopatol. Bras*., 31, 564–571.

Crossa, J., 1990. Statistical analyses of multilocation trials. *Advances in agronomy*, 44, pp.55-85.

Crossa, J., Fox, P., Pfeiffer, W., Rajaram, S. and Gauch Jr, H., 1991. AMMI adjustment for statistical analysis of an international wheat yield trial. *Theoretical and Applied Genetics*, *81*(1), pp.27-37.

Crossa, J., Vargas, M. and Joshi, A. 2010. Linear, bilinear, and linear-bilinear fixed and mixed models for analyzing genotype environment interaction in plant breeding and agronomy. *Canadian Journal of Plant Science*. 90(5), 561-574

Cruz, C., Torres, R. and Vencovsky, R., 1989. An alternative approach to the stability analysis proposed by Silva and Barreto. *Revista Brasileira de Genética*, *12*(3), pp.567-580.Davidian, M. and Giltinan, D., 1995. Nonlinear models for repeated measurement data CRC press, Lendon. 361p

Daviron, B. and Ponte, S., 2005. The coffee paradox: Global markets, commodity trade and the elusive promise of development. Zed books.

Davis, A., Tosh, J., Ruch, N. and Fay, M., 2011. Growing coffee: Psilanthus (Rubiaceae subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of Coffea. *Botanical Journal of the Linnean Society*, *167*(4), pp.357-377.

Davis, L., 1957. Flowering and alternate bearing. *Proc. A mer. Soc.Hart. Sci.* 70:545-556.

de Resende, M., Thompson, R. and Welham, S., 2006. Multivariate spatial statistical analysis of longitudinal data in perennial crops. *Rev. Mat. Estat*, *24*(1), pp.147-169.

Dean, A., 1995. A molecular investigation of genotype by environment interactions. *Genetics. 139*(1), pp.19-33.

Dejene, D., 2016. Genotype by Environment Interaction and Yield Stability Analysis of Ethiopian Bread Wheat Using Mixed Model. An MSc thesis presented to school of graduate studies of Haramaya University, 94p.

Eberhart, S. and Russell, W., 1966. Stability parameters for comparing varieties. *Crop science*, *6*(1), pp.36-40.

ECEA., 2008. Understanding Commodities to be traded at Ethiopia Commodity Exchange, Policy Analysis and Economic Research Team, Addis Ababa, Ethiopia

Finlay, K. and Wilkinson, G., 1963. The analysis of adaptation in a plant-breeding programme. *Crop and Pasture Science*, *14*(6), pp.742-754.

Fitzgmaurice, G., Laird, N. and Ware, J., 2004. Applied longitudinal analysis. Hoboken.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. eds., 2008. Longitudinal data analysis. CRC Press.

Gauch Jr, H. and Zobel, R., 1988. Predictive and postdictive success of statistical analyses of yield trials. *Theoretical and Applied Genetics*, *76*(1), pp.1-10.

Gauch Jr, H., 1988. Model selection and validation for yield trials with interaction. *Biometrics*, *44(3)*, pp.705-715.

Gauch Jr, H., 1992. AMMI analysis on yield trials. *CIMMYT Wheat Special Report (CIMMYT)*.

Gauch, H., 2013. A simple protocol for AMMI analysis of yield trials. *Crop Science*, *53*(5), pp.1860-1869.

Geisser, S. and Greenhouse, S., 1958. An extension of box's results on the use of the $ F $ distribution in multivariate analysis. *The Annals of Mathematical Statistics*, *29*(3), pp.885-891.

Getachew, W., Sentayehu, A., Taye, K. and Tadesse, B. (2013), Genetic Diversity Analysis of Some Ethiopian Specialty Coffee (*Coffea arabica* L.) Germplasm Accessions Based on Morphological Traits*, Time Journals of Agriculture and Veterinary Sciences, 1*(4), pp47-54.

Gichuru, E., Agwanda, C., Combes, M., Mutitu, E., Ngugi, E., Bertrand, B. and Lashermes, P., 2008. Identification of molecular markers linked to a gene conferring resistance to coffee berry disease (Colletotrichum kahawae) in Coffea arabica. *Plant pathology*, *57*(6), pp.1117-1124.

Girma, N., 2011. Marketing information operation in Ethiopia with special reference to the Ethiopia Commodity Exchange (ECX) Coffee Trading. An MSc. thesis. Presented in Swedish University of Agricultural sciences Uppsala, Sweden. Department of Urban and Rural development. XX, pp75.

Girma, T., 2010. Spatial Modelling Techniques To Improve Data Analysis of Field Trials. OmniScriptum GmbH & Co. KG, USA.

Gole, T., Denich, M., Teketay, D. Vlek, P., 2002. Human impacts on Coffea arabica gene pools in Ethiopia and the need for its in situ conservation. In: R. Rao, A. Brown, M. Jackson (eds.) Managing Plant Genetic Diversity. CABI and IPGRI, 237-247.

Gollob, H., 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, *33*(1), pp.73-115.

Gomez, K. and Gomez, A., 1984. Statistical procedures for agricultural research. John Wiley & Sons, Tornoto, Canada

Gómez-Ruiz, J., Leake, D. and Ames, J., 2007. In vitro antioxidant activity of coffee compounds and their metabolites. *Journal of agricultural and food chemistry*, *55*(17), pp.6962-6969.

Gómez-Ruiz, J., Leake, D. and Ames, J., 2007. In vitro antioxidant activity of coffee compounds and their metabolites. *Journal of agricultural and food chemistry*, *55*(17), pp.6962-6969.

Gray, Q., Tefera, A. and Tefera, T., 2013. Ethiopia: Coffee Annual Report. GAIN Report No. ET 1302.

Greenhouse, S. and Geisser, S., 1959. On methods in the analysis of profile data. *Psychometrika*, 24(2), pp.95-112.

Gurka, M., 2006. Selecting the best linear mixed model under REML. *The American Statistician*, *60*(1), pp.19-26.

Harville, D., 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*(358), pp.320-338.

Hilten, J., 2002.CoffeeProduction. Retrieve from www.worldmapper.org/posters/worldmapper_1037_**coffee_production**_ver5.pd

Hu, Q. and Buyanovsky, G., 2003. Climate Effects on Corn Yield in Missouri. *Journal of Applied Meteorology*, *42*(11), pp.1626-1635.

Institute of Biodiversity Conservation (IBC), 2008. Country report on the state of plant genetic resources for food and agriculture, Addis Ababa.

International Coffee Organization (ICO), 2015. Sustainability of the coffee sector in Africa, International Coffee Council, 115th Session ,2-6 March 2015, London, united kingdom

International Trade Centre (ITC)., 2010. Climate Change and the Coffee, Industry, ITC, Geneva, vi, 28 pages (Technical paper)

Iscaro, J., 2014. The Impact of Climate Change on Coffee Production in Colombia and Ethiopia. *Global Majority E-Journal*, *5*(1), p.33.

Jaramillo, J., Setamou, M., Muchugu, E., Chabi-Olaye, A., Jaramillo, A., Mukabana, J., Maina, J., Gathara, S. and Borgemeister, C., 2013. Climate change or urbanization? Impacts on a traditional coffee production system in East Africa over the last 80 years. *PloS one*, *8*(1), pp51815.

Jennrich, R. and Schluchter, M., 1986. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, *42*(4), pp.805-820.

Jonkers, H. 1979. Biennial bearing in apple and pear: a literature survey. *Scientia Hart.* *11*(4), pp303-317.

Kassa, L., Smith, M. and Fufa, H., 2006. Stability analysis of grain yield of tef (Eragrostis tef) using the mixed model approach. *South African Journal of Plant and Soil*, *23*(1), pp.38-42.

Kassahun, T., Tamiru, O., Kim. G., Endeshaw, B. and Thomas, B., 2008. Genetic diversity and population structure of wild Coffea arabica population in Ethiopia using molecular marker. In: Girma, A., Bayetta, B., Tesfaye., S., Endale., T and Taye., K. (eds.). Coffee Diversity and Knowledge. Proceedings of a National Workshop Four Decades of Coffee Research and Development in Ethiopia, 14-17 August 2007, Addis Ababa, Ethiopia. pp.35-49.

Kodama, Y., 2009. The effects of fair trades on Coffee producers: A case study of Ethiopian Coffee Cooperatives. Procceedings of the 16 th International Conference of Ethiopian Studies ed. by Svein Ege, Harald Aspen, Birhanu Teferra and Shiferaw Bekele, Trondheim 2009.

Koebler, J., 2013. Buzzkill? How Climate Change Could Eventually End Coffee. USNews.com, US News & World Report (27 March 2013)

Kwanchai, A. and Arturo, A. 1984. Statistical Procedures for Agricultural Research 2nd ed. Johan Wiley & sons, Inc., New York. pp.166

Laird, N. and Ware, J., 1982. Random-effects models for longitudinal data. *Biometrics*, *38*(4), pp.963-974

Lande, R. and Shannon, S., 1996. The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, *50*(1), pp.434-437.

Lemi, B. and Ashenafi, A., 2016. Genetic Variability, Heritability and Genetic Advance for Yield and Yield Components of Limmu Coffee (Coffea arabica L.) Accessions in South Western Ethiopia. *Middle-East Journal of Scientific Research*, *24*(6), pp.1913-1919.

Lentner, M. and Bishop, T. (1993) Experimental design and analysis (SecondEdition). Valley Book Company, Blacksburg, VirginiaCochran, WG. and Cox, DR. (1957) Experimental designs (Second Edition). John Wiley andSons, Inc., New York.

Lin, C., Binns, M. and Lefkovitch, L., 1986. Stability analysis: where do we stand? *Crop science*, 26(5), pp.894-900.

Lone, A., Khan, M., Warsi, M. and Nehvi, F., 2009. Ammi Analysis for Anthesis Silking Interval Stability of Maize (Zea Mays L.) Under Normal and Excess Soil Moisture Conditions. *Crop Improvement*, *36*(2), pp.18-21.

Luquet, D., Dingkuhn, M., Kim, H., Tambour, L. and Clement-Vidal, A., 2006. EcoMeristem, a model of morphogenesis and competition among sinks in rice. 1. Concept, validation and sensitivity analysis. *Functional Plant Biology*, *33*(4), pp.309-323.

Mather, K. and Caligari, P., 1976. Genotype x environment interactions IV. The effect of the background genotype. *Heredity, 36*(1): 41-48.

Meaza , D., Mesfin , K. and Girma, T., 2011. Additive Main Effects and Multiplicative Interaction Analysis of Coffee Germplasms from Southern Ethiopia. *SINET: Ethiopian Journal of Science*, *34*(1), pp.63-70.

Milosav, B., Anelkovi, V. and Vojka, B., 2008. Genotype by environment interaction in maize breeding. *Genetika*, *40*(3), pp.303-312.

Ministry of Trade. 2012. Coffee Opportunities in Ethiopia. Presentation of the Ministry of Trade of the FDRE. February 2012, Addis Ababa.

MoARD (Ministry of Agriculture and Rural Development), 2007. Agriculture Marketing Sector, Coffee. The gift of Ethiopia to the world. Retrieve from http://etd.aau.et/dspace/bitstream

Modur, S., 2010. Missing Data Methods for Clustered Longitudinal Data (Doctoral dissertation, The Ohio State University)

Morrell, C., 1998. Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics*, *54*(4) pp.1560-1568.*1560*

Nakagawa, S. and Schielzeth, H., 2013. A general and simple method for obtaining $R^2$ from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), pp.133-142.

Nuhu, A., 2014. Bioactive micronutrients in coffee: recent analytical approaches for characterization and quantification. ISRN nutrition, 2014. Package in Yergacheffe District, Gedeo Zone, SNNP Regional State, Ethiopia.

Paramaguru, Kharunya (2012). Coffee Under Threat: How Wild Arabica Could Go Extinct. Time.com. TIME, news story (November 16, 2012)

Pendegrast, M. (1999). Uncommon Grounds: The History of Coffee and how it Transformed our World. Basic books, New York, USA.

Petit, N., 2007. Ethiopia's coffee sector: A bitter or better future?. *Journal of Agrarian Change*, 7(2), pp.225-263.

Picini, A., Camargo, M., Ortolani, A., Fazuoli, L. and Bollergallo, P., 1999. Test and analysis of agrometeorological models for predicting coffee yield. *Bragantia*, 58(1), pp157–170.

Piepho, H. and Eckl, T., 2014. Analysis of series of variety trials with perennial crops. *Grass and Forage Science*, *69*(3), pp.431-440.

Piepho, H., 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics*, *97*(1-2), pp.195-201.

Pinheiro, J. and Bates, D., 2000. Mixed-Effects Models in S and S-PLUS, Springer-Verlag, Berlin

Ponte, S., 2002. The latte revolution'? Regulation, markets and consumption in the global coffee chain. *World development*, *30*(7), pp.1099-1122.

Quintin, G.,Abu,T. and Taddy,T,.2013. Global agricultural net work, USDA Agricultural service report, ET-1302

Rashidi, M., Farshadfar, E. and Jowkar, M., 2013. AMMI analysis of phenotypic stability in chickpea genotypes over stress and non-stress environments. *International Journal of Agriculture and Crop Sciences*, *5*(3), p.253.

Ramirez-Villegas, J., Salazar, M., Jarvis, A. and Navarro-Racines, C., 2012. A way forward on adaptation to climate change in Colombian agriculture: perspectives towards 2050. *Climatic Change*, *115*(3-4), pp.611-628.

Rashidi, M., Farshadfar, E. and Jowkar, M., 2013. AMMI analysis of phenotypic stability in chickpea genotypes over stress and non-stress environments. *International Journal of Agriculture and Crop Sciences*, *5*(3), p.253

Rodrigues, P., Malosetti, M., Gauch, H. and van Eeuwjik, F., 2014. A weighted AMMI algorithm to study genotype-by-environment interaction and QTL-by-environment interaction.*Crop Science*, *54*(4), pp.1555-1570.

Rodrigues, W., Tomaz, M., Ferratilde, R., Ferratilde, M., da Fonseca, A. and Martins, L., 2013. Crop yield bienniality in groups of genotypes of conilon coffee. *African Journal of Agricultural Research*, *8*(34), pp.4422-4426.

Schabenberger, O., 2004. Mixed model influence diagnostics in proceedings of the twenty-ninth annual SAS users group international conference. 189, p. 29. In Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference.

Scheff´e, H. 1959. The Analysis of Variance. Wiley, New York

Scheffe, H., 1956. Alternative models for the analysis of variance. The Annals of Mathematical Statistics, pp.251-271.

Scheffe, H., 1999. The analysis of variance (Vol. 72). John Wiley & Sons, Toronto, Canada. 479p.

Shukla, G., 1972. Some statistical aspects of partitioning genotype environmental components of variability. *Heredity*, *29*(2), pp.237-245.

Singh, L. 1948. Studies in biennial bearing. II. A review of the litera-ture. *J. Hort. Sci.* 24, pp45-65.

Singh, R., 1971. Biennial bearing in fruit trees-accent on apple and mango. *Ind. Counc. Agr. Res. Bul. 30.*

Smith, A., Cullis, B. and Thompson, R., 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, *57*(4), pp.1138-1147.

Sugiura, N., 1978. Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's. *Communications in Statistics-Theory and Methods*, *7*(1), pp.13-26.

Taye, K., Ashenafi, A., Alemseged, Y., Teshome, K. and Wondiyfraw, T., 2011a. The contribution of coffee research for coffee seed development in Ethiopia.*E3 Journal of Agricultural Research and Development*, *1*(1), pp.009-016.

Taye, K., Tesfaye, S., *Alemseged*, Y., *Anteneh*, N. and *Endale*, T., 2001b. The impact of close spacing on yield of arabica coffee under contrasting agroecologies of Ethiopia. *Afr. Crop Sci. J*. 9, pp.401-409.

Tesfaye, K., 2006. Genetic Diversity of wild Cofea arabica populations in Ethiopia as a contribution to conservation and use planning. Ecology and Development Series. University of Bonn,Germany. No.44.

Tesfaye, K., 2006. Genetic Diversity of Wild Coffea arabica Populations in Ethiopia as a Contribution for Conservation and Use Planning. Ecology and Development Series, No.44, Centre for Development Research, University of Bonn.

Tesfaye, S., Alemseged, Y., Anteneh, N., Endale, T. and Taye, K. 2008. Fruit physiology and factors affecting bean sizes in Arabica coffee. Coffee diversity and Knowledge Workshop. EIAR, Addis Ababa, p.163-168

Tesfaye. S., Taye., K. and Alemseged,Y. 2002. The effect of established shade trees on the growth and yield of arabica coffee in two planting patterns. In Proceedings of the 19th International Conference on Coffee Science (ASIC), May 14th-18th, 2001, Trieste, Italy.

Toler, J. and Burrows, P., 1998. Genotypic performance over environmental arrays: a non-linear grouping protocol. *Journal of Applied Statistics*, *25*(1), pp.131-143.

Verbeke, G. and Molenberghs, G., 2009. Linear mixed model for longitudinal data. Springer Science & Business Media, New York, USA. 567p

West, B., Welch, K. and Galecki, A., 2014. Linear mixed models: a practical guide using statistical software. CRC Press, Ann Arbor, Michigan. 414p

Yates, F. and Cochran, W., 1938. The analysis of groups of experiments. *The Journal of Agricultural Science*, *28*(4), pp.556-580.

Yonas, B. and Tarekegn, A. 2015. Genetic Variation and Extent of Heritability of the Various Agronomic Traits of Arabica Coffee (*Coffea Arabica* L.) Genotypes Grown across different Environments in Ethiopia. *Greener Journal of Plant Breeding and Crop Science*, *3*(1), pp. 1-10

Yonas, B., 2005. Genotype by environment interaction and stability analysis of coffee (*Coffea arabica* L.) bean yield and yield related traits An MSc thesis presented to school of graduate studies of Haramaya University.

Yonas, B. and Bayetta, B., (2008). Genotype by environment interaction and Stability analysis of Arabica genotypes In: Girma, A., Bayetta, B., Tesfaye., S., Endale., T and Taye., K. (eds.). Coffee Diversity and Knowledge. Proceedings of a National Workshop Four Decades of Coffee Research and Development in Ethiopia, 14-17 August 2007, Addis Ababa, Ethiopia. pp. 71-79.

Yonas, B., Bayeta, B. and Fininsa, C., 2014a. Stability Analysis of Bean Yields of Arabica Coffee Genotypes across Different Environments. *Greener Journal of Plant Breeding and Crop Science*, *2*(2), pp. 018-026

Yonas, B., Bayeta, B. and Fininsa, C., 2014b. Performance evaluation of indigenous Arabica coffee genotypes across different environments. *Journal of Plant Breeding and Crop Science*, *6*(11), pp.171-178.

Zambolim, L., do Vale, F., Costa, H., Pereira, A. and Chaves, G. , 2002. Epidemiology and Integrated Control of Coffee Rust. In The State of the Art Technology in the Production of Coffee; Zambolim, L., Ed., UFV: Viçosa, MG., Brazil, pp. 369–450.

Zeger, S. and Liang, K., 1986. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*(1), pp.121-130.

Zelalem, T., 2011. Statistical analysis of genotype by environment interactions and grain yield stability in bread wheat using ANOVA and AMMI models. An MSc thesis presented to school of graduate studies of Addis Ababa University, 86p.

Zobel, R.W., Wright, M.J. and Gauch, H.G., 1988. Statistical analysis of a yield trial. *Agronomy Journal*, *80*(3), pp.388-393.

# 7. APPENDIX

**Appendix A. Figures**

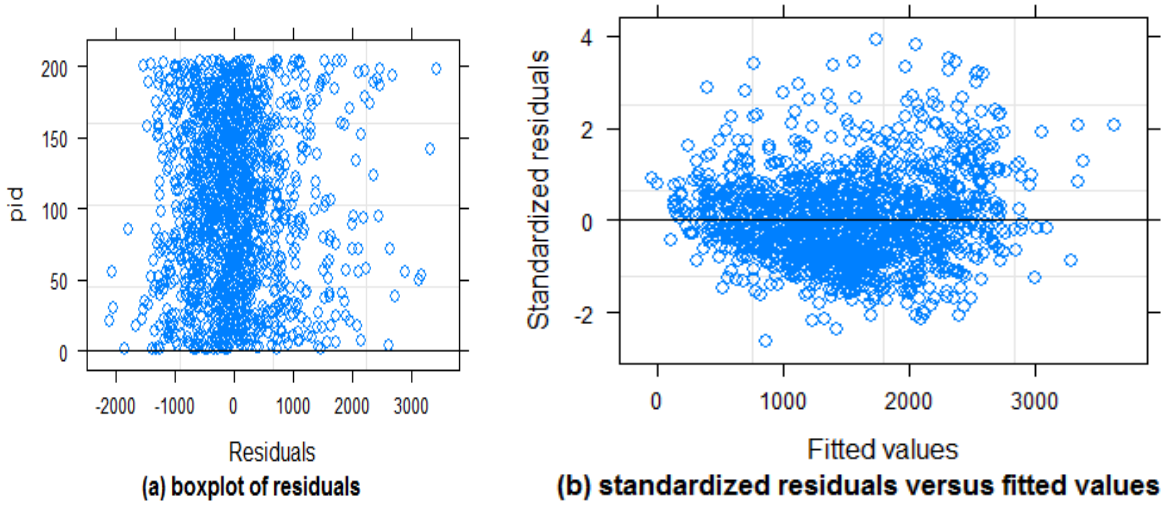



Figure 9: Boxplots of the residuals by subject(coffee trees/pid) (a) and standardized residuals versus fitted values(b) for the final fitted linear mixed model

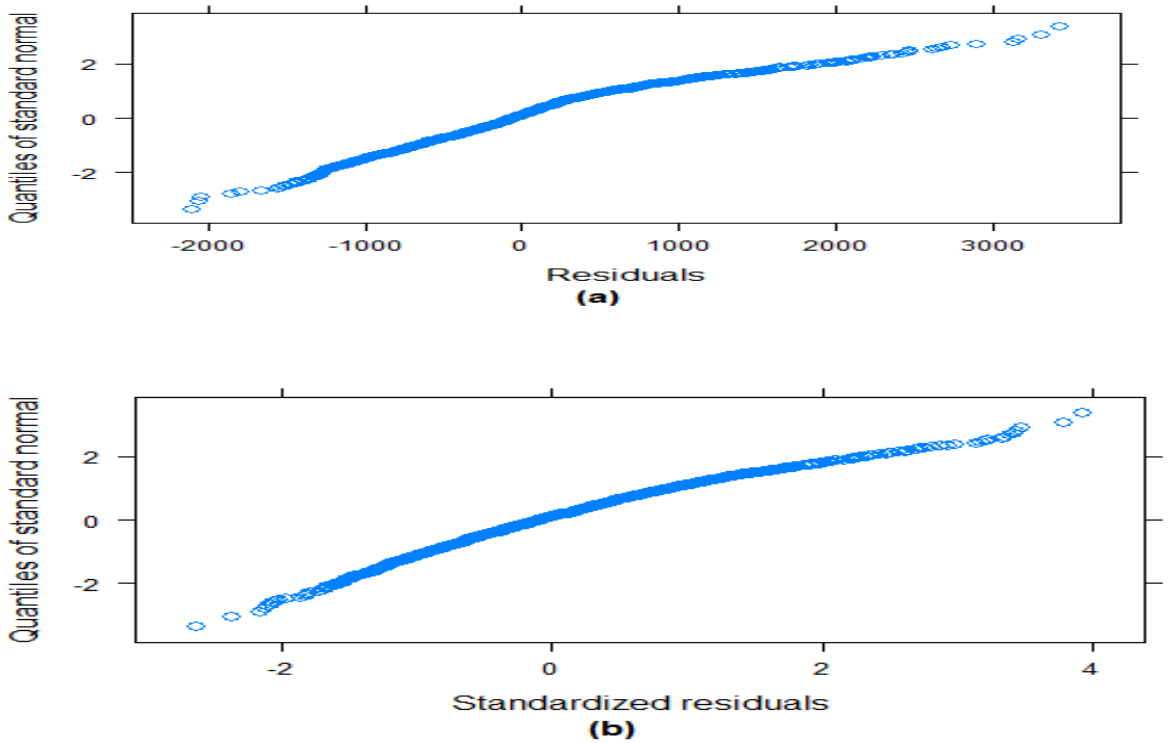



Figure 10: Normal plot of residuals (a) and standardize residuals (b) for the final fitted linear mixed model

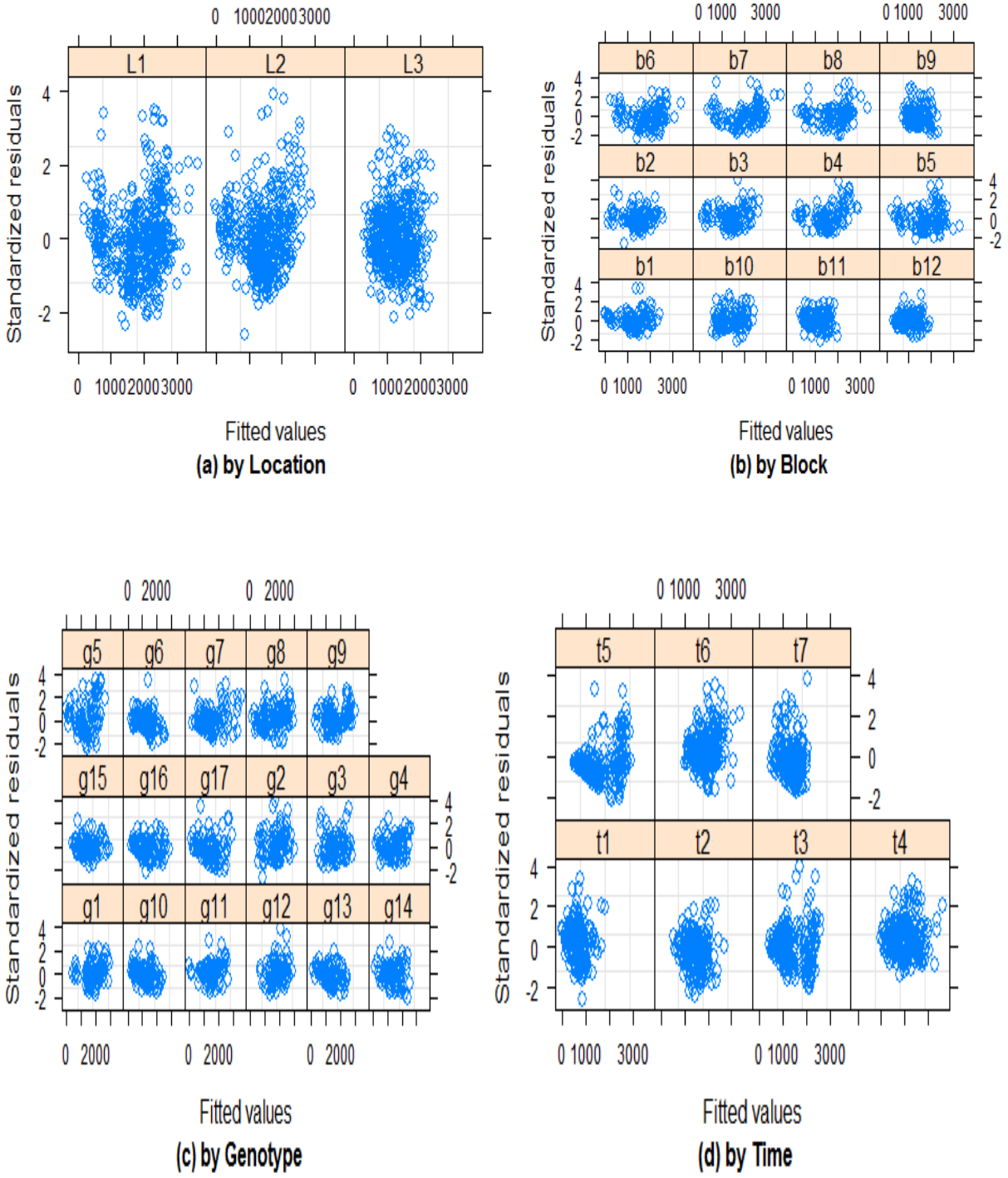



Figure 11: Scatter plots of standardized residuals versus fitted values for the final fitted linear mixed model by Location (a), Block (b), Genotype(c), and Time (d)

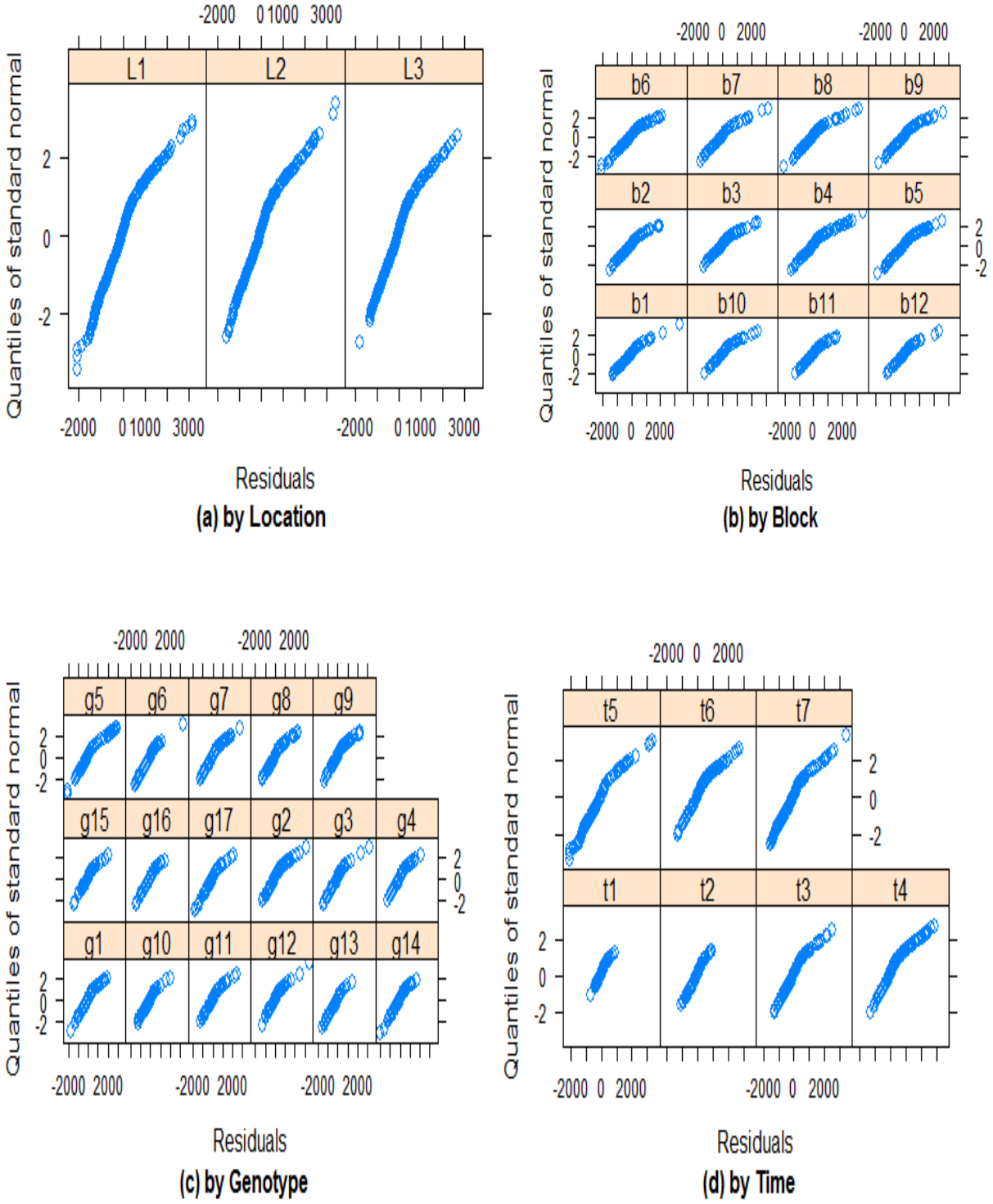



Figure 12: Normal plot of residuals for the final fitted linear mixed model by Location (a), Block (b), Genotype(c), and Time (d).

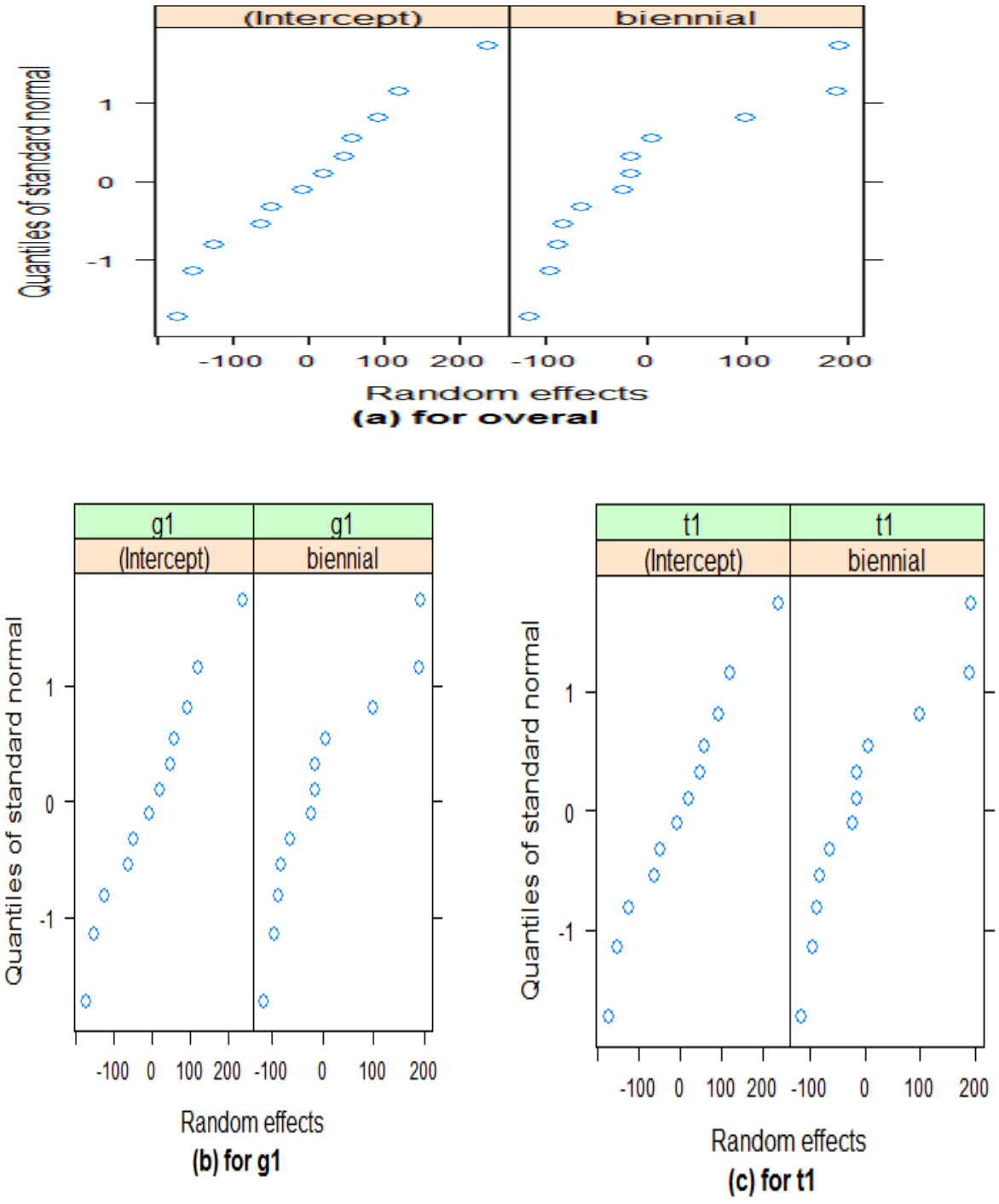



Figure 13: Normal plot of random effects of block associated with intercept and biennial: for overall (a) and for groups (example, for genotype1 (G1) (b) and for time1 (at year1) (c))

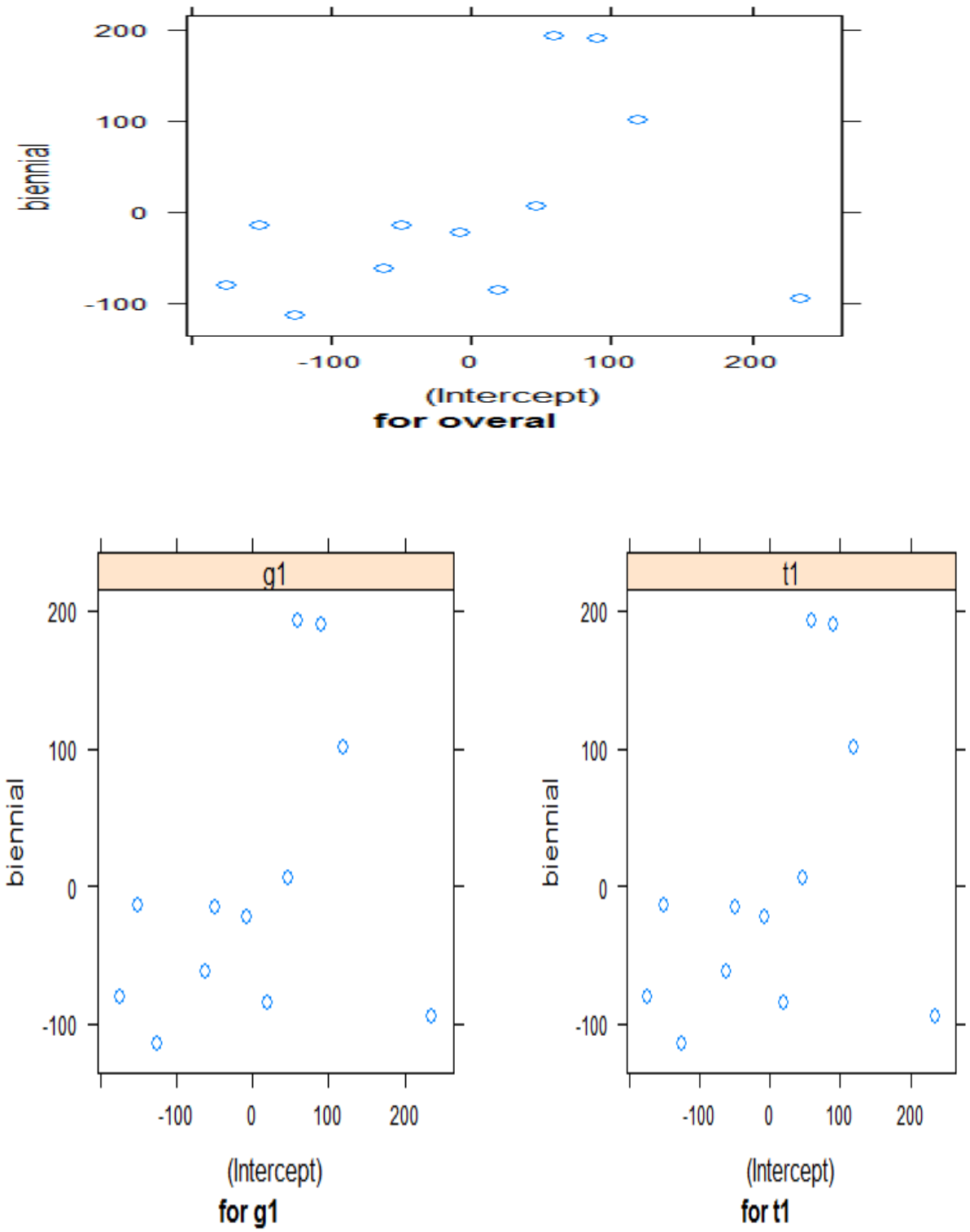
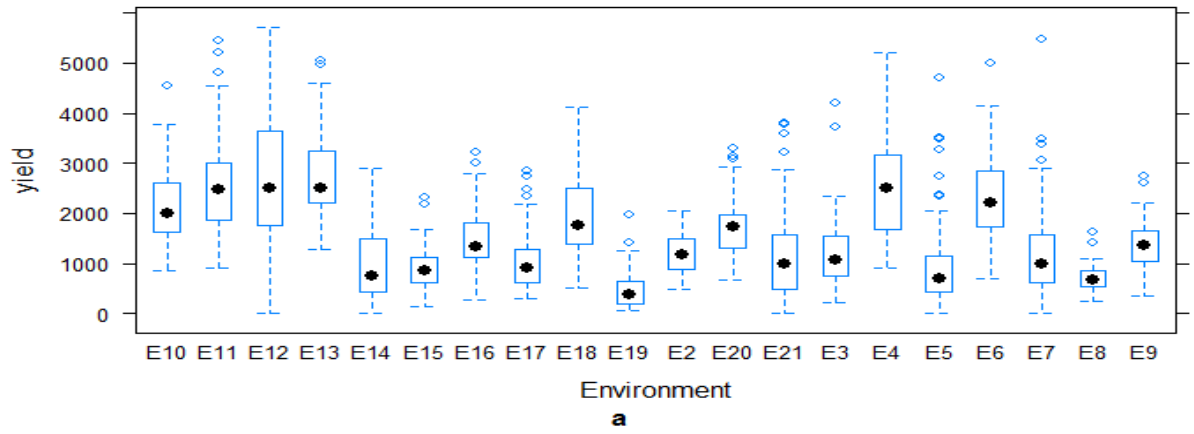
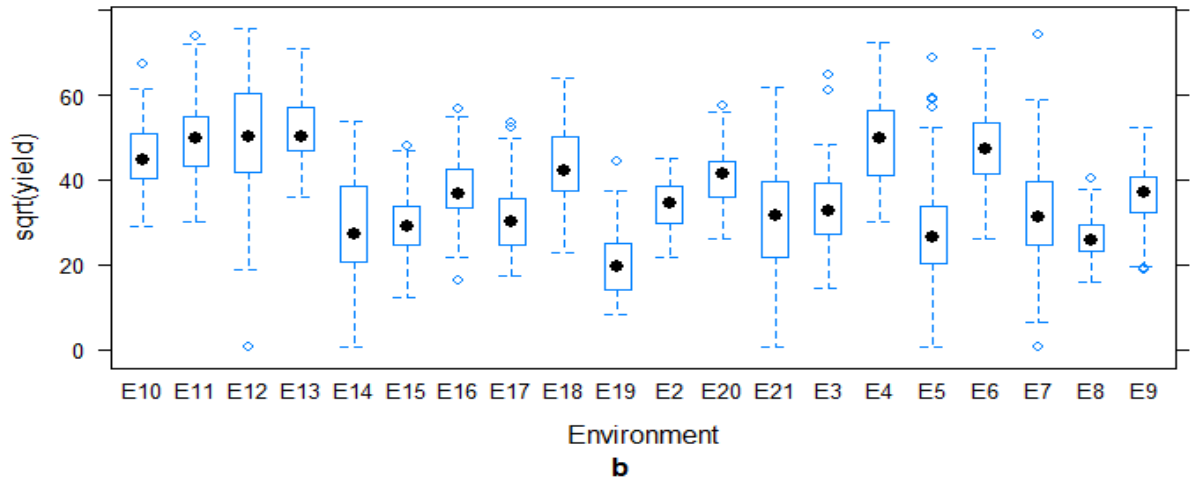



Figure 14: Scatter plot of random effects: for overall (a) and for groups (b) and (c)



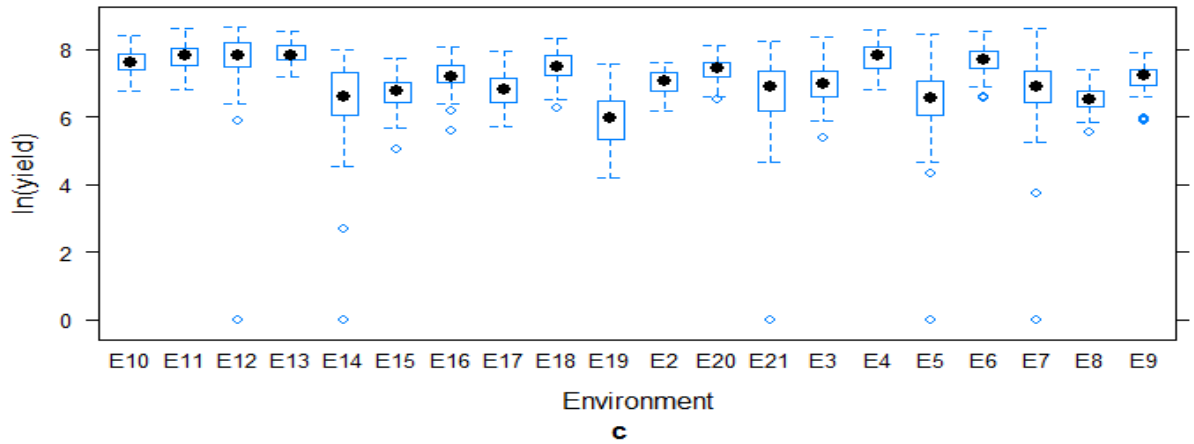




Figure 15: Box-plot of coffee yield measurements :(a) actual yield, (b) Square root transformation and (c) Logarithm transformation

**Appendix B Tables**

Table 15: Parameter estimates and their corresponding standard error, t-value, degree of freedom (DF) and p-value for fixed effects from the final fitted LMM

| | Jimma | | | Agaro | | | Metu | | |
|---|---|---|---|---|---|---|---|---|---|
| | In the presence of biennially | | | | | | | | |
| | estimate | t-value | p-value | estimate | t-value | p-value | estimate | t-value | p-value |
| G2 | 169.17 | 0.63 | 0.526 | -501.24 | -1.33 | 0.184 | -144.79 | -0.38 | 0.701 |
| G3 | -92.15 | -0.35 | 0.730 | -331.71 | -0.88 | 0.380 | -263.02 | -0.70 | 0.486 |
| G4 | -512.63 | -1.92 | 0.055 | -110.16 | -0.29 | 0.770 | 446.70 | 1.18 | 0.237 |
| G5 | -239.04 | -0.90 | 0.371 | -417.74 | -1.11 | 0.269 | 187.70 | 0.50 | 0.619 |
| G6 | -840.21 | -3.15 | 0.002 | -315.65 | -0.84 | 0.403 | 471.01 | 1.25 | 0.212 |
| G7 | 786.67 | 2.95 | 0.003 | -1093.77 | -2.90 | 0.004 | -1057.55 | -2.80 | 0.005 |
| G8 | 61.05 | 0.23 | 0.819 | -791.40 | -2.10 | 0.036 | 184.95 | 0.49 | 0.624 |
| G9 | 121.05 | 0.45 | 0.650 | -387.09 | -1.03 | 0.305 | -657.84 | -1.74 | 0.082 |
| G10 | -671.76 | -2.52 | 0.012 | -384.01 | -1.02 | 0.309 | 130.87 | 0.35 | 0.729 |
| G11 | -214.11 | -0.80 | 0.423 | -401.09 | -1.06 | 0.288 | -274.42 | -0.73 | 0.467 |
| G12 | 38.19 | 0.14 | 0.886 | -399.01 | -1.06 | 0.291 | -199.73 | -0.53 | 0.597 |
| G13 | -409.24 | -1.53 | 0.125 | -760.31 | -2.01 | 0.044 | -72.27 | -0.19 | 0.848 |
| G14 | -502.00 | -1.88 | 0.060 | -716.27 | -1.90 | 0.058 | 293.61 | 0.78 | 0.437 |
| G15 | -484.75 | -1.82 | 0.070 | -63.39 | -0.17 | 0.867 | -177.83 | -0.47 | 0.638 |
| G16 | -12.56 | -0.05 | 0.963 | -658.05 | -1.74 | 0.082 | -386.35 | -1.02 | 0.306 |
| G17 | 127.53 | 0.48 | 0.6329 | -965.96 | -2.56 | 0.011 | -374.20 | -0.99 | 0.322 |
| Df | | 1311 | | | 1311 | | | 1311 | |
| Std.Error | | 266.90 | | | 377.46 | | | 377.46 | |
| | In the absence of biennially | | | | | | | | |
| G2 | -230.89 | -0.70 | 0.484 | 869.29 | 1.86 | 0.063 | -527.54 | -1.13 | 0.258 |
| G3 | 163.44 | 0.50 | 0.620 | 295.51 | 0.63 | 0.527 | -415.40 | -0.89 | 0.373 |
| G4 | 481.63 | 1.46 | 0.144 | -33.87 | -0.07 | 0.942 | -1330.06 | -2.85 | 0.004 |
| G5 | -188.79 | -0.57 | 0.567 | 395.15 | 0.85 | 0.397 | -650.98 | -1.40 | 0.163 |
| G6 | 765.69 | 2.32 | 0.020 | 210.95 | 0.45 | 0.651 | -1415.89 | -3.04 | 0.002 |
| G7 | -746.74 | -2.26 | 0.024 | 898.55 | 1.93 | 0.054 | 285.10 | 0.61 | 0.541 |
| G8 | 169.47 | 0.51 | 0.608 | 293.34 | 0.63 | 0.530 | -1335.93 | -2.86 | 0.004 |
| G9 | -24.71 | -0.07 | 0.940 | 115.99 | 0.25 | 0.804 | 76.85 | 0.16 | 0.869 |
| G10 | 425.15 | 1.29 | 0.198 | 447.29 | 0.96 | 0.338 | -975.11 | -2.09 | 0.037 |
| G11 | 239.37 | 0.73 | 0.468 | -32.82 | -0.07 | 0.944 | -257.23 | -0.55 | 0.581 |
| G12 | 22.30 | 0.07 | 0.946 | 851.00 | 1.82 | 0.068 | -6.99 | -0.01 | 0.988 |
| G13 | 192.49 | 0.58 | 0.560 | 711.33 | 1.53 | 0.128 | -389.62 | -0.84 | 0.404 |
| G14 | 382.72 | 1.16 | 0.246 | 612.59 | 1.31 | 0.189 | -898.78 | -1.93 | 0.054 |
| G15 | 728.99 | 2.21 | 0.027 | -151.41 | -0.32 | 0.746 | -868.63 | -1.86 | 0.063 |
| G16 | -91.88 | -0.28 | 0.781 | 586.29 | 1.26 | 0.209 | -367.66 | -0.79 | 0.431 |
| G17 | -158.58 | -0.48 | 0.631 | 719.79 | 1.54 | 0.123 | -539.92 | -1.16 | 0.247 |
| Df | | 1311 | | | 1311 | | | 1311 | |
| Std.Error | | 329.8124 | | | 466.4252 | | | 466.4252 | |

Table 16: Parameter estimates and their corresponding standard error, t-value, degree of freedom (DF) and p-value for the remaining fixed effects (which are not presented in Table 15) from the final fitted LMM

| | Estimate | Std.Error | DF | t-value | p-value |
|---|---|---|---|---|---|
| Intercept | 2623.77 | 220.32 | 1311 | 11.91 | 0.000 |
| Time | 158.92 | 13.70 | 1311 | 11.60 | 0.000 |
| $Time^2$ | -151.51 | 8.11 | 1311 | -18.67 | 0.000 |
| Biennial | -103.42 | 247.34 | 1311 | -0.42 | 0.676 |
| Agaro | -32.82 | 311.57 | 9 | -0.11 | 0.918 |
| Metu | -745.35 | 311.57 | 9 | -2.39 | 0.040 |
| Agaro*Biennial | -879.54 | 349.79 | 1311 | -2.51 | 0.012 |
| Metu*Biennial | -46.78 | 349.79 | 1311 | -0.13 | 0.894 |
| Agaro*Time | -0.87 | 19.38 | 1311 | -0.05 | 0.964 |
| Metu*Time | -127.84 | 19.38 | 1311 | -6.60 | 0.000 |
| Agaro* $Time^2$ | 85.47 | 11.47 | 1311 | 7.45 | 0.000 |
| Metu*$Time^2$ | 146.52 | 11.47 | 1311 | 12.77 | 0.000 |

Table 17: Individual (separate) analyses of variance (RCB design) for a trial with 17 genotypes and four replications, by 21 environment

| Env | DF | Source of variation | SS | MS | Env | Source of Variation | DF | SS | MS |
|---|---|---|---|---|---|---|---|---|---|
| E1 | 3 | rep | 207.77 | 69.26 | E12 | rep | 3 | 485.96 | 161.99 |
| | 16 | genotype | 1417.41 | 88.59 | | genotype | 16 | 9243.10 | 577.69 |
| | 48 | error | 1281.19 | 26.69 | | error | 48 | 4974.71 | 103.64 |
| E2 | 3 | rep | 503.74 | 167.91 | E13 | rep | 3 | 131.80 | 43.93 |
| | 16 | genotype | 909.71 | 56.86 | | genotype | 16 | 2617.10 | 163.57 |
| | 48 | error | 757.81 | 15.79 | | error | 48 | 1584.25 | 33.01 |
| E3 | 3 | rep | 56.07 | 18.69 | E14 | rep | 3 | 19.06 | 6.35 |
| | 16 | genotype | 3639.98 | 227.50 | | genotype | 16 | 4501.82 | 281.36 |
| | 48 | error | 1694.69 | 35.31 | | error | 48 | 4219.84 | 87.91 |
| E4 | 3 | rep | 1592.81 | 530.94 | E15 | rep | 3 | 823.94 | 274.65 |
| | 16 | genotype | 3934.33 | 245.90 | | genotype | 16 | 1732.46 | 108.28 |
| | 48 | error | 1637.93 | 34.12 | | error | 48 | 789.36 | 16.45 |
| E5 | 3 | rep | 271.86 | 90.62 | E16 | rep | 3 | 778.59 | 259.53 |
| | 16 | genotype | 3335.88 | 208.49 | | genotype | 16 | 1007.06 | 62.94 |
| | 48 | error | 8015.22 | 166.98 | | error | 48 | 2170.95 | 45.23 |
| E6 | 3 | rep | 574.82 | 191.61 | E17 | rep | 3 | 527.61 | 175.87 |
| | 16 | genotype | 3163.40 | 197.71 | | genotype | 16 | 2303.85 | 143.99 |
| | 48 | error | 2121.14 | 44.19 | | error | 48 | 2106.89 | 43.89 |
| E7 | 3 | rep | 779.03 | 259.68 | E18 | rep | 3 | 115.10 | 38.37 |
| | 16 | genotype | 5982.80 | 373.93 | | genotype | 16 | 3191.81 | 199.49 |
| | 48 | error | 4189.70 | 87.29 | | error | 48 | 1742.99 | 36.31 |
| E8 | 3 | rep | 211.92 | 70.64 | E19 | rep | 3 | 253.71 | 84.57 |
| | 16 | genotype | 313.44 | 19.59 | | genotype | 16 | 1580.55 | 98.78 |
| | 48 | error | 860.88 | 17.94 | | error | 48 | 2262.19 | 47.13 |
| E9 | 3 | rep | 277.54 | 92.51 | E20 | rep | 3 | 423.2512 | 141.084 |
| | 16 | genotype | 1955.02 | 122.19 | | genotype | 16 | 326.5068 | 20.4067 |
| | 48 | error | 868.11 | 18.09 | | error | 48 | 2888.346 | 60.1739 |
| E10 | 3 | rep | 406.14 | 135.38 | E21 | rep | 3 | 2536.725 | 845.575 |
| | 16 | genotype | 2935.22 | 183.45 | | genotype | 16 | 2089.682 | 130.605 |
| | 48 | error | 1621.66 | 33.78 | | error | 48 | 9946.011 | 207.209 |
| E11 | 3 | rep | 841.44 | 280.48 | | | | | |
| | 16 | genotype | 3507.79 | 219.24 | | | | | |
| | 48 | error | 1599.54 | 33.32 | | | | | |

Df = degree of freedom; SS= sum square; MS mean square; Env =environment

## Appendix C. R code

#LONGITUDINAL ANALYSIS

```
library(foreign)
library(lattice)
library(nlme)
library(faraway)
library(MASS)
rm(list=ls())
my<- read.csv(file="C:\\Users\\user\\Desktop\\\\Tdata\lmm.csv")
attach(my)
View(my)
```

##Exploratory data analysis

```
#############exploring the individual profile ##############
xyplot (y ~ t0, data = m, type = "a", group = pid, xlab = "Time in year"
      , col.line = "gray20",ylab = " Yield(kg/ha)", sub="b")

xyplot (y ~ t0|bid, data = m, type = "a", group = pid, xlab = "Time in year"
      , col.line = "gray20",ylab = " Yield(kg/ha)", sub="b")

####################exploring the mean profile####################
interaction.plot(t0,Location,y , fun=mean, lwd = 3, xlab= "Time in year",
            ylab= "   Yield(kg/ha)", las=1,sub="a")
interaction.plot(t0,Genotype,y , fun=mean,  lwd = 3, xlab= "Time in year",
            ylab= "yield(kg/ha)", las=1, sub="b")

mean1<-tapply(y, t0, mean)
age1<-as.numeric(unique(t0))
plot(age1,mean1,type= "l",ylim=c(0,3000),col="1",lwd = 2,xlab="Time in Year",ylab="Coffee
grain yield(kg/ha) ")

####################smoothing plot####################

 title(main="Mean profile plot of coffee yield by genotype using loess smoothing")
plot(t0,y,col = "gray50",lwd = 1, pch = ".", main="General mean profile plot of coffe yield by
using loess smoothing ",
            xlab = "Time(years) since base line",ylab = "Coffee    yield  (kg/hectar")
             with(m,{lines(loess.smooth (t0,y ,family = "gaussian"),
             lwd = 4,lty = 1,col = 1)})

mean1<-tapply(y, t0, mean)
age1<-as.numeric(unique(t0))
plot(age1,mean1,type= "l",ylim=c(0,3000),col="1",lwd = 2,xlab="Time in Year",ylab="Coffee
grain yield(kg/ha) ")
```

```
#############Exploring Correlation##############
par(mfrow=c(3,3))
CD4.lm <- lm (y ~ t, data = m)
m$lmres <- resid (CD4.lm)
m$roundyr <- round(m$t)

## Reshape the data to wide format
CD4w <- reshape(m[,c("pid", "lmres", "roundyr")],
direction = "wide",
v.names = "lmres", timevar = "roundyr",
idvar = "pid")
## Put histograms on the diagonal
panel.hist <- function(x, ...) {
usr <- par("usr"); on.exit(par(usr))
par(usr = c(usr[1:2], 0, 1.5) )
h <- hist(x, plot = FALSE)
breaks <- h$breaks;
nB <- length(breaks)
y <- h$counts;
y <- y/max(y)
rect(breaks[-nB], 0, breaks[-1], y, col="cyan", ...)}
## Put (absolute) correlations on the upper panel, w/ size prop. to correlation.
panel.cor <- function(x, y, digits=2, prefix="", cex.cor) {
usr <- par("usr"); on.exit(par(usr))
par(usr = c(0, 1, 0, 1))
r <- abs (cor(x, y, use = "pairwise.complete.obs"))
txt <- format(c(r, 0.123456789), digits=digits)[1]
txt <- paste(prefix, txt, sep="")
if(missing(cex.cor)) cex <- 0.8/strwidth(txt)
text(0.5, 0.5, txt, cex = cex * r)
}
pairs (CD4w[,c(2:8)], upper.panel = panel.cor,diag.panel = panel.hist)

##################correlation of the observed data################
t=as.factor(t)
t1<-y[Time==1]
t2<-y[Time==2]
t3<-y[Time==3]
t4<-y[Time==4]
t5<-y[Time==5]
t6<-y[Time==6]
t7<-y[Time==7]
distance1<-cbind(t1,t2,t3,t4,t5,t6,t7)
cov(distance1)
panel.hist <- function(x, ...)
{
```

```
usr <- par("usr"); on.exit(par(usr))
par(usr = c(usr[1:2], 0, 1.5) )
h <- hist(x, plot = FALSE)
breaks <- h$breaks; nB <- length(breaks)
y <- h$counts; y <- y/max(y)
rect(breaks[-nB], 0, breaks[-1], y, col="cyan", ...)
}
pairs(distance1, panel=panel.smooth, cex = 1.5, pch = 24,
bg="light green",diag.panel=panel.hist,
cex.labels = 2, font.labels=2)
```

## ##Model building

```
#########selection of fixed effect structure##############
f0=gls(y~l*g*(bb +t + t2),
     data=my,method = "ML")
f1=stepAIC(f0,direction="backward")

###############selection of random effects####################
rb0=gls(y ~ l + g + bb + t + t2 + l:g + l:bb + l:t + l:t2 +
          g:bb + l:g:bb,data=my,method = "REML")
rb1=lme(y ~ l + g + bb + t + t2 + l:g + l:bb + l:t + l:t2 +
          g:bb + l:g:bb,data=my,method = "REML",random=list(bid=pdSymm(~1)))
rb2=update(rb1,random=list(bid=pdSymm(~bb)),weights=varIdent(~1|l))
rb3=update(rb1,random=list(bid=pdSymm(~bb+t)),weights=varIdent(~1|l))
rb4=update(rb1,random=list(bid=pdSymm(~bb+t+t2)),weights=varIdent(~1|l))
anova(rb0,rb1,rb2,rb3,rb4)
rp1=update(rb2,random=list(bid=pdSymm(~bb),pid=~1))
rp2=update(rb2,random=list(bid=pdSymm(~bb),pid=pdSymm(~bb)),weights=varIdent(~1|l))
rp3=update(rb2,random=list(bid=pdSymm(~bb),pid=pdSymm(~bb+t)),weights=varIdent(~1|l))
rp4=update(rb2,random=list(bid=pdSymm(~bb),pid=pdSymm(~bb+t+t2)),weights=varIdent(~1|l
))
anova(rb2,rp1,rp2,rp3,rp4)
#################selection of the variance function###################
c=lmeControl(maxIter=50000, msMaxIter=200, tolerance=1e-4, niter=50,
     msTol=1e-5, nlmStepMax=500,msVerbose=TRUE,returnObject=TRUE)
v1=update(rb2,weights=varFixed(~t0),control=c)
v2=update(rb2,weights=varPower(form=~t0),control=c)
v3=update(rb2,weights=varPower(form=~t0|l),control=c)
v4=update(rb2,weights=varExp(form=~t0),control=c)
v5=update(rb2,weights=varExp(form=~t0|l),control=c)
v6=update(rb2,weights=varIdent(form=~1|l),control=c)
v7=update(rb2,weights=varIdent(5,form=~1|t0),control=c)
v8=update(rb2,weights=varIdent(5,form=~1|t0*l),control=c)
anova(rb2,v1,v2,v3,v4,v5,v6,v7,v8)
###################selection of the correlation structure############
```

```
c1=update(v7,corr=corSymm(form=~1|bid/pid))
c2=update(v7,corr=corAR1(-0.5,form=~1|bid/pid))
c3=update(v7,corr=corARMA(c(-0.3,0.2),form=~1|bid/pid,p=2))
c4=update(v7,corr=corARMA(c(-0.2,0.15,0.1,0.09),form=~1|bid/pid,p=3))
c5=update(v7,corr=corARMA(c(-0.2,0.15,0.1,0.09),form=~1|bid/pid,p=4))
c6=update(v7,corr=corCompSymm(form=~1|bid/pid))
c7=update(v7,corr=corExp(form=~t0|bid/pid))
c8=update(v7,corr=corGaus(form=~t0|bid/pid))
anova(v7,c1,c2,c3,c4,c5,c6,c7,c8)
########final fitted model##############
c4=lme(y ~ l + g + bb
 + t + t2 + l:g + l:bb + l:t + l:t2 +
          g:bb + l:g:bb,data=my,method = "REML",random=list(bid=pdSymm(~biennial)),
        weights=varIdent(5,form=~1|t0),control=c, corr=corARMA(c(-
0.2,0.15,0.1),form=~1|bid/pid,p=3))
```

## ##Model diagnosis

```
plot( c4, pid~resid(.), abline = 0, sub="(a) boxplot of residuals" )
plot( c4, sub="(b) standardized residuals versus fitted values ")

plot( c4, resid(., type = "p") ~ fitted(.)|l,id = 0.000001, adj = -0.3, sub="(a) by Location")
plot( c4, resid(., type = "p") ~ fitted(.)|bid,id = 0.000001, adj = -0.3, sub="(b) by Block")
plot( c4, resid(., type = "p") ~ fitted(.)|g,id = 0.000001, adj = -0.3, sub="(c) by Genotype")
plot( c4, resid(., type = "p") ~ fitted(.)|ts,id = 0.000001, adj = -0.3, sub="(d) by Time")

qqnorm(c4,~resid(.),sub="(a)")
qqnorm(c4,sub="(b)")

qqnorm(c4,~resid(.)|l,sub="(a) by Location")
qqnorm(c4,~resid(.)|bid,sub="(b) by Block")
qqnorm(c4,~resid(.)|g,sub="(c) by Genotype")
qqnorm(c4,~resid(.)|ts,sub="(d) by Time")

qqnorm( c4,~ranef(.,level = 1)|biennial,id =0.05)
pairs( c4, ~ranef(.,augFrame=T), id = 0.1, adj=-0.5)
```

## #GEI ANALYSIS

```
################## data transformation ##############################

library(agricolae)
my<- read.csv(file="C:\\Users\\user\\Desktop\\Tdata\\ammi.csv")
attach(my)
View(my)
bwplot( y~ ENV, data=m, xlab="Block",ylab="y",
       sub="a")
```

```
bwplot( sy~ ENV, data=m, xlab="Block",ylab="sqrt(y)",
       sub="b")
bwplot( ly~ ENV, data=m, xlab="Block",ylab="ln(y)",
       sub="c")

########################Ammi analysis####################

model<- with(my,AMMI(ENV, GEN, REP, y0, console=T,PC=T))
model$analysis
# see help(plot.AMMI)
# biplot
plot(model)
# triplot PC 1,2,3
plot(model, type=2, number=F)
# biplot PC1 vs Yield
plot(model, first=0,second=3, number=F)
pc(model)
```