academicresearchJournals

Vol. 6(7), pp. 370-379, September 2018
DOI: 10.14662/ARJASR2018.049

ISSN: 2360-7874
http://www.academicresearchjournals.org/ARJASR/Index.htm

Academic Research Journal of Agricultural Science and Research

Full Length Research

Longitudinal Analysis of Arabica Coffee Bean Yield: Application of Linear Mixed Model for Clustered Longitudinal Data

Tarekegn Argaw[1], Girma Taye[2], Dechasa Bedada[3], Ashenafi Ayano[4]

[1]Biometrics, Ethiopian Institute of Agricultural Research, P. O. Box 2003, Addis Ababa, Ethiopia
[2]Associate Professor, School of Public Health, Addis Ababa University, P. O. Box 9086, Addis Ababa, Ethiopia
[3]Department of Statistics, College of Natural Science, Jimma University, P.O. Box 378, Jimma, Ethiopia
[4] Ethiopian Institute of Agricultural Research ,Jimma Agricultural Research Center, P.O. Box, 192, Jimma, Ethiopia
Corresponding author's E-mail: tare.aragaw@gmail.com

Accepted 26 September 2018

The study aimed to do longitudinal analysis to investigate the effect of time, *biennial*, and correlation on Arabica coffee bean yield by using Exploratory Data Analysis (EDA) and Linear Mixed Model (LMM). The data for this study came from coffee variety field trials conducted by Jimma Agricultural Research Center (JARC) over 7 years during 2005-2011 in south west Ethiopia across 3 coffee growing areas (Jimma, Agaro, and Metu). The experimental design of the trial was RCBD with 4 replications and 17 Arabica coffee genotypes. The LMM results revealed that the heterogeneous variance function (varIdent) and autoregressive order three (AR3) were, respectively, found to give better fit to the variance and correlation structure among measurements of coffee bean yield. 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. It was observed that the measurements of coffee bean yield obtained from Arabica coffee tree over time induced an autocorrelation which is known as serial correlation. There was initially an increasing and gradually a decreasing trend in Arabica coffee bean yield over time/years with linear rate of growth. There was also a differential response of genotypes and environments in the presence and absence of *biennially*. The effects of correlation among measurements, time, and *biennial* have to be considered in Arabica coffee breeding research to improve the precision and accuracy of research outcomes by using advanced statistical models.

Key Words: Arabica Coffee, Biennial, Clustered Longitudinal Data

Cite this article as: Argaw, T., Taye, G., Bedada, D., Ayano, A. (2018). Longitudinal Analysis of Arabica Coffee Bean Yield: Application of Linear Mixed Model for Clustered Longitudinal Data. Acad. Res. J. Agri. Sci. Res. 6(7): 370-379

INTRODUCTION

Arabica coffee (*Coffeaarabica* 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).

Economically, coffee is the second most exported commodity after oil, and employs over 100 million people worldwide (Gray *et al*., 2013). Coffee is a perennial crop with more than 124 species of which Arabica coffee is economical important (Gichuru et al, 2008). In fact, 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).

Field trials with perennial crops give rise longitudinal measurements taken on the same plot on several occasions (Piepho and Eckl, 2014). 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). Like annual crops, coffee breeders generate multi-location trial data over year to evaluate the yield performance of coffee genotypes across location over year. The statistical methods which commonly used to analyze such data are 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). 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. 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).

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*., 2001). It also acutely reported that such improved technologies can be obtained through rigorous breeding procedures and efficient statistical design and modeling (Girma, 2010). However, 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.

DATA AND METHODS

The data for this study came from coffee variety field trials conducted by Jimma Agricultural Research Center (JARC) over several years in south west Ethiopia. The field trial was conducted across three locations (Jimma, Agaro and Metu). These locations have different soil type and altitudes and could also possibly be differentiated with their mean seasonal rainfall and temperature. Seven year coffee bean yield data collected during 2005 to 2011 were used in this study. These yield measurements were obtained from a total of 204 coffee trees with 7measurements per coffee tree (over 7 year's period).

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

Exploratory data analysis

Before mixed model analysis, exploratory data analysis was used to explore 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 results of such exploration were used in order to choose a fixed-effects structure for linear mixed model.

Linear mixed model

According to Modur (2010), a linear mixed effect model for cluster longitudinal coffee bean yield data set given as

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

[1]

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 ϵ_{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 [1] can be represented as follows:

$$\begin{cases} Y_{ji} = X_{ji}\beta + Z_{ij}b_{ji} + \epsilon_{ji} \\ b_{ji} = (\mathrm{M_i}{}^{t}\mathrm{C_{ji}}{}^{t})^{t} \sim \mathrm{MVN}(0, \mathrm{D} = \mathrm{D_M} \oplus \mathrm{D_C}) \\ \epsilon_{ji} = MVN(0, R_{ji}) \end{cases} \quad [2]$$

Variance and correlation functions for modeling heteroscedasticity and dependency

Variance and correlation functions are used to model the variance and correlation 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). Table 1 shows the most standard variance correlation function classes which are built in R computing statistical package

Method of parameter estimations

Maximum likelihood estimation and restricted maximum likelihood estimation are the two commonly used

Table 1: Standard correlation function classes

Name	Expression
Standard correlation function classes	
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
Standard variance function classes	
varFixed	fixed variance
varIdent different	variances per stratum
varPower	power of covariate
varExp	exponential of covariate

methods of estimations in linear mixed model. ME estimation also provides estimators of the fixed effects. 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.

Model selection

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:

$$LRT - 2\log(L\_nested/Lreference) = -2\log(L\_nested) - (-2\log 〖(L\_reference))\sim 〖x\text{^}2〗\_df 〗 \quad [3]$$

In Equation [3], 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. 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. The likelihood ratio tests that we use to test linear hypotheses about fixed-effect parameters in an LMM are based on ML estimation. When testing hypotheses about covariance parameters in linear mixed model, REML estimation should be used for both the reference and nested models.

Another set of tools useful in model selection are referred to as information criteria. 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)

$$AIC = -2\ x\ l(\hat{\beta},\hat{\theta}) + 2p[4]$$

In Equation 4, 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:

$$BIC = -2l(\hat{\beta},\hat{\theta}) + p\ x\ ln(n)[5]$$

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.

Computing soft ware

All analysis were done with the help of R soft ware

RESULT AND DISCUSSION

The base line (year 1) is the time when the age of coffee trees was 5 years after planted on the field. 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.

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.

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. On the mean trend, it was shown that there is up and down trajectory in the evolution of coffee yield over time, and it is evident for the presence of biennial effect on coffee yield. This factor was coded and used as indicator variable in LMM for the adjustment of biennial effect.

Figure 3 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 3). This suggests that there is different coffee bean yield growth trend among coffee growing areas over year.

Selection of the fixed effects structure

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 possible terms in the fixed effect structure 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 2, 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.

Selection of the random effects structure

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

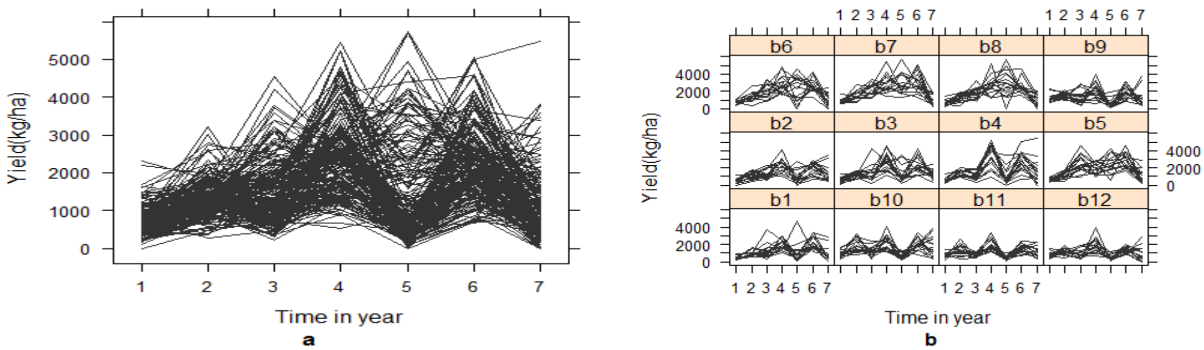

Figure 1: Individual profile plots of yield by coffee tree (a) and coffee tree nested in block (b)

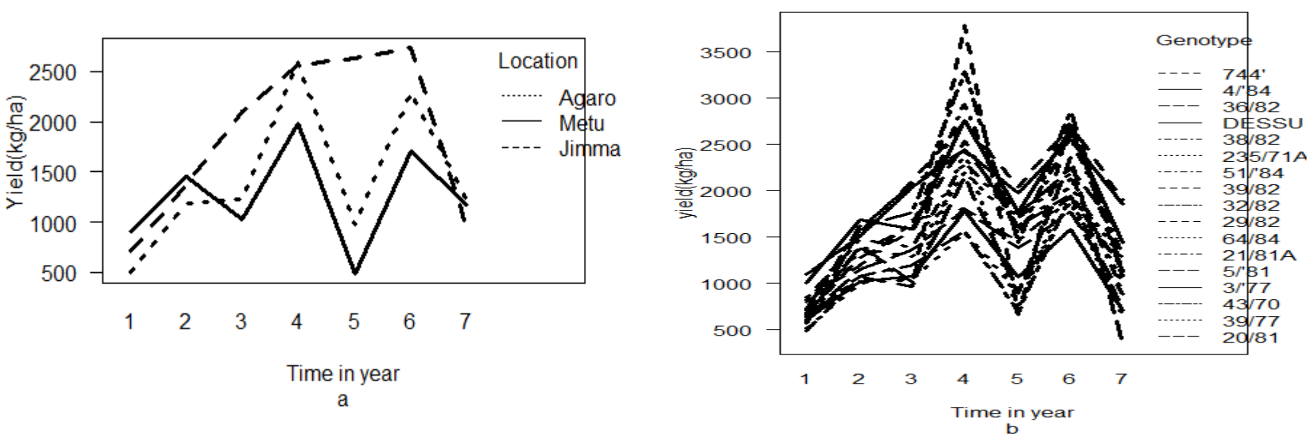

Figure 2: The mean profile plot of coffee yield by location (a) and genotype (b)

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.

The choice was made with AIC and BIC values for which smaller value is considered as better. Table 3 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.

Selection of the variance and correlation functions

After the selection of fixed and random effects, 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 4). 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 4). 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

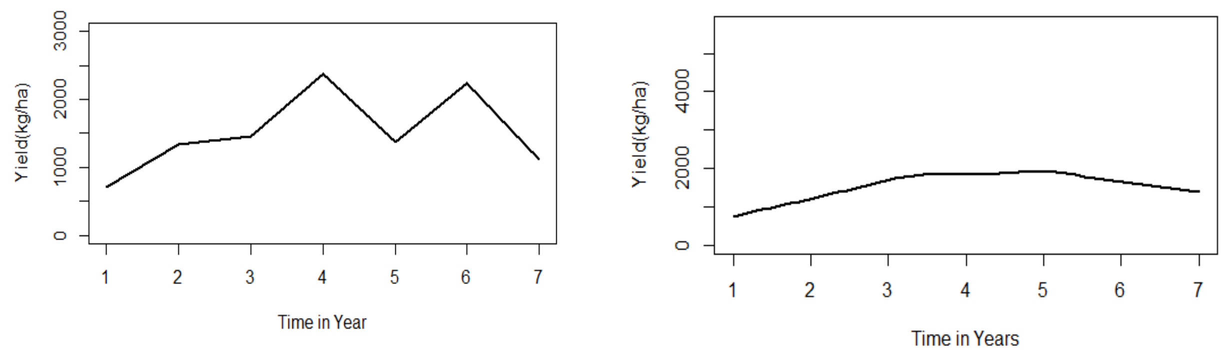

Figure 3: General mean profile plot of yield with and without loess smoothing

Table 2: 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

value since it applies a greater penalty for models with more parameters than does the AIC. Therefore, the heterogeneous variance function (varIdent) 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 (4) 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

Table 3: Selection of random effects to be included in the linear mixed model

No	Random Effects	AIC	BIC	Log Lik	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; log Lik= log likelihood; L.Ratio= likelihood ratio

Table 4: 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;

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).

Results of the final fitted linear mixed model

The output of the final fitted linear mixed model is summarized in two tables (Table 5 and Table 6). 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 5 and Table 6) , and the parameter estimate of random effects with 95% CI (Table 6).

Accordingly, this study showed evidence for the presence of serial correlation among repeated measurements of

Table 5: 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

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*) (Table.5). Despite coffee species and 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). 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 (Table 6), 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 study also revealed a significant location by linear and quadratic time effect interaction. From Table 6, 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

Table 6: Parameter estimates and their corresponding 95% CI for both random effects and the remaining fixed effects (which are not presented in table 5) 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		
σ (b_{0j})	221.81	(129.03	381.28)	
σ (b_{3j})	145.24	(68.48	308.05)	
corr($b_{0i,}$ b_{3j})	-0.78	(-0.96	-0.13)	
σ (ϵ_{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				

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.

CONCLUSIONS

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

LIST OF ACRONYMS

AIC	Akaike Information Criterion
AR1	Autoregressive Order One
BIC	Bayes Information Criterion
BLUP	Best Linear Unbiased Prediction
CBD	Coffee Berry Disease
CBY	Coffee Bean Yield
EDA	Exploratory Data Analysis
EIAR	Ethiopian Institute of Agricultural Research
GDP	Gross Domestic Product
JARC	Jimma Agricultural Research Center
LMM	Linear Mixed Model
LRT	Likelihood Ratio Test
MLE	Maximum Likelihood Estimator
RCBD	Randomized Complete Block design
REML	Restricted Maximum Likelihood Estimation

REFERENCES

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.

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.

Chauhan, R., Hooda, M. and Tanga, A., 2015. Coffee: The Backbone of Ethiopian

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

Davidian, M. and Giltinan, D., 1995. Nonlinear models for repeated measurement data CRC press, Lendon. 361p

Davis, L., 1957. Flowering and alternate bearing. *Proc.A mer. Soc.Hart. Sci.* 70:545-556. Economy. *International Journal of Economic Plants., 1*(2), 082-086

Gichuru, E., Agwanda, C., Combes, M., Mutitu, E., Ngugi, E., Bertrand, B. and Lashermes, P.,

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 (Colletotrichumkahawae) in Coffeaarabica. *Plant pathology*, *57*(6), pp.1117-1124.Piepho, H. and Eckl, T., 2014.Analysis of series of variety trials with perennial crops.*Grass and Forage Science*, *69*(3), pp.431-440.

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 (Colletotrichumkahawae) in Coffeaarabica. *Plant pathology*, *57*(6), pp.1117-1124.

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

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

Laidig, F., Piepho, H. P., Drobek, T., & Meyer, U. (2014).Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. *Theoretical and Applied Genetics*, *127*(12), 2599-2617.

Mesfin, K. and B. Bayetta, 2005. Genetic divergence of Hararge Coffee (Coffeaarabica L.) germplasm accessions at pre-bearing stage. Proceedings of the 20 International conferences on Coffee Science, Oct.11-15, Bangalore, India. pp: 1107-1112.

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

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.

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.

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.

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.