USMANU DANFODIYO UNIVERSITY, SOKOTO

(POSTGRADUATE SCHOOL)

STUDY OF MULTICOLLINEARITY ON THE EFFECT OF CLIMATIC CONDITIONS ON OIL PALM YIELD AT NIFOR NIGERIA

A Dissertation

Submitted to the

Postgraduate School

USMANU DANFODIYO UNIVERSITY, SOKOTO, NIGERIA

In Partial Fulfilment of the Requirements

For the Award of the Degree of

MASTER OF SCIENCE (STATISTICS)

BY

ESIOVWA, ABEL

(112103061236)

DEPARTMENT OF MATHEMATICS

JUNE, 2015

DEDICATION

This research study is dedicated to my parents (Mr Owin Esiovwa and Mrs Elizabeth Esiovwa), my beloved princess (Miss Stella Owie) and all who stood by me during my trialling period in the cause of undergoing this study for their unflinching support, encouragement, care and love they showed to me during the course of my study.

CERTIFICATION

This dissertation by Esiovwa, Abel (112103061236) has met the requirements for the award of the Degree of Master of Science (Statistics) of the Usmanu Danfodiyo University, Sokoto, and is approved for its contribution to knowledge.

Prof. R. A. Ipinyomi External Examiner	Date
Dr. A. Danbaba Major Supervisor	Date
Dr. R. V. K. Singh Co – Supervisor I	Date
Dr. A. D. Zagga Co –Supervisor II	Date
Prof. I. J. Uwanta Head of Department	Date

ACKNOWLEDGEMENTS

I thank the Almighty God for delivering me from the unfortunate incidence that almost claimed my life a day to the commencement of my second semester exams and for sustaining me in good health since then. I also thank God for granting me success in my studies and for His unending love, divine protection, and favour throughout my stay in Usmanu Danfodio University.

I wish to express my gratitude to my major supervisor and the Head of Statistics unit Dr. A. Danbaba who was always ready at all time and was never weary or tired throughout his supervision of this dissertation, despite his busy schedules made out time for the guidelines, comments, corrections and suggestions which were helpful for the achievement of this dissertation.

Special thanks also goes to my co-supervisor I Dr. R.. V. K. Singh and my co supervisor II D.r A. D. Zagga for their untied effort and useful contributions towards the success of this research work.

I also appreciate the efforts of all my lecturers and particularly the Head of Department in person of Prof. I. J. Uwanta, and the former Head of Statistics unit prof. S. U. Gulumbe for the opportunity to learn from their wealth of knowledge and making this programme a fruitful one for me and my other colleagues.

My thanks also go to Dr. Aliyu Usman from the department of mathematics and computer science of federal polytechnic Kaduna, ex-masters students' Isa Kargi and Isiyaku Amadu for the role they played towards the success of this research work.

This is to appreciate Mr. A. A. Edokpayi H.O.D. statistics division, Mr. Ngbede Godwin, data processing ass. (MET) NIFOR, Mr A. Oribhabor, H.O.D. harvesting division NIFOR and other staff of the harvesting division especially Mrs. E. Uwadiae, Mrs E Okodhi, Miss U. O. Obiozor, Mrs C. O. Izomon for assisting me to collect the relevant data for this study.

I also express my profound gratitude to my parents, closed friend Osakpamwan Osaigbovo and to my princess Owie Stella, I say thank you all for your support both morally, materially and spiritually.

I won't forget to acknowledge all my lovely course mates especially malam Abubakar Adamu, Tasiu Musa and Charity Sunday for their unwavering support and for standing by me when I needed them most.

My sincere thanks also goes to all UCC members and exco of 2012/2013 , Ecwa church members especially bro Shuaibu Mani, pastor Dauda Mani , Dr. S. S. matazu of department of science education, Dr. Kiro of French department for the love they showed to me during my challenging time.

With humility and a sense of obligation, I express my gratitude to all those who have directly or indirectly contributed to the success of this research dissertation. Those not mentioned here, I love you all.

TABLE OF CONTENTS

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS/NOTATIONS

FFB - Fresh fruit bunch of oil palm yield

OLS - Ordinary Least Squares

R.H. – Relative Humidity

S.R. – Solar Radiation

R.F. – Rainfall

S.S. – Sunshine

A.T. – Air Temperature

VIF – Variance Inflation Factor

EV – Eigenvector

PC – Principal Component

PCR – Principal Component Regression

STEP – Stepwise Regression

RR – Ridge Regression

PLSR – Partial Least Square Regression

NA – Not Available

SE - Standard Error

MSE – Mean Square Error

NIFOR – Nigeria Institute for Oil palm Research

CI – Condition Index

I(S.S-A.T) and I(S-T) – I(SUNSHINE - TEMPERATURE)

Varia – Variables

Interc – Intercept

ABSTRACT

One of the major problems of multiple linear regression analysis is multicollinearity of the independent variables. The existence of multicollineariity on climate variables such as relative humidity, solar radiation, rainfall, sunshine and temperature on the response of agricultural output may lead to inflation of standard error of the regression coefficients or false non-significant p-value. In this study, monthly data spanning from 1980-2012 obtained from the Nigeria Institute for Oil Palm Research (NIFOR) on relative humidity, solar radiation, rainfall, sunshine, temperature and oil palm yield were used to examine the probable effects of climate conditions/climate change would have on oil palm yield. The estimation of parameters of climatic variables in multiple linear regression appears to have suffered severe distortions due to multicollinearity. This research study resort to principal component regression, ridge regression and stepwise regression to stabilized the parameter estimate. Ridge regression was used to estimate the effect of climate conditions on oil palm yield because it performed better than others due to its lower measure of accuracy. It was observed that average relative humidity and rainfall had positive significant effect while solar radiation, mean sunshine hour and average air temperature had negative significant effect on oil palm yield.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background to the Study

Statistical models utilize the information from independent variables to predict, understand relation or control a dependent variable. Regression analysis is one of the most widely used of all statistical methods for model building. Multiple regression models are models containing a number of predictor variables (Neter *et al,* 2005). The multiple linear regression models are used to study the relationship between a dependent variable and more than one independent variable (Greene, 2003). For instance, agriculture is an economic activity that is highly dependent upon weather or other climate variables in order to produce the food and fibre necessary to sustain human life. Not surprisingly, agriculture is deemed to be an economic activity that is expected to be vulnerable to climate variability and changes. One of the biggest long-term risks to global development is climate change. Choices and investment made in climate change mitigation and adaption are vital for ensuring sustainable and inclusive growth. Anon. (2014a). Any unfavorable climate will negatively affect agricultural growth (Murad *et al*, 2010). Therefore, climate change and climatic conditions phenomenon are important issues that should be taken into account in maintaining the sustainability and productivity of agricultural crops. There are various measures for crop cultivation which could be employed to adapt to the current climate change event in order to minimize crop damage in the event of unexpected bad weather (Adger *et al*, 2007). In order to identify how climate change and climate conditions could negatively impact the Nigeria, Malaysian and other nation's socio-economy, it becomes necessary to understand the nature of climate variability. The description of the changing pattern of the climate could be understood by analyzing the pattern of daily temperature and rainfall for certain period of years (Al-Amin *et al*, 2011). Plantation agriculture specifically oil palm is one aspect of agriculture that has been greatly influenced by climate change and climate conditions. It is difficult to talk about oil palm without referring to Malaysia, since they are the largest world producer of it. Oil palm is best suited to the humid climate in Malaysia where the rain occurs at night and the days are bright and sunny. For optimum yield, the minimum rainfall required is around 1,500 mm $year^{-1}$ with an absence of dry season and an evenly distributed sunshine exceeding 2000 h $year^{-1}$ (Basiron, 2007). A mean maximum temperature range between 29 to 33° C and minimum temperature between 22 to 24° C favor the highest oil palm bunch production (Corley and Tinker, 2003). Oil palm yields can be complicated by a number of factors (i) the interaction of climatic factors with each other (ii) the oil palm, being a perennial, yields for many years and the climatic influences may be complicated by inter-bunch competition. One of the problems in regression analysis and Correlations per se can be the biased due to linear dependencies among independent variables or factors, leading to multicollinearity and spurious results. (Oboh and Fakorede, 1999). Multicollinearity refers to a situation in which two or more explanatory variables in a multiple regression model are highly linearly related. We have perfect multicollinearity if, for example, the correlation between two independent variables is equal to 1 or -1. In practice, we rarely face perfect multicollinearity in a data set. More commonly, the issue of multicollinearity arises when there is an approximate linear relationship among two or more independent variables. In the presence of multicollinearity, the estimate of one variable's impact on the dependent variable Y while controlling for the others tends to be less precise than if predictors were uncorrelated with one another. The usual interpretation of a regression coefficient is that it provides an estimate of the effect of a one unit change in an independent variable, X_1, holding the other variables constant. If

X_1, is highly correlated with another independent variable, X_2, in the given data set, then we have a set of observations for which X_1 and X_2 have a particular linear stochastic relationship. Some problems of multincollinearity are as follows:

i. One of the features of multicollinearity is that the standard errors of the affected coefficients tend to be large. In that case, the test of the hypothesis that the coefficient is equal to zero may lead to a failure to reject a false null hypothesis of no effect of the explanatory variable.

ii. High variance of coefficients may reduce the precision of estimation.

iii. Multicollinearity can result in coefficients appearing to have the wrong sign.

iv. A principal danger of such data redundancy is that of over-fitting in regression analysis models. The best regression models are those in which the predictor variables each correlate highly with the dependent (outcome) variable but correlate at most only minimally with each other. Such a model is often called "low noise" and will be statistically robust (that is, it will predict reliably across numerous samples of variable sets drawn from the same statistical population).

v. The presence of multicollinearity could also be misleading with the significance test telling us that some relevant variables are not needed in the model. Multicollinarity causes a reduction of statistical power in the significance of a statistical test.

vi. The unique solution for the parameter estimator is very unstable. The parameter estimators would change drastically when little changes occur in the independent or dependent variable. This also leads or explains the high variances in the parameter estimators. The variance of the parameter estimators for the explanatory or independent variables that led to multicollinearity would be very high or large. The result of having high variances is that the width or size of the confidence intervals for the parameters will also be inflated or increased abnormally. Therefore the effect of multicollinearity is grievous if the primary or major interest of research study is in estimating the parameters and identifying the significant 7 variables in the process. Anon. (2013a).

There are various formal and alternative techniques that have been developed for detecting the presence of serious multicollinearity. One of the most commonly used is variance inflation factor (VIF) that measures how much the variances of the estimated regression coefficients are inflated compare to when the independent variables are not linearly related Neter, *et. al.* (1990). The problem of multicollinearity can be averted or solved using some method of estimation or some modifications of the method of least squares for estimating the regression coefficients.

Some remedial measures to multicolinearity according to (Harshada, 2012) include:

1. **Stepwise Regression:** One remedy to multicolinearity is to drop one or several predictor variables and re-specify the regression in order to lessen the multicollinearity. This is done through variable selection which is intended to select the “best” subset of predictors. Stepwise regression (Backward Elimination) is the simplest of all variable selection procedures and can be easily implemented without special software.
2. **Alternative Methods:** If none of the predictor variables can be dropped, alternative methods of estimation are ridge regression and principal component Regression.

Ridge Regression: Ridge regression provides an alternative estimation method that can be used where multicollinearity is suspected. It gives an alternative estimator (k) that has a smaller total mean square error value.

Principal Component Regression: Another remedy suggested and could be used for the multicollinearity problem is the Principal Component Regression (PCR). Every linear regression model can be restated in terms of a set of orthogonal explanatory variables. These new variables are obtained as linear combinations of the original explanatory variables. They are referred to as the principal components. The principal component regression approach combats multicollinearity by using less than the full set of principal components in the model. To obtain the principal components estimators, assume that the regressors are arranged in order of decreasing eigenvalues, $\lambda 1 \geq \lambda 2 \ldots\ldots \geq \lambda p > 0$. In principal components regression, the principal components corresponding to near zero eigenvalues are removed from the analysis and least squares applied to the remaining components.

1.2 Statement of Research Problem

Nigeria remained the largest producer of crude oil palm in the world over from 1950s and till mid-1960s. It had a market share of 43.0%, supplying 645,000 Metric Tons (MT) of palm oil, on annual basis, across the globe. The civil war which began in 1967 and lasted till 1970 changed all of that. The war destroyed almost all of the oil palm plantations and dispersed the small land holders of oil palm, who till date, accounts for 80.0% of the oil palm produced locally. The war though ended but left behind a legacy of crippled oil palm industry. Today, from being the largest producer of oil palm, Nigeria is now a net importer of palm oil. The domestic palm oil produced totaled 850,000 (MT) in 2012. The growth in oil palm has stagnated at 850,000 (MT) since 2009. The consumption of palm oil in Nigeria amounts to 1.0 million (MT) per annum. The official figures states that the shortage in oil palm industry is estimated to be around 150,000 (MT) annually. Nigeria today produces only 1.7% of the world's consumption of palm oil which is insufficient to meet its domestic consumption which stands at

2.7%. Thus, the question of net exports doesn't arise. Reasons for the decline in oil palm production include small farm holdings, transportation mode, unavailability of human labour, low capital and variability in climatic factors or conditions. However, several research studies have reviewed that oil palm is highly sensitive to variation in climatic factors most especially rainfall, temperature and sunshine hours. Several views have been expressed about the impacts of irregularity of climate on oil palm production, some claimed that rural and poor oil palm farmers are most affected; some said that farmers who depend on traditional livelihood system such as farming, fishing and pastoral practices are most affected while some other researchers claimed that subsistence oil palm farmers are the most affected. Anon. (2013b).

With the foregoing, can we say these claims are really true of climate change/climate conditions? This study therefore, endeavors to find answers to the following research questions.

What are the effects of climate change and climatic conditions on oil palm production in Nigeria?

What coping strategies could be adopted by oil palm farmers in sustaining oil palm yield loses?

What is the influence of multicollinearity on the estimate of the effect of climate conditions on oil palm production in Nigeria?

1.3 Scope and Limitation of the Research

This research work is limited to data obtained at the Nigeria Institute for Oil Palm Research (NIFOR) only. The research is based on studying and handling problems of multicollinearity in regression analysis as it relate to climatic variables using secondary data obtained from NIFOR.

Multiple regression technique was reviewed and use to establish a linear relationship between the variables under study while stepwise Regression was use to examine the impact of each variable to the model.

Features from principal component regression and multiple regressions are used together to solve the problem of multicollinearity. The Principal components regression was used to compute the linear combinations of the predictive variables or regressors;

Ridge regression was also another method treated. In this case, a small biased estimator is added to the diagonal elements of the matrix (X'X) to be inverted which is the modifications of the least square estimator.

1.4 Relevance and Significance of the Study

Many ecological studies include the collection and use of data to investigate the relationship between a response variable and a set of explanatory factors (predictor variables). If the predictor variables are related to one another, then a situation commonly referred to as multicollinearity results. Then results from many analytic procedures (such as linear regression) become less reliable. In this research study, the usefulness of stepwise regression, principal component regression and ridge regression to solving problem on multicolinearity was examined with specific interest on the effect of climatic conditions in oil palm production. Findings in this study could stimulate other researchers to look into alternative methods and how to apply them to solving problem of muticolinearity in other relevant area of studies. While farmers, society and government will be better informed on the effect of climate change and climate conditions on oil palm production and be able to seek appropriate measure to address the problem.

1.5 Justification

Past studies have used a variety of approaches to capture climate change effects on agriculture (Parry *et al*, 2009; Wang *et al*, 2009; Deressa and Hassan, 2010). These approaches range from simply equating average future impacts to yield losses observed in historical droughts to more quantitative crop simulation modelling, statistical time series and cross-sectional analyses. Particularly, correlations, stepwise regression, multiple regressions, Ricardian model and path coefficient analysis have been used extensively to determine the relationship between climatic variables and yield components in the oil palm, by other researchers (Oboh *et al*, 1999., Zahid *et al*, 2012., Okpamen *et al,* 2012.) without taken, a look at the problem of multicolinearity. It is with regard to this that the research is bent on studying the multicollinarity on the effect of climatic conditions on oil palm yield.

1.6 Aim and Objectives of the Study

The aim of this study is to examine or investigate the multicolinearity on the estimation of effects of climatic conditions on fresh fruit bunch of oil palm yield. The objectives are to;

i. Assess or estimate the impact of, rainfall, relative humidity, average air temperature, sunshine hour and solar radiation variability on oil palm production.

ii. Determine the degree of multicolinearity among the independent variables (temperature, rainfall, sunshine, relative humidity and solar radiation).

iii. Compare principal component regression (PCR), ridge regression (RR), and stepwise regression models to dealing with multicolinearity.

iv. Assess the degree of efficiency between the methods (PCR, and RR).

1.7 Definition of Terms used in the Study`

Some of the basic terms in this study are listed below

Relative Humidity /Average Relative Humidity - The American Heritage® Dictionary of the English Language (2000), defines relative humidity as the ratio of the amount of water vapour in the air at a specific temperature to the maximum amount that the air could hold at that temperature, expressed as a percentage., While average relative humidity is the averaging of monthly relative humidity ranging between two periods of time.

According to Anon. (2014b):

Rainfall - It refers to liquid water droplets that fall from the atmosphere, having diameters greater than drizzle (0.5 mm).

Solar Radiation - This is the energy emitted in the form of electromagnetic waves from the sun.

Air Temperature /Average Air Temperature – Air temperature is a measure of the warmth or coldness of an object or substance with reference to a standard value while average air temperature is the average daily maximum and minimum temperatures. (Jones *et al,* 1999)

Sunshine / Sunshine hours / Monthly Mean Sunshine hours - sunshine is the direct irradiance from the Sun measured on the ground of at least 120 watts per square meter W/m2. In simple terms, it is approximately the sunlight strong enough to cause shadows on the ground. Anon. (2014c)).

sunshine duration or sunshine hours is a climatologically indicator, measuring duration of sunshine in given period (usually, a day or a year) for a given location on Earth, typically expressed as an average of several years. It is a general indicator of cloudiness of a location. Anon. (2014d).

Monthly mean sunshine hours is the average number of hours of bright sunshine each month in a calendar year, calculated over the period of record. Anon. (2007).

Oil Palm / Fresh Fruit Bunchh (FFB) - Oil palm *(Elaeis guineensis)* is a tall palm tree native to tropical Africa, having nutlike fruits that yield commercially valuable oil. Anon. (2011).

Fresh Fruit Bunch (FFB) - is the initials stand for fresh fruit bunch, and refers to the bunch harvested from the oil palm. Anon. (2012).

CHAPTER TWO

2.0 REVIEW OF LITERATURE

2.1 Multicolinearity

The absence of multi-collinearity is essential to a multiple regression model. In regression when several predictors are highly correlated, this problem is called multicollinearity or collinearity. When things are related, we say they are linearly dependent on each other because you can nicely fit a straight regression line to pass through many data points of those variables. Collinearity simply means co-dependence. Why is co-dependence of predictors detrimental? Think about a couple in a jury. If two persons who are husband and wife are both members of a jury, the judge should dismiss either one of them, because their decisions may dependent on each other and thus bias the outcome. Collinearity is problematic when one's purpose is explanation rather than mere prediction (Vaughan & Berry, 2005). Collinearity makes it more difficult to achieve significance of the collinear parameters. Multicollnearity has the following consequences.

1. Variance (SEE) of the model and variances of coefficients are inflated. As a result, any inference is not reliable and the confidence interval becomes wide.

2. Estimates remain BLUE, so does R^2

3 $R^2_{yxi\ldots xk} < r^2_{yx1} + r^2_{yx2} + \ldots + r^2_{yxk}$

2.2 The Concept of Climate, Climate Change and Climate Conditions

According to Anon. (2013c). Climate in a narrow sense is usually defined as the average weather, or in a more scientifically way, as the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The classical period for averaging these variables is 30 years. The relevant quantities are most often surface variables such as temperature, precipitation, sunshine, humidity and wind. Climate in a wider sense is the state, including a statistical description, of the climate system. But climate change refers to a statistically significant variation in either the mean state of the climate or in its variability, persisting for an extended period (typically decades or longer). Climate change may be due to natural internal processes or external forcing or to persistent anthropogenic changes in the composition of the atmosphere or in land use.

According to Anthony (2013).Suggested that climate change is a complex phenomenon. Consequently it is difficult concept to define. A general definition of this topic can be stated as a short-term or long-term alteration of the statistical properties of a climate system. Such a change can be temporary or permanent. It can occur regionally or globally. However, in recent times the main focus is on human activity that is responsible for climate change.

Climate conditions typically refer to various aspects and patterns of weather in a given area, and the potential consequences and effects that such weather can create. The area in which such conditions may be considered can be relatively small, though accurate understanding of climate in any area typically considers worldwide conditions as well. These conditions are often used as the basis upon which weather predictions and disaster warnings are formed, and provide possible causal or corollary data for events that occur. Climate conditions can refer to the actual weather itself, as well as possible results of the weather such as droughts. One of the simplest explanations of climate conditions is that they are the weather conditions found in a given area. This can include basic aspects of weather such as wind, rain, and snow, or somewhat more complicated elements of weather systems such as tornadoes, hurricanes, droughts, and rising sea levels. Understanding, analyzing, and predicting climate changes and conditions often involves a variety of aspects of weather and how air moves and acts across the surface of the Earth. Atmospheric pressure, precipitation, and jet streams are all considered and utilized to better understand climate conditions and how they create other effects. Anon. (2003).

2.3 Climate Change / Climatic conditions in some Specific Places

In many parts of the country of the world, the most important impact of climate change during 21st century will be its effect on the supply of water. Recent droughts in the Southeast and in the West have underscored our dependence on the fluctuating natural supply of freshwater. Any changes in water supply will quickly ripple through the nation's farms as well. The National Academy of Sciences, a lead scientific body in the U.S., determined that the Earth's surface temperature has risen by about 1 degree Fahrenheit in the past century, with accelerated warming during the past two decades. There is new and stronger evidence that most of the warming over the last 50 years is attributable to human activities. Anon. (2013d).

In Nigeria, analysis of long- term meteorological data (temperature, rainfall, dust haze) show discernable evidence of climate change (NIMET, 2008 cited in WEP, 2011).

The effect of climate change in Nigeria is already contributing to extreme weather events: amount of rainfall, proliferation of pests, crop diseases and high temperature effects (NEST, 2004).

2.4 The Effect of Climate Change / Climate Conditions in Agriculture

According to Margaret (2008) "Climate change is one of the greatest challenges of our time. Climate change will affect, in profoundly adverse ways, some of the most fundamental determinants of health: food, air, water".

According to (IPCC, 2007; Deressa *et al.,* 2008; BNRCC, 2008). "There is a growing consensus in the scientific literature that in the coming decades the world will witness higher temperatures and changing precipitation levels. The effects of this will lead to low/poor agricultural products".

According to (SPORE, 2008; Apata *et al*, 2009). Evidence has shown that climate change is already affecting crop yields in many countries. This is particularly true in low-income countries, where climate is the primary determinant of agricultural productivity and adaptive capacities are low.

According to Idowu *et. al*, (2011). Climate change and global warming if left unchecked will cause adverse effects on livelihoods in Nigeria, such as crop production, livestock production, fisheries, forestry and post-harvest activities, because the rainfall regimes and patterns will be altered, floods which devastate farmlands would occur, increase in temperature and humidity which increases pest and disease would occur and other natural disasters like floods, ocean and storm surges, which not only damage Nigerians' livelihood but also cause harm to life and property, would occur.

2.5 The Impact of Climate Change / Climate Conditions in Oil Palm Yield

The impart of current global climate change in Agriculture is not restricted to atmospheric alteration in temperature and rainfall intensities rather has crept into soil nutrients status resulting in depletion and oil palm output decline (Henson and Chang, 2007).

Projections by MOSTE (now called MOSTI or the Ministry of Science, Technology and Innovation) in 2000 established that oil palm was relatively robust to climate change compared to rice, rubber, and cocoa in Malaysia. Oil palm yield was rather insensitive to increases to air temperature by up to 1.4 degrees Celsius. Oil palm was instead sensitive to drier weather. Oil palm yield declined by about 1% for every 1% reduction in the amount of rain. Higher amounts of rain led to increases in oil palm yield unless the increased rainfall led to flooding. Expectedly, increases in ambient air

CO_2 concentration led to increases in oil palm yield by about 10% for every 200 ppmv (parts per million by volume) increase in CO_2 concentration.

The report suggested or indicated that oil palm yield is, as expected, sensitive to CO_2 concentration and air temperature. The yield benefits of the increase CO_2 is counterbalanced by the increase in air temperature. Less expectedly, however, wind speed plays a big role in affecting oil palm yields. Stomata conductance increases with increasing wind speeds, which in turn increases photosynthesis; thus, increasing the growth rate and yield of oil palm. The role of wind speed on oil palm is under-estimated and under-appreciated. In Malaysia, coastal soils often give higher oil palm yields as compared to that on inland soils. This is often attributed to higher soil fertility for coastal soils than for inland soils. But another possible reason could be that coastal soils experience higher wind speeds than those inland soils. When air temperature increases, this would increase water evaporation, which would lead to more cloud formation and rain. Greater amount of clouds, however, would reduce the amount of solar radiation reaching the oil palm fields. So, would oil palm yields subsequently fall? The results however indicate that a change in sunshine hour has little impact on oil palm yields. (Christopher, 2012).

2.6 Vulnerability and Adaptation of Oil Palm to Climate Change and Climate Conditions

The Intergovernmental Panel on Climate Change (IPCC), in its Second Assessment Report, defines vulnerability as “the extent to which climate change may damage or harm a system.” It adds that vulnerability “depends not only on a system’s sensitivity, but also on its ability to adapt to new climatic conditions” (Watson et al. 1996).

Adaptation means any adjustment, whether passive, reactive or anticipatory that is proposed as a means for ameliorating the anticipated adverse consequences associated with climate change. Oyekale *et al*, (2009)

According to Idowu *et al,* (2011).Suggested that Nigeria, at present does not enjoy food security and therefore is more vulnerable to the effect of climate change. The climatic threats to food security are due to extreme weather events, e.g. drought, floods and erosion, Variability in the onset and cessation of rainfall and rainfall amounts, Proliferation of pests and diseases affecting agricultural production, high temperature which depress production of crop.

Subsequently, in order to assuage the impacts of climate change in Nigeria, the under listed adaptation strategies among others should be undertaken: Provision of foot-bridges across road tracks/roads and road passages for use in times of floods especially in the farming communities, Improved presence of local government personnel to promote Enlightenment/campaigns, Provision of government subsidized (at least 50%) of all Agricultural Inputs (Seeds, Fertilizers, Agro-chemicals, etc.) for all stakeholders in the farming communities and Support for stakeholders through empowerment, training, equipment provision, credit assistance and training workshop support provision.

2.7 Modelling the Effect of Climate Change and Climate Conditions on Oil Palm Yield

According to Okpamen *et al,* (2012).There are two major classes of models which are talked about and have common similarities. These are experimental designed model and regression models. The similarities in these models are that all the models assumed data representing output variables for observed yield. He suggested that due to errors that may arise from individual or single application and prediction, an attempt is made to expand the modelling into several non -linear or multiple application or prediction. This will help take care of the incidence of antagonism of one element with another. It will help solve problem of joint application which is usually the case in oil palm fields and plantations. The successful application of multiple regression analysis for data statistically processed and transformed field results would graduate into a possible research experimental model adequate for prediction in plantations with similar environmental condition specific to it (soil and climate factors).

According to Majid and Moh'd, (2007).The kernel methods are helpful for clustering complex and high dimensional and non-linearly separable data. Consequently for developing a system for oil-palm yield prediction, an algorithm, weighted kernel k-means incorporating spatial constraints, is presented which is a central part of the system. The proposed algorithm has the mechanism to handle spatial autocorrelation, noise and outliers in the spatial data.

According to Guiot *et al*, (1982).When the independent variables show mild collinearity, coefficients of a response function may be estimated using the classical method of least squares, because climatic variables are often highly inter-correlated.

According to Oboh and Fakorede (1999). In perennial species, consideration should be given to the differences in weather from year to year; otherwise large unexpected errors in yield estimates may become unavoidable. In the oil palm, correlation and multiple regression have been extensively used to study the effect of climatic factors on fresh fruit bunch (FFB) yield. Solar radiation has been implicated as a factor causing major fluctuations in the yield of oil palm (Sparnnaij *et al*, 1963). Ferwerda (1977) opined that the effect of solar radiation may be due to moisture stress associated with the high temperature resulting from solar radiation.

According to (Broekmans, 1957; Spamnaij *et al, 1963; Ong,* 1982). There has been a tendency in previous studies on climatic effects on oil palm to relate each climatic factor to yield in isolation without regard to possible overlapping or interaction effects with other climatic factors. In taking into consideration this fact, not only correlations were used but also stepwise multiple regression and path coefficient analysis to comprehensively explain climatic influences on oil palm yield. Correlations per se can be biased due to dependency among factors, which lead to multicollinearity and spurious results. However, stepwise multiple regression is effective in considering the influence of each climatic factor on yield without any interactive effect. 'Path coefficient analysis as developed by Wright (1923) further helps to eliminate any spurious effects detected by correlation and regression as it determines the direct effect of the factor through other climatic factors.

According to Draper and Smith (1981). When the climatic variables exhibit multicollinearity, OLS inflates the percentage of variation in annual radial growth accounted for by climate (R^2 climate). Therefore, using ordinary regression procedures under high levels of correlation among the climatic variables affects the four characteristics of the model that are of major interest to dendroecologists: magnitude, sign, and standard error of the coefficients as well as R^2 climate. This will generally leads to problem of multicollinearity.

Similarly, Rolf (2002),suggested that when regressors are near-collinearity, in multiple regression, so called regularised or shrinkage techniques can be most or highly preferable to ordinary least squares by trading the biased for variance. Also continuum regression, introduced by Stone and brooks in (1990), combines several more classical regularized regression techniques, such as, PCR, RR and PLSR.

According to Mathias and Duk (2006), as suggested in his study, observed that the result of prediction task, OLS is never worse than other techniques and in many cases better than PLSR or RR. This may be as a result of the fact that PCR, RR, and PLSR spread strongly in the prediction matrix, but are nevertheless significant.

According to Li (2010) suggested in his comparative study of principal component regression (PCR), ridge regression (PR) and partial least squares regression (PLSR) stated that; the solution and performance of PCR, and RR, and PLSR seems to be quite similar in practical data sets. PLSR is preferable to PCR with a smaller number of factors. But RR performs slightly better according to the root mean of the predicted residual sum of squares. While in studies on simulation shows that PLSR prediction ability is more precise and more stable. As a result, there is need for strong verification or confirmation for any claim as it relate with the superiority of any of the three biased regression techniques.

CHAPTER THREE

3.0 MATERIALS AND METHOD

3.1 The Method of Data Collection

Secondary means of data collection was employed in this research work. The data collected and used in this dissertation was obtained from record and files of statistics and harvesting division of Nigeria Institute for Oil Palm Research ((NIFOR). NIFOR was founded in 1939 by the colonial masters. It is located in KM 7, Benin City, Edo state, Nigeria. The data was made up of monthly data of five climatic variables (Humidity, Air Temperature, Solar Radiation, Rainfall, Sunshine) and monthly yield of fresh fruit bunch of oil palm yield spanning the period of 1980-2012 (33 years). The data was collected mainly through documentations i.e. information that is collected by consulting previous records documents. Government is the most important routine compilers and suppliers of this type of data. Its advantages include; it is less expensive, it is time saving and better result can be obtained since those collecting the data are most likely to be professionals. Its set back involves – records on relevant items might not be available and the authenticity of the data cannot be guarantee by the researcher. The data used in this study are presented in tables as appendix C. The statistical packages used for the analysis were R and SPSS statistical package.

3.2 Methods of Detecting Multicollinearity

The methods of detecting multicollinearity include the Kleins Test, Frisch Confluence Analysis Test, Eigenvalues & Condition Index Test, Bunch – Map Analysis Test, Farrar–Glauber Test and Variance Inflation Analysis Test for multicollinearity. Some of the robust multicollinearity tests are discussed below.

3.2.1 Kleins Multicollinearity Test

Kleins test is a robust test for multicollinearity. (Klein, 1955) suggested that multicollinearity is not necessarily a problem unless the pair wise coefficients of determination are high relative to the overall multiple coefficient of determination. That is, Klein argues that collinearity is harmful if the following condition holds:

$$r^2_{X_i X_j} \geq R^2_{Y.X_1,X_2,\ldots,X_K}$$

Where;

$R^2_{Y.X_1,X_2,\ldots,X_K}$ = multiple coefficient of determination.

$r^2_{X_i X_j}$ = coefficient of determination between any two explanatory variables

(X_i and X_j)

3.2.2 Frisch's Confluence Analysis Multicollinearity Test

Frisch's confluence analysis is a suitable test for all regression models. (Frisch, 1934) suggested that the seriousness of the effects of multicollinearity seems to depend on the degree of inter-correlations $(r_{X_i X_j})$ as well as the overall correlation coefficient i.e. $(\sqrt{R^2_{Y.X_1,X_2,\ldots,X_K}} = R)$. Thus, the standard errors, the partial correlation coefficients $(r_{X_i X_j}\text{'s})$ and overall correlation coefficient R may be used for testing multicollinearity. The implication is that:

Large standard error may imply multicollinearity.

High inter-correlations of the explanatory variables may also imply multicollinearity.

Smaller R relative to the $(r_{X_i X_j}\text{'s})$ may also imply multicollinearity.

3.2.3 Eigenvalues & Condition Index Multicollinearity Test

From the eigenvalues of the $(X'X)$ matrix, we can respectively derive what is known as the condition number K and condition index $CI = \sqrt{K}$ as follows:

$$K = \frac{\max imum \quad eigenvalue}{\min imum \quad eigenvalue}$$

And

$$CI = \sqrt{K} = \sqrt{\frac{\max imum \quad eigenvalue}{\min imum \quad eigenvalue}}$$

Then we have this rule of thumb. If K is between 100 and 1000, there is a moderate to strong multicollinearity and if it exceeds 1000, there is severe multicollinearity. Alternatively, if the $CI = \sqrt{K}$ is between 10 and 30, there is moderate to strong multicollinearity and if it exceeds 30, there is severe multicollinearity. Some statisticians believe that the condition index $(CI = \sqrt{K})$ is the best available multicollinearity diagnostics for all kinds of regression models (Gujarati, 2003).

3.2.4 Bunch-Map Analysis Multicollinearity Test

A revised version of Frisch's confluence analysis is Bunch-Map analysis (Frisch, 1934) However, the combination of the three criteria set up by Frisch may detect multicollinearity. In order to gain as much knowledge as possible as to the seriousness of multicollinearity, the Bunch-Map analysis is hereby suggested.

The procedure is to regress the dependent variable on each one of the explanatory variables separately. Thus we obtain all the basis of a priori and statistical criteria. We choose the elementary regression which appears to give the most plausible results on both a priori and statistical criteria. Then we gradually insert additional variables and we examine their effects on the individual coefficients on their standard errors and on the overall R^2. A new variable is classified as useful, superfluous or detrimental when any of the following condition happens:

If the new variable improves R^2 without rendering the individual coefficients unacceptable (wrong) on a priori considerations, the variable is considered useful and is included among the explanatory variables. If the new variable does not improve R^2 and does not affect, to any considerable extent, the values of the individual coefficients, it is considered as superfluous and is rejected (i.e. is not included among the explanatory variables).

If the new variable affects considerably the signs or the values of the coefficients, it is considered as detrimental. If the individual coefficients are affected in such a way as to become unacceptable on theoretical, a priori considerations, then we may say that this is a warning that multicollinearity is a serious problem. The new variable is important, but because of inter-correlations with the other explanatory variables its influence cannot be assessed statistically by OLS. This does not mean that we must reject the detrimental variable. If we did so, we would ignore information valuable to our attempts of approaching, as best as we can the true specification of the relationship. In order to avoid the complications of multicollinearity and take into account the influence of detrimental variable, we have to find a remedy. If we omit the detrimental variable completely in an attempt to avoid its detrimental influence on the other coefficients, we must bear in mind that in so doing we simply leave its influence to be absorbed by the other coefficients.

This method differs from the Frisch's Confluence Analysis in that the latter estimates all possible regressions between the variables which are present in a relationship. It takes each variable successively as the dependent variable and considering all possible regressions of each variable on all others which are gradually introduced in the analysis. Thus, it is obvious that the Bunch-Map Analysis Multicollinearity Test requires much more computations but suitable for all kinds of models.

3.2.5 Farrar-Glauber Multicollinearity Test

A comprehensive statistical test for multicollinearity has been recently developed by Farrar and Glauber. It is really a set of three tests i.e. three test statistics are used for testing for multicollinearity. The first test is a $\chi^2 - test$ for the detection of the existence and the severity of multicollinearity in a model including several explanatory variables. The second test is an $F - test$ for locating which variables are multicollinear. The third test is a $t - test$ for finding out which variables are responsible for the appearance of multicollinear variables.

Farrar and Glauber consider multicollinearity in a sample as a departure of the observed $X's$ from orthogonality. Their approach emerged from the general idea that if multicollinearity is perfect then the regression coefficients become indeterminate. And that the inter-correlations among the various explanatory variables can be measured by multiple correlation coefficients and partial correlation coefficients. The test may be outlined as follows:

Firstly: A $\chi^2 - test$ for the presence and severity of multicollinearity in a model with several explanatory variables. The hypothesis being tested at this stage is that the sample $X's$ are orthogonal $(r_{X_i X_j} = 1 \ and \ r_{X_i X_j} = 0)$. For this, it is convenient to standardize the variables for sample size and for standard deviation. Standardization is implemented through division of all observations of each X, expressed in deviation from its mean, by $\sqrt{n}$ times the standard deviation of X. That is, the Standardized value of the mth observation of the jth variable is as follows:

$$\psi = \frac{X_{jm} - \overline{X}_j}{(S_{Xj})\sqrt{n}}$$

This operation is equivalent to division of each element of the determinant of the sums of squares and sums of products of the $X's$ (expressed in deviation form) by the square roots of the sums of squared deviations of the variables appearing in the element. In a three-variable model, the determinant of $X's$ (in deviation form) is given by the determinant below:

$$\varphi = \begin{vmatrix} \sum X_1^2 & \sum X_1 X_2 & \sum X_1 X_3 \\ \sum X_1 X_2 & \sum X_2^2 & \sum X_2 X_3 \\ \sum X_1 X_3 & \sum X_2 X_3 & \sum X_3^2 \end{vmatrix}$$

To obtain the Standardized form of this determinant, we divide the first element by $\left(\sqrt{\sum X_1^2}\sqrt{\sum X_1^2}\right) = \left(\sqrt{\sum X_1^2}\right)^2$, the second element by $\left(\sqrt{\sum X_1^2}\sqrt{\sum X_2^2}\right)$ and so on. In general, the element $\sum X_i X_j$ is divided by $\left(\sqrt{\sum X_i^2}\sqrt{\sum X_j^2}\right)$ to give the corresponding element of the Standardized determinant. For the three-variable model, the Standardized determinant is:

$$\varphi' = \begin{vmatrix} \dfrac{\sum X_1^2}{\left(\sqrt{\sum X_1^2}\right)^2} & \dfrac{\sum X_1 X_2}{\sqrt{\sum X_1^2}\sqrt{\sum X_2^2}} & \dfrac{\sum X_1 X_3}{\sqrt{\sum X_1^2}\sqrt{\sum X_3^2}} \\ \dfrac{\sum X_1 X_2}{\sqrt{\sum X_1^2}\sqrt{\sum X_2^2}} & \dfrac{\sum X_2^2}{\left(\sqrt{\sum X_2^2}\right)^2} & \dfrac{\sum X_2 X_3}{\sqrt{\sum X_2^2}\sqrt{\sum X_3^2}} \\ \dfrac{\sum X_1 X_3}{\sqrt{\sum X_1^2}\sqrt{\sum X_3^2}} & \dfrac{\sum X_2 X_3}{\sqrt{\sum X_2^2}\sqrt{\sum X_3^2}} & \dfrac{\sum X_3^2}{\left(\sqrt{\sum X_3^2}\right)^2} \end{vmatrix}$$

The Standardized determinant of the denominators of the least squares estimates may be rewritten in a slightly different form. Bearing in mind that the main diagonal elements are equal to unity and the off-diagonal elements are the simple correlations coefficients among the explanatory variables. Thus, the Standardized determinants are called correlation determinants. In the three -variable model, the Standardized determinant will be as below:

$$S = \begin{vmatrix} 1 & r_{X_1X_2} & r_{X_1X_3} \\ r_{X_1X_2} & 1 & r_{X_2X_3} \\ r_{X_1X_3} & r_{X_2X_3} & 1 \end{vmatrix}$$

From these forms, we can easily examine the two extreme cases of orthogonality and perfect multicollinearity. In the case of perfect multicollinearity, the simple correlation coefficient $r_{X_1X_2}, r_{X_2X_3}$ and so forth are equal to unity. Hence, the value of the Standardized (correlation) determinant is equal to zero. For the Three-variable model we have:

$$S = \begin{vmatrix} 1 & r_{X_1X_2} & r_{X_1X_3} \\ r_{X_1X_2} & 1 & r_{X_2X_3} \\ r_{X_1X_3} & r_{X_2X_3} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0$$

In the case of orthogonality of the $X's$, the simple correlation coefficient for each pair of $X's$ is equal to zero. Hence, the value of the Standardized (correlation) determinant is equal to unity. In the Three-variable model, we have the determinant below:

$$S = \begin{vmatrix} 1 & r_{X_1X_2} & r_{X_1X_3} \\ r_{X_1X_2} & 1 & r_{X_2X_3} \\ r_{X_1X_3} & r_{X_2X_3} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1$$

It follows that if the value of the Standardized (correlation) determinant lies between zero and unity, there is some degree of multicollinearity in the model. The above analysis suggests that multicollinearity may be considered as a departure from orthogonality. That is, the closer the value of the determinant is to zero, the stronger the degree of multicollinearity and vice versa. Starting from this fact, Farrar and Glauber suggested the following $\chi^2 - test$ for detecting the strength of multicollinearity over the whole set of explanatory variables. The hypothesis is as follows:

$H_0 : S = 0$ (The $X's$ are orthogonal i.e. there is no multicollinearity).

H_1 : S $\neq 0$ (The $X's$ are not orthogonal i.e. there is multicollinearity)

The Farrar Glauber test statistic is given by the formula below:

$$*\chi^2 = -\left[n - 1 - \frac{1}{6}(2K + 5)\right]\log_e S \quad ,$$

Where;

$S =$ Value of the standardized correlation determinant,

$n =$ Sample size, $k =$ Number of independent variable s,

$*\chi^2 =$ The value of χ^2 computed from the sample.

If the null hypothesis is true, then $*\chi^2$ follows a χ^2 distribution with $\nu = \frac{1}{2}K(K-1)$ degrees of freedom. Where χ^2 is the value that defines the critical region of the χ^2 with $\nu = \frac{1}{2}K(K-1)$ degrees of freedom. If the observed $*\chi^2$ is greater than the theoretical value of χ^2, the null hypothesis of orthogonality (H_0) is rejected otherwise accepted. That is, that there is multicollinearity in the model. Obviously, the higher the observed $*\chi^2$, the more severe the multicollinearity.

Secondly: A $F - test$ for the location of multicollinearity. To locate the factors which are multicollinear, Farrar and Glauber suggest the computation of the multiple coefficients of determination among the explanatory variables ($R^2_{X_1.X_2,...,X_K}$, $R^2_{X_2.X_3,...,X_K}$, and in general $R^2_{X_i.X_1,X_2,...,X_K}$). Then the test of the statistical significance of these coefficients of determination with an $F - test$. Thus, for each coefficient of determination the observed $F*$ is computed using the statistic below:

$$F* = \frac{\left(R^2_{X_i.X_1,X_2,...,X_K}\right)/(k-1)}{\left(1 - R^2_{X_i.X_1,X_2,..,X_k}\right)/(n-k)}$$

Where; as defined above,

$n =$ Sample size, $k =$ Number of independent variable s,

Thus, the hypothesis is as follows:

$H_0 : R^2_{X_i.X_1,X_2,...,X_K} = 0$ (The variable X_i is not collinear)

$H_1 : R^2_{X_i.X_1,X_2,...,X_K} \neq 0$ (The variable X_i is collinear)

The observed value of F^* is compared with the theoretical F with $\nu_1 = (k-1)$ and $\nu_2 = (n-k)$ degrees of freedom (at the chosen level of significance). The theoretical F value is the value of F that defines the critical region of the test. If $F^* > F$, H_0 is rejected and accept that the variable X_i is multicollinear. Else if $F^* < F$, H_0 is accepted and that the variable X_i is not multicollinear.

Thirdly: A $t-test$ for the pattern of multicollinearity. This is a $t-test$ which aims at the detection of the variables which causes multicollinearity. To find which variables are responsible for the multicollinearity, we compute the partial correlation coefficients among the explanatory variables and test their statistical significance with the $t-statistic$. Recall that the partial correlation coefficient between any two variables, X_i *and* X_j, shows the degree of correlation between these two variables, all others being kept constant. For two-variable model, the partial correlation coefficient is the same as the simple correlation coefficient. For three-variable model, the partial correlation coefficients are obtained by the formulae below:

$$r_{X_1X_2.X_3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}, \quad r_{X_1X_3.X_2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}} \text{ and } r_{X_2X_3.X_1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}}$$

For models involving more than three explanatory variables, similar formulae can be developed. The hypothesis to be tested in this case is:

$H_0 : r_{X_i X_j . X_1, X_2, \ldots, X_K} = 0$ (X_i, X_j are not responsible for multicollinearity)

$H_1 : r_{X_i X_j . X_1, X_2, \ldots, X_K} \neq 0$ (X_i, X_j are responsible for multicollinearity)

Having estimated the partial correlation coefficients, their significance is tested by computing for each of their $t - statistic$ below:

$$t^* = \frac{(r_{X_i X_j . X_1, X_2, \ldots, X_K})\sqrt{n-k}}{\sqrt{1 - r^2_{X_i X_j . X_1 X_2, \ldots, X_k}}}$$

Where;

$r_{X_i X_j . X_1, X_2, \ldots, X_K} =$ The partial correlation coefficients between X_i *and* X_j.

The observed value t^* is compared with the theoretical $t - value$ with $\nu = n - k$ degrees of freedom (at the chosen level of significance). If $t^* > t$ we accept that the partial correlation coefficient between the variables X_i *and* X_j is significant. That is, the variables X_i *and* X_j are responsible for the multicollinearity in the model. Else if $t^* < t$, we accept H_0, that is X_i *and* X_j are not the cause of multicollinearity, since their partial correlation coefficient is not significant. With the above three statistic we can find the severity, the location and the pattern of multicollinearity.(Farrar *et al*, 1967).

3.2.6 Variance Inflation Factor (VIF)

Understanding multi-collinearity should go hand in hand with understanding variation inflation. Variation inflation is the consequence of multicollinearity. We may say multicollinearity is the symptom while variance inflation is the disease. In a regression model we expect a high variance explained (R-square). The higher the variance explained is, the better the model is. However, if collinearity exists, probably the variance, standard error, parameter estimates are all inflated. In other words, the high variance is not a result of good independent predictors, but a mis-specified model that carries mutually dependent and thus redundant predictors! VIF shows how multicollinearity has increased the instability of the coefficient estimates (Freund and Ramon, 2000). Put differently, it tells you how "inflated" the variance of the coefficient is, compared to what it would be if the variable were uncorrelated with any other variable in the model. Variance inflation factor (VIF) is the common way for detecting multicollinearity. It can be computed as follow:

$$\text{VIF} = \frac{1}{1-R_j^2}$$

Where R_j^2 is the coefficient of determination in the regression of explanatory variable Xj on the remaining explanatory variables of the model. Generally, when VIF > 10, we assume there exists highly multicollinearity (Sana and Eyup, 2008).

3.3 Alternative Methods for Detecting Multicollinearity

Other rules of thumb, which do not necessarily involve serious inferential methodology for detecting multicollinearity, are also available. These rules may be used to strengthen the inferential methods earlier discussed in this study. Available literatures have revealed the following methods in this case. (Gujarati, 2003).

3.3.1 High R^2 but Few Significant t-Ratio

A classical symptom of multicollinearity is when R^2 is high, say $(R^2 > 0.8)$, the $F-test$ in most cases will reject the null hypothesis that the partial slopes coefficients are simultaneously equal to zero. But the individual $t-test$ will show that none or very few of the partial slope coefficients are statistically different from zero (Gujarati, 2003).

3.3.2 High Pair-Wise Correlations among Independent variables

Another suggested rule of thumb is that if the pair-wise or zero-order correlation coefficient between two independent variables is high say $(r_{X_iX_j} > 0.8)$, the multicollinearity is a serious problem. The problem with this criterion is that, although high zero-order correlations may suggest collinearity, it is not necessary that they be high to have collinearity in any specific case. To put the matter somewhat technically, high zero-order correlations are a sufficient but not a necessary condition for the existence of multicollinearity. This is because multicollinearity can exist even though the zero-order or simple correlations are comparatively low say $(r_{X_iX_j} < 0.5)$. (Gujarati, 2003).

3.3.3 Examination of Partial Correlations

Because of the problem earlier mentioned in relying on zero-order correlations, Farrar and Glauber have suggested that one should look at the partial correlation coefficients. Thus in the regression of the variables $Y \quad on \quad X_2, X_3 \quad and \quad X_4$, when $R^2_{1.234}$ is very high but $r^2_{12.34}, r^2_{13.24}, and. r^2_{14.23}$ are comparatively low may suggest that the variables $X_2, X_3 \quad and \quad X_4$ are highly inter-correlated and that at least one of these variables is superfluous. Although a study of the partial correlations may be useful, there is no guarantee that they will provide an infallible guide to multicollinearity, since it may happen that both R^2 and all the partial correlations are sufficiently high in some cases. (Gujarati, 2003).

3.3.4 Auxiliary Regression

Recall that multicollinearity arises because one or more of the independent variables are exact or approximately linear combinations of the other independent variables. One way of finding out which X variable is related to other X variables is to Regress each Xi on the remaining X variables and compute the corresponding R^2, which we designate as ; R_i^2 . Each one of these regressions is called an auxiliary regression; auxiliary to the main regression of $Y \quad on \quad the \quad X's$. Klien's rule of thumb says that multicollinearity may be a troublesome problem only if the R^2 obtained from an auxiliary regression is greater than the overall R^2. Then, following the relationship between $F \quad and \quad R^2$, the variable F_i follows the F distribution with $\nu_1 = (k-1)$ and $\nu_2 = (n-k)$ degrees of freedom.

Where F_i is given below:

$$F_i = \frac{\left(R^2_{X_i.X_1,X_2,\ldots,X_K}\right)/(k-1)}{\left(1-R^2_{X_i.X_1,X_2,\ldots,X_k}\right)/(n-k)}$$

Where;

$n =$ Sample size,

$k =$ Number of independent variable,

$R^2_{X_i.X_1,X_2,\ldots,X_K} =$ Coefficient of determination in the regression of X_i on the remaining X variables.

If the computed F_i exceeds the critical F at the chosen level of significance, then the particular X_i is collinear with the other $X's$. If it does not exceed the critical F, we say that it is not collinear with other $X's$. In which case we may retain that variable in the model (Gujarati, 2003).

3.3.5 Comparison of F and T test

Checking the relationship between the F and T tests might provide some indication of the presence of multicollinearity. If the overall significance of the model is good by using F-test, but individually the coefficients are not significant by using t-test, then the model might suffer from multicollinearity. (Anon. 2006).

3.4 Review of Related and Relevant Statistical Methods to Multicollinearity

The least squares estimators of the regression coefficients are the best linear unbiased estimators. That is, of all possible estimators that are both linear functions of the data and unbiased for the parameters being estimated, the least squares estimators have the smallest variance. In the presence of collinearity, however, this minimum variance may be unacceptably large. Relaxing the least squares condition that estimators be unbiased opens for consideration a much larger set of possible estimators from which one with better properties in the presence of collinearity might be found. Biased regression refers to this class of regression methods in which unbiasedness is no longer required. Such methods have been suggested as a possible solution to the collinearity problem. The motivation for using biased regression methods rests in the potential for obtaining estimators that are closer, on average, to the parameter being estimated than are the least squares estimators. Ridge regression and principal component regression are two commonly used biased regression methods. The biased regression methods attack the collinearity problem by computationally suppressing the effects of the collinearity. Ridge regression does this by reducing the apparent magnitude of the correlations. Principal component regression attacks the problem by regressing Y on the important principal components and then parcelling out the effect of the principal component variables to the original variables. (John et al, 1998), Sundberg, (1993), Bjoksilon and Sundberg,(1999) sited in Isa(2013) proposed partial least squares regression, ridge regression and principal component regression as some of the methods that have been developed to overcome the deficiencies of multicollinearity.

3.5 Classical Multiple Linear Regression

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable *x* is associated with a value of the dependent variable *y*. The population regression line for *p* explanatory variables $x1, x2, \cdots, xp$ is defined to be $y = b_0 + b_1 x1 + b_2 x2 + \cdots + b_p xp$. This line describes how the mean response μ_y changes with the explanatory variables. The observed values for *y* vary about their mean μ_y and are assumed to have the same standard deviation σ. The fitted values $b_{0,} b_{1,} b_{2,} \ldots, b_p$ estimate the parameters $\beta_{0,} \beta_{1,} \beta_{2,\ldots,} \beta_p$ of the population regression line. Since the observed values for *y* vary about their means μ_y, the multiple regression model includes a term for this variation. In words, the model is expressed as DATA = FIT + RESIDUAL, where the "FIT" term represents the expression $\beta_0 + \beta_1 x1 + \beta_2 x2 + \cdots + \beta_p xp$. The "RESIDUAL" term represents the deviations of the observed values *y* from their means μ_y, which are normally distributed with mean 0 and variance σ. The notation for the model deviations is ε.

Formally, the model for multiple linear regression, given *n* observations, is $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$ for i = 1, 2,…, n.

In the least-squares model, the best-fitting line for the observed data is calculated by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared, then summed, there are no cancellations between positive and negative values. The least-squares estimates b_0, b_1, ... b_p are usually computed by statistical software. The values fit by the equation $b_0 + b_1 x_{i1} + \cdots + b_p x_{ip}$ are denoted $\hat{y}_{i,}$ and the residuals ε_i are equal to $y_i - \hat{y}_i$, the difference between the observed and fitted values. The sum of the residuals is equal to zero.

The variance σ^2 may be estimated by $s^2 = \frac{\Sigma e_i^2}{n-p-1}$, also known as the mean-squared error (MSE). The estimate of the standard error *s* is the square root of the MSE.

Anon. (2013e).

3.6 Relating the Model to oil Palm Yield and Climate Variables

Consider plantation research in which the data consists of palm oil yield (i.e., the response variable *y*) that spans monthly for *n* years (33 years) and *k* climatic variables *x*1, *x*2, …, *x5*. Assume that in the region of the *x*'s defined by the data, *y* is related approximately linearly to the climatic variables. Multiple linear regression attempts to model the relationship between two or more of the explanatory variables and a response variable y by fitting a linear equation to observed data. Every value of the independent variable *x* is associated with a value of the dependent variable *y*. Recall, The linear additive model for relating a dependent variable to *p* independent variables is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_5 x_{i5} + \varepsilon_i \qquad \text{for i = 1,2, ... 396.} \qquad (3.7.0)$$

where the response variable *y* is the oil palm yield , the independent variables *x*1, *x*2, …, *x5* are monthly mean temperature, average monthly rainfall, monthly total relative humility, monthly sunshine and monthly solar radiation respectively.

$\beta_{0,}\ \beta_{1,}\ \beta_{2,}\ \ldots\ \beta_{5,}$ are the regression coefficients to be estimated, *n* is the number of years (33 years) times 12 months, and ε_i is the i^{th} year model error, assumed uncorrelated from observation to observation, with mean zero and constant variance. Here y_i is a measure of fresh fruit bunch of oil palm yield at the i^{th} year, X_{ji} is the i^{th} year reading on the j^{th} climatic variable. In addition, for the purpose of testing hypotheses and calculating confidence intervals, it is assumed that ε_i is normally distributed, ε_i ~ N (0, δ^2). Using matrix notation, the model in Eq. 3.7.0 can be written as: Y = Xβ + ε (3.7.1)

Where

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & \mathrm{X}_{n2} & \cdots & X_{np} \end{bmatrix}_{n\times(p+1)} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

The subscript *i* denote the observational unit from which the observations on *Y (oil palm yield)* and the *p (5)* independent variables were taken. The second subscript designates the independent (climatic) variables. The sample size is denoted with *n (396), i* = 1, . . . , *396*, and *p* denotes the five climatic variables. The parameters β_j , *j* = 0, . . . , *p* to be estimated when the linear model includes the intercept β_0. We assume that $n >$ p+1, four matrices are needed to express the linear model in matrix notation: Y : the $n\times 1$ column vector of observations on the dependent variable Y_i, *X*: the $n\times(p+1)$ matrix consisting of a column of ones, which is labeled 1**,** followed by the *p* column vectors of the observations on the independent variables; *β*: the $(p+1)\times 1$vector of parameters to be estimated; and ε : the *n* × 1 vector of random errors.

The least squares estimator $b = (b_0\ b_1\ b_2\ \cdots\ \mathrm{b}_p)'$ of the regression coefficients of the climatic variables is (assuming X is of full column rank) $\hat{\beta} = b = (X'X)^{-1} X'Y$ and the variance covariance matrix of the estimated regression coefficients in vector b is

$$\mathrm{Var}(b) = \sigma^2 (X'X)^{-1}$$

(Draper and Smith 1981, Myers 1986). Each column of **X** represents measurements for a particular climatic variable.

3.7 Stepwise Regression

Collinearity happens to many inexperienced researchers. A common mistake is to put too many regressors into the model. Inevitably many of those independent variables will be too correlated. In addition, when there are too many variables in a regression model i.e. the number of parameters to be estimated is larger than the number of observations, this model is said to be lack of degree of freedom and thus over-fitting. One common approach to select a subset of variables from a complex model is stepwise regression. A stepwise regression is a procedure to examine the impact of each variable to the model step by step. The variable that cannot contribute much to the variance explained would be thrown out. There are several versions of stepwise regression such as forward selection (add in variables one at time), backward elimination (start with full model and remove insignificant variables and stepwise (combination of forward and backward). Many researchers employed these techniques to determine the order of predictors by its magnitude of influence on the outcome variable (Leigh, 1996).

3.8 Principal Component Regression (PCR)

Principal Components Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value.

In ordinary least squares, the regression coefficients are estimated using the formula

$$\widehat{B} = (X'X)^{-1} X'Y$$

Note that since the variables are standardized, $X'X$ = R, where R is the correlation matrix of independent variables.

To perform principal components (PC) regression, we transform the independent variables to their principal components. Mathematically, we write

$$X'X = PDP' = Z'Z$$

where D is a diagonal matrix of the eigenvalues of $X'X$, P is the eigenvector matrix of $X'X$ and Z is a data matrix (similar in structure to X) made up of the principal components. P is orthogonal so that $P'P$ = I.

The new variables Z are created as weighted averages of the original variables X. Since these new variables are principal components, their correlations with each other are all zero. If we begin with variables X1, X2, and X3, we will end up with Z1, Z2, and Z3.

Severe multicollinearity will be detected as very small eigenvalues. To rid the data of the multicollinearity, we omit the components (the z's) associated with small eigenvalues. Usually, only one or two relatively small eigenvalues will be obtained. For example, if only one small eigenvalue were detected on a problem with three independent variables, we would omit Z3 (the third principal component).

When we regress Y on Z1 and Z2, multicollinearity is no longer a problem. We can then transform our results back to the X scale to obtain estimates of B. These estimates will be biased, but we hope that the size of this bias is more than compensated for by the decrease in variance. That is, we hope that the mean squared error of these estimates is less than that for least squares.

Mathematically, the estimation formula becomes

$$\widehat{A} = \left(Z'Z\right)^{-1} Z'Y = D^{-1}Z'Y$$

because of the special nature of principal components. Notice that this is ordinary least squares regression applied to a different set of independent variables.

Similarly, two sets of regression coefficients, A and B, are related using the formulas

$$A = P'B$$

and

$$B=PA$$

Omitting a principal component may be accomplished by setting the corresponding element of **A** equal to zero. Hence, the principal components regression may be outlined as follows:

1. Complete a principal components analysis of the **X** matrix and save the principal components in **Z**.

2. Fit the regression of **Y** on **Z** obtaining least squares estimates of **A**.

3. Set the last element of **A** equal to zero.

4. Transform back to the original coefficients using **B** = **PA**. Anon. (2013f).

3.9 The Ordinary Ridge Regression (ORR).

When multicollinearity occurs, the variances are large and thus far from the true value. Ridge regression is an effective counter measure because it allows better interpretation of the regression coefficients by imposing some bias on the regression coefficients and shrinking their variances (Morris, 1982; Pagel & Lunneberg, 1985; Mooney & Duval, 1993).Consider the standard model for multiple linear regression in equation (3.7.1)

$$Y = X\beta + E \qquad (3.9.1)$$

Where Y is (n × 1) vector of the dependent variable values, X is (n×p) matrix contains the values of P predictor variables and this matrix is full Rank (matrix of rank p), β is a (p × 1) vector of unknown coefficients, and E is a (n × 1) vector of normally distributed random errors with zero mean and common variance $\sigma^2 I$. Note that, both X's and Y have been standardized. The OLS estimate of $\hat{\beta}$ of β is obtained by minimizing the residual sum of squares, and are given by:

$$\left(Y - X\hat{\beta}\right)' \left(Y - X\hat{\beta}\right),$$

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y$$

$$Var\left(\hat{\beta}\right) = \hat{\sigma}^2 \left(X'X\right)^{-1}.$$

Where σ^2 is the mean squares error. This estimator $\hat{\beta}$ is unbiased and has a minimum variance. However, if $X'X$ is ill-conditioned (singular), the OLS estimate tend to become too and it makes some of the coefficients to have wrong sign (Wethrill, 1986). In order to prevent these difficulties of OLS, Hoerl and Kennard (1970), suggested the ridge regression as an alternative procedure to the OLS method in regression analysis, especially, when multicollinearity exist. The ridge technique is based on adding a biasing constants K's to the diagonal of $X'X$ matrix before computing $\hat{\beta}'s$ by using method of Hoerl and Kennard (2000). Therefore, the ridge solution is given by:

$$\hat{\beta}(K) = (X'X + KI)^{-1} X'Y, \qquad K \geq 0 \qquad (3.9.3)$$

Where K is ridge parameter and I is identity matrix. Note that if K = 0, the ridge estimator become as the OLS. If all K's are the same, the resulting estimators are called the ordinary ridge estimators (John, 1998) sited in (El-Dereny *et al*, 2011).

3.10 Method of Estimating K

The question in ridge regression is how to determine the parameter k. In general, Ridge Trace, generalized Cross Validation (GCV) and Mallow Cp are widely used. The ridge regression parameter is also chosen automatically using the method proposed by Cule *et al* (2012) in the package called ridge. In this case, the package automatically choose 0.01 as the ridge parameter, so the result is little different from the output of lm.ridge. *k* is a positive quantity less than one (usually less than 0.3).

3.11 Performance Measure

A major problem to consider in solving multicolinearity in regression analysis is to select the best approach or method that will handle the problem for various types of data for different numbers of independent variables. For the purpose of this study, the best was one that effectively provided solution to the problem of multicolinearity better than the others. The methods were therefore accessed based on their ability and performance to solve multicollinearity. The efficiency of the methods was evaluated by means of the residual standard error, which other regression packages call the root mean square error or RMSE or the standard error of the regression and the R^2. By definition the residual standard error is given as

$\sqrt{\frac{RSS}{df}}$. Clearly, it's good when this is small.

Where

$$RSS = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n}\left(y_i - \left(a + bX_i\right)\right)^2$$

and df = model degrees of freedom.

R^2 close to 1 is achieved when the fit of the model data is perfect, that is, all residuals are zero. The equation of R^2 is defined as follows: $R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the result and observation for the analysis of the data that was carried out in this research. Pearson correlations and OLS among variables were calculated to check and establish a linear relationship among the variables. multicolinearity diagnostics was conducted to identify the multicollinearity problem among independent variables using Eigen Values and Several Condition Numbers and variance inflation factor. Stepwise Regression, Ridge Regression and Principal Component Regression were conducted to overcome the multicollinearity problem. Comparison on stepwise regression, ridge regression, and principal component regression was carried out to identify the best remedial method to multicollinearity. Finally discussion about the outcome of the analysis was done. 0.05 alpha level of significant was set throughout the analysis.

4.2 Data Analysis

The results of the analysis are presented as follows

Table 4.1: Correlation matrix

	FFB	R.H.	S.R.	R.F.	S.S.	A.T.
FFB	1.0000000	0.5030185	-0.8964355	0.5101200	-0.6264534	-0.9355382
R.H.	0.5030185	1.0000000	-0.5010845	0.6211155	-0.4746649	-0.5254503
S.R.	-0.8964355	-0.5010845	1.0000000	-0.5108292	0.6234412	0.9646603
R.F.	0.5101200	0.6211155	-0.5108292	1.0000000	-0.5043177	-0.5367012
S.S.	-0.6264534	-0.4746649	0.6234412	-0.5043177	1.0000000	0.6359592
A.T.	-0.9355382	-0.5254503	0.9646603	-0.5367012	0.6359592	1.0000000

Table 4.1 shows the Pearson correlations or correlation matrix among the variables. All the predictor variables had a strong except air temperature (A.T.) that has a very strong relationship on the dependent variable i.e. fresh fruit bunch of oil palm (FFB). It shows that the Pearson's correlation coefficients of relative humidity (R.H.) and rainfall (R.F.) had positive values which are indication of increasing relationships between them and the fresh fruit bunch of oil palm while solar radiation (S.R.), sunshine (S.S.) and air temperature (A.T.) had negative Pearson's correlation coefficients which indicates a decreasing relationships between them and the fresh fruit bunch of oil palm. It can also be seen that, there was high positive correlations in terms of air temperature against solar radiation. All the independent variables except the correlation of relative humidity (R.H.) on sunshine (S.S) had significant relationship with one another.

Table 4.2: Model Summary of Multiple Rregression

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.937	0.8776	0.876	51.94

The R square, R adjusted, R and std. error of the estimate shows how good is the model used to fit the data. The R square indicates a 87.76% change in FFB due to the impact of the predictor variables. This implies that the model is a good fit.

Table 4.3: Multiple Rregression Analysis

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.836e +03	2.098e+02	18.283	<2e-16 ***
HUMIDITY	1.246e-01	4.195e-01	0.297	0.7667
RADIATION	2.396e-01	1.705e-01	1.406	0.1607
RAINFALL	-6.511e-03	2.890e-02	-0.225	0.8219
SUNSHINE	-4.884e+00	2.151e+00	-2.270	0.0237 *
TEMPERATURE	-1.323e+02	9.254e+00	-14.301	<2e-16 ***

Multiple regression was used to estimate the relationship between the dependent and the independent variables. It was also used to determine the effect of climate change on oil palm yield in the presence of possible multicollinearity problem. The t-test results indicated that only the variable sunshine (S.S) and air temperature (A.T.) in the model were statistically significant while relative humidity (R.H.), solar radiation (S.S.) and rainfall (R.F.) were not significant. The linear relationship between the variables can be written as:

FFB = 3836 + 0.125(HUMIDITY) + 0.240 (RADIATION) -0.007(RAINFALL)

- 4.884(SUNSHINE) - 132.3(TEMPERATURE)

Table 4.4: Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
HUMIDITY	1	2174353	2174353	805.8893	< 2.2e-16 ***
RADIATION.	1	4764476	4764476	1765.8774	< 2.2e-16 ***
RAINFALL.	1	10314	10314	3.8229	0.0512712 .
SUNSHINE	1	40177	40177	14.8910	0.0001333 ***
TEMPERATURE	1	551772	551772	204.5057	< 2.2e-16 ***

The f-test results indicated that all variables in the model were statistically significant. This does not tally with the t- test results which say only two of the five variables are significant. Thus implying possible multicollinearity problem.

4.3 Multicollinearity Diagnostics

Eigen values / condition index and variance inflation factor were used to diagnose the degree of multicollinearity in this study.

Table 4.5: Eigen Values and Several Condition Index

Eigenvalues λ_j	Condition Index (CI) $\sqrt{k_j}$
$\lambda_1 = 7.843241e+07 = 78432410$	$\sqrt{k_1} = 1.000000$
$\lambda_2 = 1.132837e+07 = 11328370$	$\sqrt{k_2} = 2.631262$
$\lambda_3 = 3.241683e+06 = 3241683$	$\sqrt{k_3} = 4.918838$
$\lambda_4 = 2.270487e+04 = 22704.87$	$\sqrt{k_4} = 58.774409$
$\lambda_5 = 5.993117e+02 = 599.3117$	$\sqrt{k_5} = 361.760714$

Table 4.5 shows that $\sqrt{k_4} = \sqrt{\left(\lambda_1 / \lambda_4\right)}$ and $\sqrt{k_5} = \sqrt{\left(\lambda_1 / \lambda_5\right)}$ are very high and there is a possibility that multicollinearity occurs and the result of parameter estimation in table 4.3 is unstable. There is a wide range in the eigenvalues and two of the condition index is larger than 30 signifying severed multicollinearlity problems in at least two linear combination of the independent variables.

Table 4.6: Variance Inflation Factor

X_j	VIF_j
X_1 : HUMIDITY	VIF_1 = 1.801491
X_2 : RADIATION	VIF_2 = 14.471980
X_3 : RAINFALL	VIF_3 = 1.865858
X_4 : SUNSHINE	VIF_4 = 1.820838
X_5 : TEMPERATURE	VIF_5 = 15.314389

Table 4.6 shows that VIF_2 and VIF_5 are high and X_2 and X_5 are correlated which indicate a strong or high multicolinearity. There are two variance inflations. For example, we can interpret $\sqrt{15.314389} \approx 4$ and ($\sqrt{14.471980} \approx 4$) as telling us that the standard error for both air temperature and solar radiation are 4 times larger than it would have been without collinearity. Radiation and temperature are suspected of causing multicollnearity since their VIF is greater than 10.

4.4 Counteracting Multcollinearity

Result of stepwise regression, ridge regression and principal component regression were presented below to solve the problem of multicollinearity.

Table 4.7: Model Summary of Stepwise Regression (Backward Elimination)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.937[a]	.878	.876	51.94303	.878	558.997	5	390	.000
2	.937[b]	.878	.876	51.87994	.000	.051	1	390	.822
3	.937[c]	.878	.877	51.81694	.000	.049	1	391	.826
4	.936[d]	.877	.876	51.88236	.000	1.993	1	392	.159

a. Predictors: (Constant), TEMPERATURE, HUMIDITY, SUNSHINE, RAINFALL, RADIATION

b. Predictors: (Constant), TEMPERATURE, HUMIDITY, SUNSHINE, RADIATION

c. Predictors: (Constant), TEMPERATURE, SUNSHINE, RADIATION

d. Predictors: (Constant), TEMPERATURE, SUNSHINE

e. Dependent Variable: FFB

The R square indicates 87.8% for the first three model and 87.7% for the fourth model change in fresh fruit bunch of oil palm yield (FFB) as a result of the effect of the predictor variables. This implies that the model is a good fit for the data.

Table 4.8: Coefficients of Stepwise Regression (Backward Elimination)

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	3835.506	209.781		18.283	.000		
	HUMIDITY	.125	.419	.007	.297	.767	.555	1.801
	RADIATION	.240	.170	.095	1.406	.161	.069	14.472
	RAINFALL	-.007	.029	-.005	-.225	.822	.536	1.866
	SUNSHINE	-4.884	2.151	-.054	-2.270	.024	.549	1.821
	TEMPERAT	-132.333	9.254	-.992	-14.301	.000	.065	15.314
2	(Constant)	3831.715	208.851		18.347	.000		
	HUMIDITY	.083	.376	.005	.220	.826	.690	1.450
	RADIATION	.238	.170	.094	1.401	.162	.069	14.458
	SUNSHINE	-4.800	2.116	-.053	-2.269	.024	.567	1.765
	TEMPERAT	-132.126	9.197	-.990	-14.366	.000	.066	15.164
3	(Constant)	3846.558	197.447		19.481	.000		
	RADIATION	.240	.170	.095	1.412	.159	.069	14.437
	SUNSHINE	-4.900	2.063	-.054	-2.375	.018	.594	1.683
	TEMPERAT	-132.432	9.081	-.992	-14.584	.000	.067	14.820
4	(Constant)	3591.971	80.500		44.621	.000		
	SUNSHINE	-4.757	2.064	-.053	-2.305	.022	.596	1.679
	TEMPERAT	-120.360	3.061	-.902	-39.327	.000	.596	1.679

a. Dependent Variable: FFB

Model 1 reflects the result of multiple regression. Rainfall is excluded in model 2 because is not significant in model 1. Humidity is removed from model 3 because is not significant in model 2 and Radiation is not part of the final model because is not significant in model 3. The number of predictors is reduced in order to find the most parsimonious (simplest) possible model that still guarantees a good prediction performance and obtain a good fit of the data at a relatively low model complexity.

Stepwise regression also leads to same variables that are significant in multiple regression model but with minor difference in the value of the coefficients and standard error. The model is reduced to FFB = 3591.971 - 4.757(SUNSHINE) – 120.360 (TEMPERATURE). Although the problem of multicollinearity seems to have been solved at the level of the 4^{th} model because the VIF has been reduced to a reasonable size but the result still appear not good enough.

Table 4.9: Model Summary for Ridge Regression

Model	R	R Square	Std. Error of the Estimate
1	.0.937	0.8776	4.014

The R square indicates a 87.76% change in the FFB due to the predictor variables and shows that the model is a good fit for the data.

Table 4.10: Ridge Regression Analysis

	Estimate	Scaled estimate	Std. Error (scaled)	t value (scaled)	Pr(>|t|)
HUMIDITY	7.722e-01	1.283e+02	5.044e+01	2.545	0.0109 *
RADIATION	-7.672e-01	-8.893e+02	4.031e+01	22.060	< 2e-16 ***
RAINFALL	4.538e-02	1.114e+02	5.063e+01	2.200	0.0278 *
SUNSHINE	-8.971e+00	-2.923e+02	5.105e+01	5.725	1.03e-08 ***
TEMPERATURE	-5.723e+01	-1.257e+03	3.969e+01	31.673	< 2e-16 ***

As we observed in table 4.5 and 4.6, there were multicollinearity problems in the linear regression model. Therefore, one of the biased regression methods – ridge regression was conducted to overcome the collinearity problem. All the variables appear to be statistically significant. Humidity and rainfall have positive or increasing effect on FFB while radiation, sunshine and temperature have negative or decreasing effect on FFB. These means that a unit increase in Humidity will increase fresh fruit bunch of oil palm (FFB) by 7.722e-01, a unit increase in rainfall will increase FFB by 4.538e-02. A unit increase in radiation will decrease FFB by -7.672e-01, a unit increase in sunshine will decrease FFB by -8.971e+00 and a unit increase in temperature will decrease FFB by -5.723e+01.

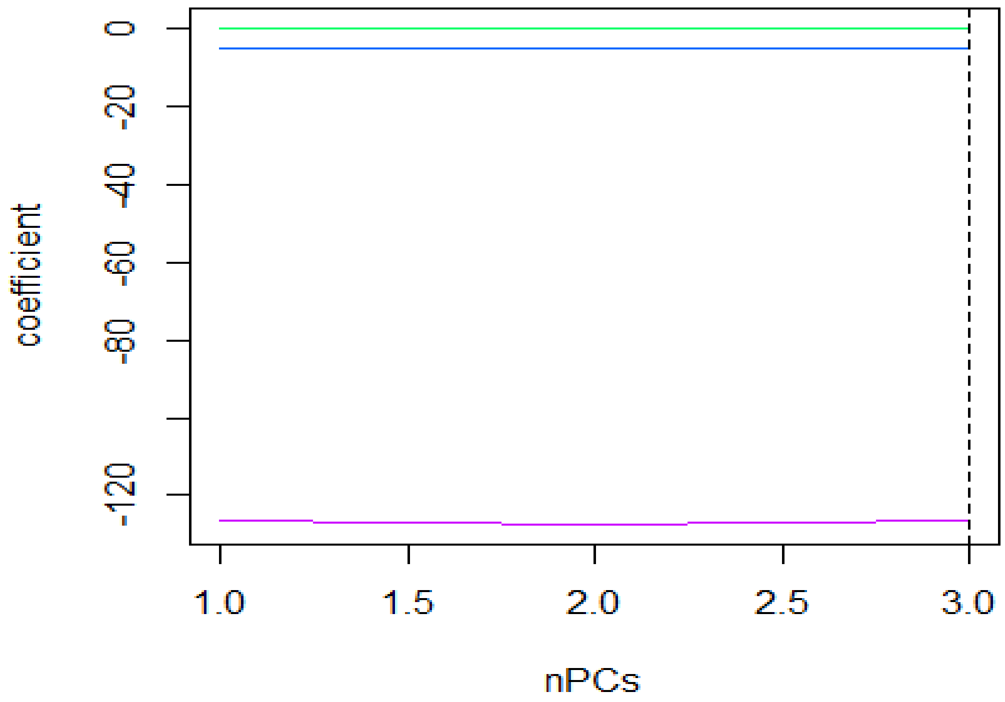

Fig. 4.1: Ridge Trace

Fig 4.1 shows how the automatic selection method of k= 0.002804394 was computed using 3 PCs first. This appears not good enough for lambda.

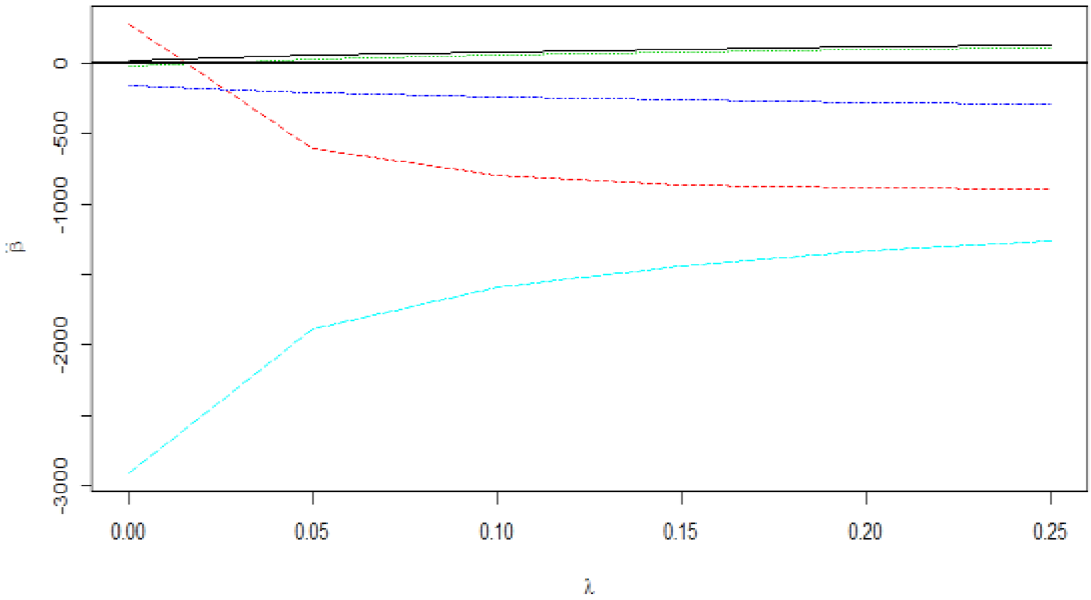

Fig. 4.2: Ridge Trace 2

Fig. 4.2 shows how the estimates of the individual regression coefficients were plotted against k or λ to give the ridge traces. A range of values (0.05 to 0.25) for various automatic selection for k or λ were use for the ridge trace. The value of 0.25 for k or λ was selected to stabilize the ridge traces.

Table 4.11: Model Summary for Principal Component Regression

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.936	0.8769	0.876	51.95

The R square indicates an 87.69% change in FFB due to the effect of the independent variables which implies that the model is a good fit for the data.

Table 4.12: Eigenvalues and eigenvectors in PC regression

	EV 1	EV 2	*EV 3*	EV 4	EV 5
HUMIDITY	0.405	-0.583	-0.227	0.667	0.015
RADIATION	-0.487	-0.431	0.302	0.037	-0.696
RAINFALL	0.413	-0.549	0.013	-0.727	0.018
SUNSHINE	-0.428	-0.135	-0.880	-0.157	-0.007
TEMPERATURE	-0.495	-0.394	0.288	0.039	0.718

The highest eigenvalue is located in the fifth eigenvector which is dominated by temperature. That is the share of temperature is relatively large.

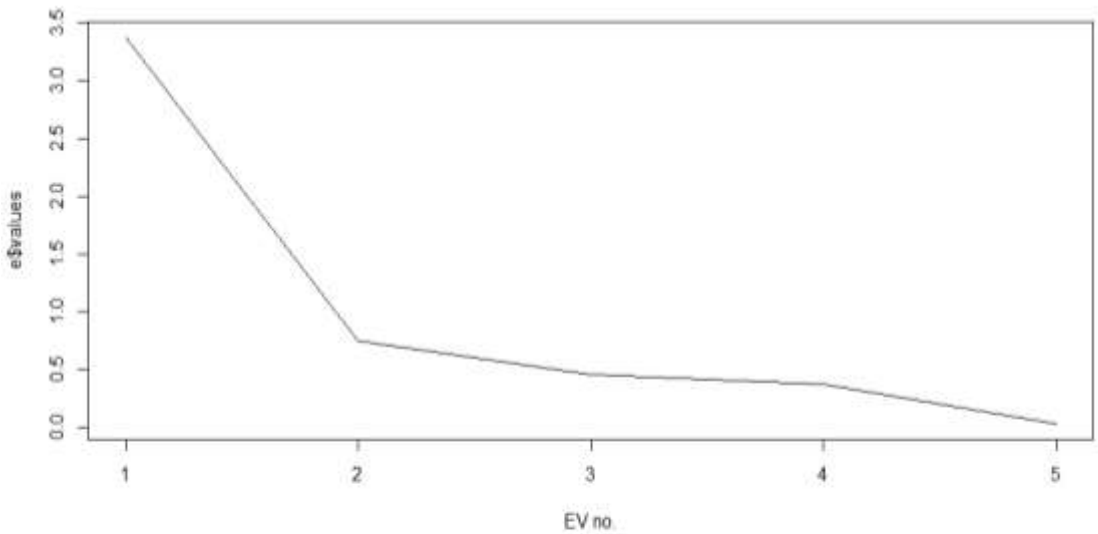

Fig. 4.3: Scree Plot

Fig 4.3 is the scree plot. One commonly used method of judging how many principal components are worth considering is the *scree plot.* Often, these plots have a noticeable “elbow” — the point at which further eigenvalues are negligible in size compared to the earlier ones. Here the elbow is at two, telling us that two principal components are enough.

Table 4.13: Transformed Z variable regression

	Estimate	Std. Error	t value	Pr(>|t|)	R Square	Standard Residual Error
(Intercept)	161.840	2.610	62.002	< 2e-16 ***	0.8776	51.94
EV 1	69.073	1.422	48.578	< 2e-16 ***		
EV 2	52.506	3.020	17.386	< 2e-16 ***		
EV 3	-31.178	3.846	-8.107	6.78e-15 ***		
EV 4	-2.636	4.262	-0.619	0.537		
EV 5	-114.659	14.030	-8.172	4.28e-15 ***		

One advantage of principal components is that it transforms the predictors to an orthogonal basis. Regressing the transformed Z variables against fresh fruit bunch (FFB) of oil palm, the above results were obtained. Notice that the p-values of all the eigenvectors are significant except the 4th one. The Standard Residual error and R Square are the same with that of the multiple regression at this point.

Table 4.14: $\left(X^T X\right)^{-1}$ Matrix

	(Intercept)	*EV 1*	*EV 2*	*EV 3*	*EV 4*	*EV 5*
(Intercept)	*0*	*0*	*0*	*0.00*	*0.00*	*0.00*
EV 1	*0*	*0*	*0*	*0.00*	*0.00*	*0.00*
EV 2	*0*	*0*	*0*	*0.00*	*0.00*	*0.00*
EV 3	*0*	*0*	*0*	*0.01*	*0.00*	*0.00*
EV 4	*0*	*0*	*0*	*0.00*	*0.01*	*0.00*
EV 5	*0*	*0*	*0*	*0.00*	*0.00*	*0.07*

Orthogonalized predictors for this data based on the eigen decomposition for the correlation matrix. Most of the variation is explained by the fifth eigenvector.

The first, second, third and fourth eigenvectos are roughly a linear combination of the 5^{th} eigenvector of the (standardized variable). The zero eigenvalues means perfect collinearity among independent variables.

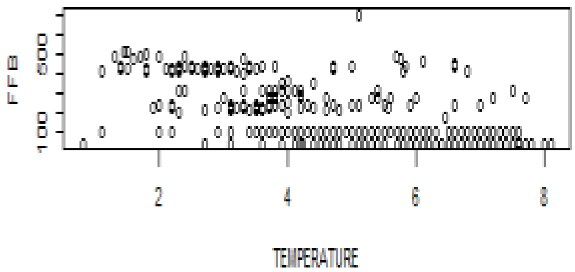

Figure 4.4: FFB & Temperature Plot

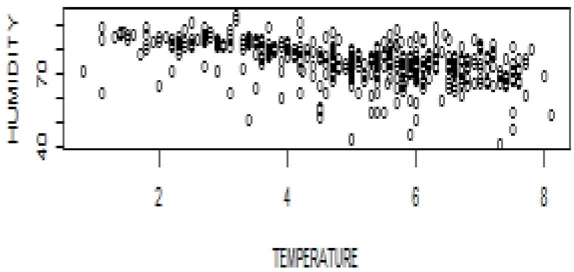

Figure 4.5: Humidity & Temperature Plot

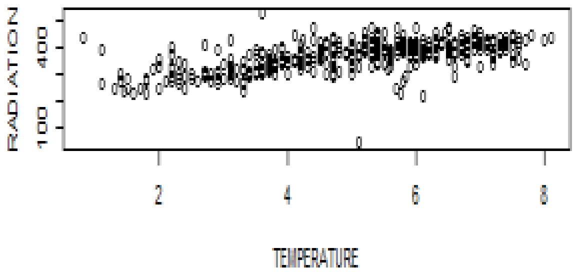

Figure 4.6: Radiation & Temperature Plot

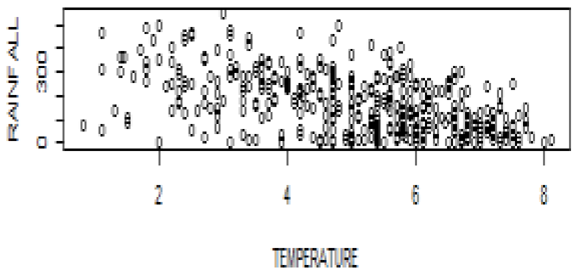

Figure 4.7: Rainfall & Temperature Plot

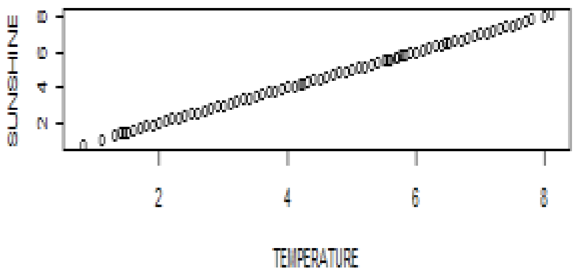

Figure 4.8: Sunshine & Temperature Plot

Fig 4.4 - 4.8 shows the plotting of each variable as they change over time. The plot of each of the variables as they change over time suggests that the fifth principal component should be identified with a temperature effect.

Table 4.15: Principal Component Regression Analysis

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.594e+03	8.859e+01	40.562	<2e-16 ***
TEMPERATURE	-1.252e+02	2.849e+00	-43.943	<2e-16 ***
RAINFALL	-1.078e-03	2.591e-02	-0.042	0.9668
I(SUNSHINE - TEMPERATURE)	-4.780e+00	2.134e+00	-2.240	0.0257 *

Table 4.15 shows the fitting of a regression with the third, fourth - the effect fifth component and fifth component. This approaches the fit of the full model of multiple regression and is easily interpretable. Only temperature and I(SUNSHINE - TEMPERATURE) is adjudged to be significant to the model.

4.5 Comparison on Stepwise Regression (STEP), Ridge Regression (RR), and Principal Component Regression (PCR)

Comparisons of the methods were made based on the Estimate of Parameters, Standard Errors of the coefficient, t-value, correlation coefficient, and Standard Residual error (SE).

Table 4.16 Comparison on Stepwise Regression (STEP), Ridge Regression (RR), and Principal Component Regression (PCR)

	Estimate of Prameters			Standard Errors		
Variables	STEP	RR	PCR.	STEP	RR	PCR
R.H.	NA	7.722e-01	NA	NA	5.044e+01	NA
S.R	NA	-7.672e-01	NA	NA	4.031e+01	NA
R.F	NA	4.538e-02	-1.078e-03	NA	5.063e+01	2.591e-02
S.S	- 4.757	-8.971e+00	NA	2.064	5.105e+01	NA
A.T.	-120.360	-5.723e+01	-1.252e+02	3.061	3.969e+01	2.849e+00
I(S.S-A.T)	NA	NA	-4.780e+00	NA	NA	2.134e+00
Intercept	3591.971	NA	3.594e+03	80.500	NA	8.859e+01

Table 4.16 (continues)

	t-value			Correlation Coefficient			Standard Residual error		
Varia	STEP	RR	PCR	STEP	RR	PCR	STEP	RR	PCR
R.H.	NA	2.545	NA	0.877	0.878	0.877	51.88	4.014	51.95
S.R.	NA	22.060	NA						
R.F.	NA	2.200	-0.042						
S.S	-2.305	5.725	NA						
A.T.	-39.327	31.673	-43.943						
I(S-T)	NA	NA	-2.240						
Interc	44.621	N.A.	40.562						

Table 4.16 shows the comparison on stepwise regression, ridge regression, and principal component regression. Although the correlation coefficient are approximately the same, the estimates to the parameter are different and the residual standard error (SE) of the parameters in RR are far more improved compared to PCR, and stepwise regression.

Stepwise regression performed better than PCR because it has a lesser standard error (SE). It seems that stepwise regression performed better than PCR with small difference in standard error (SE) because the method was not strong enough to detect that sunshine was a linear combination of Temperature. It is important to note that the difference among the estimated β_2 (Radiation) and β_5 (Temperature), affected the estimation of climate change on palm oil yield. It is reasonable that a slightly biased predictor with a very small prediction error variance can be better in terms of MSE than an unbiased prediction with a large prediction error variance.

4.6 Discussion of Result or Findings

From the study, the residual standard error value of the estimated regression coefficient was computed. The correlation matrix shows that the degree of multicollinearity was as high as 0.9. It was observed that RR has the lowest measure of accuracy among the methods that was considered. This shows that RR is more efficient in dealing with muticollinearity, stepwise regression was better than PCR in terms of measure of accuracy but PCR was better than it in terms of solving the problem of multiicolinearity as a result of the inability of stepwise regression to detect that sunshine was a linear combination of Temperature i.e. sunshine was redundant. It was noticed that the R-square was approximately the same for all the three methods. multicollinearity problems were detected in the data, by examining the correlation matrix, VIF and condition index. Since multicollinearity is present in the data, Multiple regression (OLS) estimators are imprecisely estimated as we can see from the result that shows that humidity, radiation and rainfall have no significant effect on fresh fruit bunch of oil palm yield. Multicollinearity leads to small characteristic roots and when characteristic roots are small, the standard error of $\hat{\beta}$ is large which implies an imprecision in the least squares estimation method. Stepwise regression was also looked at but it was not good enough to solving the problems of multicollinearity. Remedial measures such as ridge regression and principal component regression was use to solve the problem of multicollinearity, Ridge regression was apply to estimate the impact of climatic conditions/change on fresh fruit bunch of oil palm yield using five climatic variables because it appears to be the more efficient method of the two. It was discover that climate change has both positive and significant effect on fresh fruit bunch of oil palm yield in the sense that humidity and rainfall have positive or increasing effect on fresh fruit bunch of oil palm yield while radiation, sunshine and temperature have negative or decreasing effect on fresh fruit bunch of oil palm yield. The ridge regression model described the relationship between the independent variables and dependent variable as:

FFB = 7.722e-01 (HUMIDITY) -7.672e-01 (RADIATION) + 4.538e-02 (RAINFALL) -8.971e+00 (SUNSHINE) -5.723e+01 (TEMPERATURE)

CHAPTER FIVE

5.0 SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary

The aim of this study is to examine or investigate the multicolinearity on the estimation of the impact of climatic conditions on fresh fruit bunch of oil palm yield using NIFOR as a case study. The influence of relative humidity, solar radiation, rainfall, sunshine, and temperature were used to examine the probable effects climate change/conditions would have on oil palm yield. Ordinary least square (multiple regression) gave an imprecise and unrealistic result on the significant of the coefficient that was estimated because of multicollinearity that was present in the independent variables. Two variables (solar radiation and temperature) where considered to be highly correlated with each other and the VIF also show that there is high multicolinearity associated with both variables. Others had the existence of near-linear relationships among themselves. Stepwise regression was also considered as a way of selecting variables that are free from multicollinearity. Stepwise regression gave same significant variables as multiple regression but with an improved value of the coefficients and standard error. However since OLS created inaccurate estimates of the regression coefficients, by inflating the standard errors of the regression coefficients and stepwise regression could not adequately solve the problem of multicollinearity, as a result biased regression (ridge and principal component regression was introduced in order to check the multicollinearity. It was discovered that climate change has positive and negative significant effect on fresh fruit bunch of oil palm yield. Particularly, Humidity and rainfall have increasing effect on fresh fruit bunch of oil palm yield while radiation, sunshine and temperature have decreasing effect on fresh fruit bunch of oil palm yield.

5.2 Conclusion

The result of this study shows that increase in humidity and rainfall positively impact fresh fruit bunch of oil palm yield but increase in radiation, sunshine and temperature negatively impact fresh fruit bunch of oil palm yield. It was detected from the analysis that there is high multicolinearity in the estimation of the impact of climate Conditions / climate change on oil palm yield.

It was also observed from the results that ridge regression has a smaller residual standard error than that of principal component regression and stepwise regression, this implies that ridge regression has a lower measure of accuracy. This means that the ridge regression is a better method of dealing with multicollinearity than stepwise regression and principal component regression.

5.3 Recommendations

Based on the findings of this study, the following recommendations are hereby suggested for checking the problems of multicollinearity in the assessment of the impact of climate change on oil palm yield, minimizing climate change and climate conditions risk and hazard on oil palm yield;

1. In the presence of multicollinearity, a biased regression method in which unbiasedness is no longer required is suggested as a possible solution to the multicollinearity problems.
2. Ridge regression is recommended as the best method of solving multicollinearity problems when compare with stepwise regression and principal component regression.
3. Measures should be taken to check the negative impact of climate change and climatic conditions on oil palm yield.

4. Government must also ensure that the different research institution are researching into making new variables or breeds of palm oil that can-adapt to climate change of Nigeria for high productivity

5. Extension agents should also encourage farmers on the use of seed varieties that can adapt to different climate conditions.

5.4 Suggestion for Further Study

1. Other researchers could look at the economic impact of climate change and climatic conditions on oil palm yield with regards to

multicollinearity.

2. More than two biased method of regression should be use as remedy to multicollinearity problems after which the best selected method will be use to estimate the impact of climatic conditions on oil palm produce.

3. Other methods of detecting multicollinearity could be used by other researchers to estimate the degree of mulicollinearity on the impact of climatic conditions on oil palm yield.

4. This study should be extended to other cash crops like rubber, cocoa etc. by other researchers.

REFERENCES

Adger, W.N., Agrawala S.,. France O., Qader M.M. and Canada M. (2007). *Assessment of Adaptation practices, Options, Constraints and Capacity.* In: Change 2007: *Impacts, Adaptation and Vulnerability*, Parry, M.L., Canziani, O.F., Palutikof, J.P., van der, P.J.L. and Hanson, C.E. (Eds.). Cambridge University Press, Cambridge, UK., 717-743.

Al-Amin, W. L., Jaafar, A.H. and Ghani, A.Z. (2011). Assessing the impacts of climate in the Malaysia agricultural sector and its influences in investment decision. *Middle East J. Sci. Res.*, **7**: 225-234.

Anon. (2003). What are climate conditions? (with pictures) – Wise GEEK. At: http://www.wisegeek.com/what-are-climate-conditions.htm. Accessed on 16th October, 2014.

Anon. (2006). Multicollinearity; Causes, Effects and Remedies. At: www.iasri.res.in/seminar/AS-299/ebooks%... Accessed on16th October, 2014.

Anon. (2007). Climate Data Online- definition of other daily statistics. At: www.bom.gov.au/climate/cdo/about accessed on 8th September, 2014.

Anon. (2011). Oil Palm. At: http://www.thefreedictionary.com/oil+palm. Accessed on 16 September, 2014.

Anon. (2012). Basics to Palm Oil Terms/Definition Process. At: www.carotech.net/index/glassary.htm/ accessed on 17th September, 2014.

Anon. (2013a). Definition – Detection of multticollinearity - Consequences of multicollinearity. At: www.en.wikipedia.org/../multicollinearity. Accessed on 9th May 2013.

Anon. (2013b). Oil Palm Industry as Catalyst for economic development. At: http://www.dailyindependentnig.com/2013/12/oilpalm... Accessed on 16th October, 2013.

Anon. (2013c). What is the Climate System? - WMO. At: www.wmo.int.>programmes>wcp>ccl. Accessed on 9th May, 2013

Anon. (2013d). NASA- What's the Difference Between Weather and Climate. At: www.nasa.gov/../climate-weather.html. Accessed on 9th May, 2013.

Anon. (2013e). Multiple Linear Regression.At:www.stat.yale.edu/../linmult.htm. Accessed on 10th May, 2013.

Anon. (2013f). Principal Components Regression – NCSS.com. At: www.ncss.wpengine.netdna-cdn.com/wp-con... Obtained on 10th May, 2013.

Anon. (2014a). Climate Change/Independent Evaluation Group. At: https://ieg.worldbankgroup.org/topic/climatechange? Accessed on 16 October, 2014.

Anon. (2014b). Weather Terms – National Weather Service – NOAA. At: www.nws.noaa.gov/er/box/glossary.htm. Extracted on 15th September 2014.

Anon. (2014c). What is the definition of sunshine. At: www.answer.com/Q/what_is_the_definition_of_sunshine. Accessed on 17th September, 2014.

Anon. (2014d). Day time duration definition and measurement – Geographic distribution. At: www.en.wikipedia.org/../Sunshine_duration. Accessed on 8th September 2014.

Anthony, B. (2013). Climate change- meanings, implications, suggested solutions. Sri Lanka. At: www.jp.onkyo.com/../as02.htm. Accessed on 9th May, 2013.

Apata, T.G., Samuel, K.D., and Adeola, A.O. (2009). Analysis of Climate Change Perception and Adaptation among Arable Food Crop Farmers in South Western Nigeria. *Contributed Paper prepared for presentation at the International Association of Agricultural Economists' 2009 Conference, Beijing, China, August* **209**:16-22.

Basiron, Y. (2007). Palm oil production through sustainable planatations. *Eur. J. Lipid Sci. Techol.*, **109:** 289-295.

Bjoksilon, A. and Sundberg, R. (1999). A Generaized view on Continum regression. Scandinaviasn *journal statistics.* **25**:17-30.

Broekmans, A.F.M. (1957). Growth, flowering and yield of the oil palm in Nigeria. *J. W. Afr. Inst. Oil Palm Res.*, **2**: 187.200.

Building Nigeria's Response to Climate Change (BNRCC), (2008). 2008 Annual Workshop of Nigerian Environmental Study Team (NEST): The Recent Global and Local Action on Climate Change, held at Hotel Millennium, Abuja, Nigeria; 8-9th October, 2008. *In CBN Journal of Applied Statistic,* 2 **(1)**:32-50.

Christopher T. B. S. (2012). Effect of climate change on oil palm yield in Malaysia: Some simulations At: http://christopherteh.com/blog/2012/09/cc-oil-palm/. Accessed on 9th May, 2013.

Corley, R.H.V. and Tinker, P.B.H. (2003). *The Oil Palm*, Blackwells, Oxford, UK.

Cule, E. and De Iorio, M. (2012). A semi-automatic method to guide the choice of ridge parameter in ridge regression, stat. AP. At: www.arxiv.org/pdf/1205.0686. Accessed on 10th May, 2013.

Deressa, T.T. and Hassan, T. (2010). Economic Impact of Climate Change on Crop Production in Ethiopia: Evidence from Cross-section Measures. *Journal of African Economies*, 18 **(4)**: 529–554.

Deressa, T., Hassan, R., Alemu, T., Yesuf, M. and Ringler, C. (2008). Analyzing the determinants of farmers' choice of adaptation measures and perceptions of climate change in the Nile Basin of Ethiopia. *International Food Policy Research Institute (IFPRI)* Discussion Paper No. 00798. Washington, DC: IFPRI. *CBN Journal of Applied Statistic,* 2 **(1)**:31-50.

Draper, N.R. and Smith, H. (1981). *Applied regression analysis.* 2nd edition. New York: John Wiley and Sons.

El-Dereny, M. and Rashwan, N. I. (2011). Solving Multicollinearity Problem Using Ridge Regression Models. *Int. J. Contemp. Math. Sciences*, 6 **(12)**: 585 – 600.

Farrar, D.E. and Glauber, R.R. (1967). Multicollinearity in regression analysis: the problem revisited, *The Review of Economics and Statistics*, **49**: 92-107.

Ferwerda, J.D. (1977). Oil palm. In *Ecophysiology of Tropical Crops.* (Aluim P de T and Kozlowki T T, eds.1 Acad. Press. New York, USA. 351-382.

Freund, R.J. and Ramon, C.L. (2000). *SAS System for Regression* (Third edition). Cary, NC: SAS Institute.

Frisch, R. (1934). *Statistical Confluence Analysis by Means of Complete Regression Systems,* Oslo: University Institute of Economics.

Greene, W.H. (2003).*Econometric Methods*, fifth edition, NewYortk: Pren-tice Hall.

Guiot, J., Berger, A.L. and Munaut, A.V. (1982). Response functions. In: Hughes, M.K., Kelly, P.M., Pilcher, J.R. and LaMarche, V.C., eds. *Climate from tree rings.* Cambridge, UK: Cambridge University Press: 38-45.

Gujarati, D. (2003). *Basic Econometrics.* Third Edition, New York: John Wiley and Sons.

Harshada, J. (2012). *Multicollinearity Diagnostics in Statistical Modelling and Remedies to deal with it using SAS* . Cytel Statistical Software & Services. Session SP07 – Phuse.

Henson, I.E. and Chang, K.C. (2007). Modeling Oil Palm Nutrient Demand, Nutrient Turnover and Balance. 1-65. In oil palm plantation determined at Workshop on Nutrient Needs in Oil palm. Southeast Asia Program of PPI/PPIC and IPL. Singapore.23pp.

Hoerl, A.E. and Kennard, R.W. (2000). *Ridge Regression* : Biased Estimation for non orthogonal Problems, Technometrics*,* **42**: 80 – 86.

Hoerl, E. and Kennard, R.W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, *American Statistical Association and American Society for Quality.* 12, **(1)**: 55-67.

Idowu, A.A., Ayoola S.O., Opele, A.I. and. Ikenweiwe, N.B.(2011).Climate change in Nigeria. *Iranica Journal of Energy & Environment* 2 **(2)**: 145-152.

Intergovernmental Panel on Climate Change (IPCC), (2007): Climate Change 2007: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Parry, M.L., Canziani, O.F., Palutikof, J.P., van der L., Paul, J., and Hanson, C.E. (eds.)]. Cambridge University Press, Cambridge, United Kingdom, *CBN Journal of Applied Statistic, 2* ***(1)****:31-50.*

Isa, A.K. (2013).Comparative study between Principal Components Regression and Ridge Regression Methods in Dealing with Multicollinearity. An unpublished dissertation submitted to the postgraduate school, Usmanu Danfodiyo University.

John, O.R., Sastry G.P. and David A.D, (1998). *Applied regression analysis: a research tool*. — 2nd ed. Springer-Verlag New York, Inc.

John, T.M. (1998). Mean Squared Error Properties of the Ridge Regression Estimated Linear Probability Models**.** An unpublished Ph.D. Dissertation, University of Delaware.

Jones, P.D., New, M., Parker, D.E., Martin, S. and Rigor, I.G. (1999). Surface air temperature and its variations over the last 150 years. *Reviews of Geophysics* **37**:173-199.

Klein, L.R. (1955). "On the Interpretation of Theil's Method of Estimating Economic Relationships", *Metroeconomica*, **7**: 147-153.

Leigh , J. J. (1996). Admissibility of students to Ph.D. programs in counseling psychology. Unpublished dissertation. Boston College.

Li, Y. (2010), " A Comparison Study of principal components regression, partial Least squares regression and Ridge regression with application to FTIR Data " An unpublished Msc thesis in statistics Faculty of Social Sciences Uppsala University, Sweden.

Margaret, C. (2008). Director General's statement on World Health Day 2008. At: http://www.who.int/mediacentre/news/statements/2008/s05/en/index.html. Accessed on 9th May 2013.

Majid, A.A. and Moh'd., N. M. S. (2007). A Framework for Predicting Oil-PalmYield from Climate Data. *International Journal of Information and Mathematical Sciences* 3:2.

Mathias, S. and Duk, R. (2006). " Metabolic profiles as a reflection of physiological status" *Journal of integrative Bioimformatics. At:* www.journal.imbio.de/../jib-28.pdf . Accessed on 9th May 2013.

Mooney, C.Z. and Duval, R.D. (1993). *Bootstrapping: A nonparametric approach to statistical inference.* Newbury Park, CA: Sage Publications.

Morris, J. D. (1982). Ridge regression and some alternative weighting techniques: A comment on Darlington. *Psychological Bulletin*, **91**: 203-210.

Murad, W., Mlla R.I., Mokhtar, M.B. and Raquib A. (2010). Climate change and Agricultural growth: An examination of the link in Malaysia. *Int. J. Climate Change Strateg.Manag*. **2**:403-417.

Myers, R.H. (1986). *Classical and Modern Regression With Applications.* 2nd Ed. PWS-KENT Publishing Company.

Neter, J., Wasserman, W. and Kutner, M.H. (1990). *Applied Linear Regression Models.* 3rd Ed. IRWIN Book Team USA.

Neter, J., Kutner, M.H. and Wasserman, W. (2005). *Applied Linear Regression Models.* 5th Ed. McGraw Hill.

Nigerian Environmental Study Team (NEST), (2004). Regional Climate Modelling and Climate Scenarios Development in Support of Vulnerability and Adaptation Studies: Outcome of Regional Climate Modeling Efforts over Nigeria, NEST, Ibadan, Nigeria. *In CBN Journal of Applied Statistic*, 2 **(1)**:32-50.

Oboh, B. O. and Fakorede, M. A. B. (1999). Effeces of Weather on Yield\ Component of the Oil Palm in a Forest Location in Nigeria. *Journal of Oil Palm Research* II **(I)**:79-89.

Okpamen, U.S., Uwumarongie-Ilori, E.G., Okere, R.A., Imoisi, O.B. and Okolo, E. C. (2012). Conceptualization of Info- Crop Oil Palm Nutrient Modeling In Nigerian Institute for Oil Palm Research. Greener *Journal of Agricultural Sciences.* 2 **(8)**: 424-431.

Ong, H T (1982). System approach to climatology of oil palm. I. Identification of rainfall and dry aspects. *Oleagineux*, **37**: 93-105.

Oyekale, A.S., Bolaji, M.B. and Olowa, O.W. (2009). The effects of climate Change on cocoa production and vulnerability assessment in Nigeria. *Agricultural Journal,* 4**(2**):77-85.

Pagel, M. D. & Lunneberg, C. E. (1985). Empirical evaluation of ridge regression *Psychological Bulletin*, **97**:342-355.

Parry, M., Arnell, N., Berry, P., Dodman, D., Fankhauser, S., Hope, C., Kovats, S., Nicholls, R., Satterthwaite, D., Tiffin, R. and Wheeler T. (2009). *Assessing the Costs of Adaptation to Climate Change – A Review of UNFCCC and Other Recent Studies*, International Institute for Environment and Development and Grantham Institute for Climate Change, Imperial College London, U.K. In *CBN Journal of Applied Statistic*, 2 **(1)**:33-50.

Rolf, S. (2002). "Continuum Regression "Mathematical statistics Stockholm university, Research report No 4 . At: www.matematik.su.se/matsat. Accessed on 9th May 2013.

Sana, M. and Eyup, C. (2008). Efficient Choice of Biasing Constant for Ridge Regression**,** *Int. J. Contemp. Math. Sciences***, 3**:527 – 536.

Sparnnaij, L. D., Rees, A. R. and Chapas, L. C. (1963). Annual yield variation in the oil palm. *J.* W. *Afr. Inst. Oil Palm Res*., **4**:11.125.

SPORE, (2008); Climate Change, Spore Special Issue-August, 2008. *In CBN Journal of Applied Statistic,* 2 **(1):**31-50.

Stone, M.and Brooks, R.J. (1990). Continuum Regression .Cross-Validated Squentially constructed Prediction Embracing OLS,PLSR,PCR *Journal of royal Statistics Society* **52**:237-269.

Sundberg, R. (1993). Continuum Regression and Ridge Regression. *Journal of royal Statistics* **55**:653-659.

The American Heritage® Dictionary of the English Language, Fourth Edition copyright ©2000 by Houghton Mifflin Company. Updated in 2009. Published by Houghton Mifflin Company. At: http://www.thefreedictionary.com/Relative+air+humidity. Accessed on May, 2013.

Vaughan, T. S. and Berry, K. E. (2005). Using Monte Carlo techniques to demonstrate the meaning and implications of multicollinearity. *Journal of Statistics Education* [Online], 13 **(1).** At: www.amstat.org/publications/jse/v13n1/vaughan.html. Accessed on 9th May 2013.

Wang, J. Mendelsohn, R., Dinarc, R., Huangd, J., Rozellee, S., and Zhangd, L. (2009). The impact of climate change on China's agriculture, *Agricultural Economics* **40**: 323–337 .

Watson, R.T., Zinyoera, M.C., and Moss, R.H. 1996. Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analysis. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.

WEP (2011). Assessment of Gender Knowledge and Awareness, Vulnerability and Adaptation Strategies to the Impacts of Climate Change in Northern Nigeria. A Research conducted by Women Environmental Programme (WEP) with the support from CIDA and BNRCC.

Wethrill, H. (1986). Evaluation of ordinary Ridge Regression, *Bulletin of Mathematical Statistics***, 18**:1 – 35.

Wright, S. (1923). Correlation and causation. *J. Agric. Res.,* **20***:557-585.*

Zahid, Z., Mad, N. S., Zainal, A. M. and Shehu, U.A. (2012). Economic Impact of Climate Change on the Malaysian Palm Oil Production. *Trends in Applied Sciences Research. Academic Journals Inc.* 7(**10**):872-880.

APPENDICES

APPENDIX A

This is the result of the analysis of the of correlation matrix on Table 4.1, multiple regression (OLS) on table 4.2/4.3/4.4 and multicollinearity diagnostics on table 4.5/4.6.

Importing data from excel to R

Save data in CSV (comma delimited) (*. Csv) in excel

Then go to R environment and type: read.csv("abnew3.csv")

Note abnew3 is the file name. The following steps are also given before the data analysis in R

abnew3<-read.csv("abnew3.csv")

>names(abnew3)

>attach(abnew3)

>attach(abnew3)

Pearson Correlations using R package

>wes = cbind (abnew3)

> co = cor (wes)

> co

Multiple Rregression (OLS) using R package

>fit < - lm(abnew3) or

>fit < - lm (FFB~ HUMIDITY+RADIATION+ RAINFALL+ SUNSHINE+ TEMPERATURE)

>fit

FFB = 3.836e+03 + 0(HUMIDITY) + 0(RADIATION) - 0(RAINFALL) - 4.884e+00(SUNSHINE) - 1.323e+02 (TEMPERATURE)

>summary(fit)

```
>anova (fit)
```

Multicollinearity Diagnostics using R package

Eigen Values and Several Condition Numbers

```
>x<-as.matrix(abnew3[,-6])

>  e<-eigen(t(x)%*%x)

>  e$val

[1] 7.843241e+07 1.132837e+07 3.241683e+06 2.270487e+04 5.993117e+02

>  sqrt(e$val[1]/e$val)

[1]   1.000000   2.631262   4.918838  58.774402 361.760719
```

Variance Inflation Factor

```
>library(HH)

>vif(fit)
```

HUMIDITY	RADIATION	RAINFALL	SUNSHINE	TEMPERATURE
1.801491	14.471980	1.865858	1.820838	15.314389

APPENDIX B

This is the program that shows the result of stepwise regression, ridge regression and principal component regression used to solve the problems of multicollinearity as well as the comparisons among the methods as displayed in the tables.

Stepwise Regression (backward elimination) using SPSS

Input the data into the data view

Define the variables name in the variable view

Click on analyze and select regression, then extend the cursor to linear

Determine the dependent and independent variables as well as the relevant statistics in the regression dialog box that appears

Select backward for method in the same dialog box and click on OK for the result to be display

FFB = 3591.971 - 4.757(SUNSHINE) – 120.360 (TEMPERATURE)

Ridge Regression using R package

```
>library (ridge)

>mod <-linearRidge (FFB~.-1, data = abnew3, lambda= "automatic");

>summary (mod)

>plot (mod)

>mod1 <-linearRidge (FFB ~ . -1, data = abnew3, lambda = seq(0,0.25,0.05));
> summary(mod1)

>matplot(mod1$lambda,t(mod1$coef),type="l",xlab=expression(lambda),

ylab=expression(hat(beta)))

>  abline(h=0,lwd=2)> summary(mod1)
```

FFB = 7.722e-01 (HUMIDITY) -7.672e-01 (RADIATION) + 4.538e-02 (RAINFALL) -8.971e+00 (SUNSHINE) -5.723e+01 (TEMPERATURE)

Principal Component Regression using R package

```
> x <- as.matrix(abnew3[,-1])
> e <- eigen(cor(x))
>dimnames(e$vectors)<list(c("HUMIDITY","RADIATION","RAINFALL","SUNSHINE", "TEMERATURE"),paste("EV",1:5))
>  round(e$vec,3)
> plot(e$values,type="l",xlab="EV no.")
> nx <- scale(x)
> enx <- nx %*% e$vec
> fit1<- lm(abnew3$FFB ~ enx)
> summary(fit1)
>round(summary(fit1)$cov.unscaled,2)
>par(mfrow=c(3,2))
>for(i in 1:5) plot(abnew3[,5],abnew3[,i],xlab="TEMPERATURE",
+ ylab=names(abnew3)[i])
> summary(lm(FFB~ TEMPERATURE + RAINFALL+I(SUNSHINE-TEMPERATURE),abnew3))
```

FFB = 3.594e+03 -1.252e+02 (TEMPERATURE) -1.078e-03 (RAINFALL) -4.780e+00 I(SUNSHINE - TEMPERATURE)

APPENDIX C

OIL PALM YIELD AND CLIMATIC CONDITIONS DATA AT NIFOR NIGERIA

The data collected, processed and used for this study was obtained from Nigeria Institute for Oil Palm Research (NIFOR), composed of monthly yield of fresh fruit bunch of oil palm, average relative humidity, solar radiation, rainfall, mean sunshine and average air temperature for the period of 1980 – 2012.

Table A1: Oil Palm Yield and Climatic Conditions Data

S/NO	MONTH	YEAR	FFB	HUMID	RADIAT	RAIN	SUN	TEMP
1	January	1980	93.83	70.95	380.9	6.1	7.3	28.81
2	February		93.96	68.2	395.14	26.8	6	28.93
3	March		96.11	70.1	410.9	60.8	5.3	29.09
4	April		97.05	73.85	403.09	212.3	6.2	29.02
5	May		93.88	76.55	377.5	266.2	5.8	28.86
6	June		93.66	80.75	355.75	329	5.6	28.64
7	July		431.94	85.25	293.49	284	2.8	26.92
8	August		221.12	86.4	309.25	307.4	3.1	27.09
9	September		281.45	80.9	344.05	319.2	4.2	27.43
10	October		93.52	81.15	350.94	203.7	4.7	28.49
11	November		93.92	75.7	391.81	109.9	6.9	28.9
12	December		93.74	65.05	373.76	3.55	6.6	28.72
13	January	1981	93.65	69.8	364.38	4.3	6	28.63
14	February		36.23	60.75	422.12	2.8	6.7	29.21
15	March		36.04	70.05	403.4	57.8	5.9	29.02
16	April		36.85	72.1	384.52	99.4	6.6	29.83
17	May		93.91	75.9	390.94	294.7	5.8	28.89
18	June		93.97	76.85	386.43	254	6.5	28.95
19	July		431.96	84.2	295.33	319.3	3.5	26.94
20	August		431.11	85.9	291.2	365.1	2.7	26.9
21	September		241.48	82.4	337.2	299.2	4	27.46
22	October		93.87	77.7	376.69	187.6	5.5	28.85
23	November		36.15	68.3	414.08	23.5	6.8	29.13
24	December		93.89	73	378.1	15.3	7.2	28.87
25	January	1982	93.55	68.05	354.8	4.2	6.5	28.53
26	February		93.73	62.4	372.7	13.7	5.1	28.71
27	March		96.05	65.05	404.12	31.5	5.8	29.03
28	April		96.03	71.7	402.12	243.5	6.5	29.01
29	May		93.89	74.65	378.28	254.4	6.1	28.87
30	June		93.29	78.85	347.17	181.1	4.7	28.26
31	July		431.89	83.9	278.9	390.2	2.9	26.87

Table A1 Continued

32	August		441.97	83.35	286.5	125.9	2.4	26.95
33	September		221.18	82.75	307.86	208.5	2.9	27.16
34	October		93.53	80.2	352.7	230.4	4.3	28.51
35	November		93.81	72.5	380.4	79.8	6.6	28.79
36	December		93.77	71.45	366.96	155.1	6.7	28.75
37	January	1983	93.79	50.8	368.2	11.05	3.4	28.77
38	February		36.21	55.3	419.9	18.9	4.5	29.19
39	March		36.03	60.05	402.7	3.2	3.9	29.01
40	April		36.27	62.8	416.3	123.8	6.6	29.25
41	May		241.25	73.9	324.5	297.2	6.6	27.23
42	June		431.85	82.6	284.6	238.4	3.4	26.83
43	July		211.05	81.7	304.7	172.4	3.5	27.03
44	August		521.42	85.8	241.3	106.5	1.5	25.4
45	September		241.31	82.85	330.5	338	3.9	27.29
46	October		93.86	74.9	375.4	81.7	5.3	28.84
47	November		36.14	73.65	413.7	75.9	6.9	29.12
48	December		93.55	75.45	354.1	42.7	6.8	28.53
49	January	1984	93.63	67.1	362.2	59.3	4.7	28.61
50	February		36.12	58.4	411.4	13.5	5.9	29.1
51	March		36.26	68.45	415.5	65.9	6.2	29.24
52	April		93.83	72.6	382.4	188.5	5.5	28.81
53	May		93.94	76.2	393.1	306.5	5.8	28.92
54	June		95	75.85	398.2	136.9	5.8	28.99
55	July		93.72	80.55	371.2	270.4	5.4	28.7
56	August		93.52	80.25	351.5	316.7	4.7	28.5
57	September		93.67	78.95	356.45	213.5	4.5	28.65
58	October		93.92	77.2	391.9	272.6	5.5	28.9
59	November		36.26	70.9	416.2	19.1	5.9	29.25
60	December		93.95	65.3	394.6	3.1	5	28.93
61	January	1985	93.91	69.85	389.2	6.1	4.7	28.89
62	February		36.39	61.85	428.8	0.1	3.1	29.37
63	March		36.11	75.4	409.1	218.3	4.2	29.09
64	April		36.41	75.05	439.7	77.6	5.4	29.39
65	May		36.31	80.75	430.7	295.9	6.7	29.29
66	June		36.27	75.85	416.1	222.2	5.3	29.25
67	July		241.04	81.75	329.6	326.7	3.7	27.29
68	August		311.47	83.55	336	352.7	3.6	27.45
69	September		93.88	81.25	377.6	168.7	4.2	28.86
70	October		36.03	76.9	402.7	178.1	5.1	29.01
71	November		36.41	73.7	440.2	33.7	7.5	29.39
72	December		36.11	59	406.3	105.9	6.4	29.09
73	January	1986	36.23	66.9	421.1	3.4	6	29.21
74	February		36.31	65.05	432.6	16.6	6.8	29.29
75	March		93.95	69.65	394.7	157.3	5.4	28.93

Table A1 Continued

76	April		36.35	71.85	434	122.3	7	29.33
77	May		36.39	73.3	428.3	217.7	6.3	29.37
78	June		36.35	76.2	434.6	221.1	6.8	29.33
79	July		311.46	83.05	335.2	239.8	2.4	27.44
80	August		221.17	80.5	306.1	140.9	3.3	27.15
81	September		221.16	82.6	305.2	300.2	3.5	27.14
82	October		93.74	80.5	373.3	277.7	5	28.72
83	November		36.46	65.8	435.1	41.4	6.9	29.44
84	December		36.16	77.15	405.9	159.55	6.6	29.14
85	January	1987	36.16	67.65	405.1	65.6	7.5	29.14
86	February		36.35	69	434.8	64.8	7.4	29.33
87	March		36.13	69.55	412.2	66.4	5.8	29.11
88	April		36.51	70.25	450.8	51.4	5.2	29.49
89	May		36.36	75.4	425.9	143.8	6.1	29.34
90	June		93.91	75.75	390.9	205.7	5.6	28.89
91	July		350.58	79.1	347	270.5	4.4	27.56
92	August		211.11	85.35	310.9	349.2	2.7	27.09
93	September		350.57	81.55	346.5	345.9	3.9	27.55
94	October		36.05	74.05	404.8	133.1	5.9	29.03
95	November		36.44	71.2	443.8	2.7	7.5	29.42
96	December		36.16	68.45	405.1	67.9	7.4	29.14
97	January	1988	93.71	64.7	369.4	8.8	4.5	28.69
98	February		36.11	68.15	410.7	66.1	6	29.09
99	March		36.03	74.45	402.4	109.3	4.7	29.01
100	April		36.19	72.05	408.9	220.7	5	29.17
101	May		95	74.1	400.3	139.4	6	28.99
102	June		93.51	77.25	350.5	222.1	5	28.49
103	July		431.84	84	283.5	279.2	3.1	26.82
104	August		411.67	84.8	256.8	174.4	2.3	26.65
105	September		443	83.65	295.3	471.6	3.1	26.99
106	October		91.81	79	379.5	178.3	5.6	28.79
107	November		36.49	71.6	438.1	3.4	7.6	29.47
108	December		93.62	73.1	361.1	30.6	6.1	28.6
109	January	1989	36.49	53.05	438.7	17	8.1	29.47
110	February		36.11	65.65	410.4	0.9	5.4	29.09
111	March		93.96	67.55	385.5	58.4	5.6	28.94
112	April		36.27	68.25	416.2	87.4	7	29.25
113	May		93.88	72	377.4	217.8	6.3	28.86
114	June		271.38	75.85	327.8	410.3	5.3	27.36
115	July		241.22	79.05	321.6	371.2	5	27.2
116	August		431.88	82.2	277	546.8	3	26.86
117	September		311.48	79.95	343.3	190.3	4.1	27.46
118	October		93.71	76.8	369.7	194.1	5.4	28.69
119	November		36.25	68.75	424.4	1.6	8	29.23

Table A1 Continued

120	December		93.91	65.7	389.6	24.3	7.3	28.89
121	January	1990	93.55	71.3	354.6	5.3	7.6	28.53
122	February		456.77	69.5	411.6	43.3	6.6	29.1
123	March		36.67	65.5	456.9	0.6	7	29.65
124	April		36.11	68.55	410.7	210.4	7.3	29.09
125	May		36.05	72.2	404.9	73.9	7.1	29.03
126	June		93.61	76.7	359	175.6	5.9	28.59
127	July		491.44	85.3	243.9	497	2	25.42
128	August		443	84.85	299.3	411.6	3.4	26.99
129	September		311.44	79.2	343	343.7	3.7	27.42
130	October		91.83	77.25	381.7	231.6	5.3	28.8
131	November		36.26	76.7	415.5	42.6	7.1	29.24
132	December		93.86	77.25	375.3	79.4	7.2	28.84
133	January	1991	36.12	76.35	411	61	7.1	29.1
134	February		36.32	73.6	431.1	93.7	7.7	29.3
135	March		36.46	72.55	435.5	66	6.8	29.44
136	April		36.22	75.7	421.6	267.9	6.6	29.2
137	May		93.95	74.15	394.4	240.8	6	28.93
138	June		93.72	77.5	385	205	6.5	28.7
139	July		431.83	80.75	282.3	453.1	2.5	26.81
140	August		411.68	80.35	266	344.9	1.8	26.68
141	September		93.61	79.6	357.6	251.3	3.7	28.59
142	October		93.91	78.35	386.2	133.9	4.7	28.88
143	November		36	76.05	399.2	3.3	6.7	29
144	December		36.17	73.1	406	17.9	6.3	29.15
145	January	1992	36.19	61.85	409.1	10.6	5.4	29.18
146	February		36.77	63.95	466.7	2.3	6.5	29.75
147	March		93.85	71.7	384.7	59.8	4.2	28.83
148	April		93.84	71.25	383	157	5.4	28.82
149	May		36	74.25	398.3	234.9	6.5	29
150	June		241.3	78.75	329.8	296.5	4.7	27.28
151	July		491.44	81.5	243.4	258.5	2.4	25.42
152	August		421.76	81.2	265.3	121.2	2.9	26.74
153	September		221.19	79.95	309.9	167.6	3.7	27.18
154	October		93.82	78.45	381.9	249.7	6.2	28.8
155	November		36.16	75.6	405.3	47.1	7.4	29.14
156	December		36.37	76.2	426.5	148.4	7.7	29.35
157	January	1993	36.16	67.75	405	97.75	5	29.14
158	February		36.39	65.55	432.8	151.5	7.3	29.37
159	March		36.69	73.05	458.8	262.2	6.5	29.67
160	April		36.28	73.9	417.2	108.8	5.8	29.26
161	May		93.86	72.6	375.6	328.5	5.9	28.84
162	June		311.45	77.4	344.4	227.4	4	27.43
163	July		431.87	77.9	276.8	222	2.3	26.85

Table A1 Continued

164	August		421.72	82.55	271.6	259.7	1.8	26.7
165	September		311.42	80.05	341.8	283.4	3.8	27.4
166	October		93.8	76.95	380.8	170.7	5.8	28.79
167	November		93.71	73.1	370.4	87	6.2	28.69
168	December		95	76.3	396.4	128.85	5.9	28.99
169	January	1994	93.98	69.6	387	45.3	5	28.96
170	February		36.03	72.95	399	16.1	5.1	29.01
171	March		93.94	71.28	393	30.7	3.9	28.92
172	April		95	44.45	396	214	5.9	28.99
173	May		93.63	83.7	362.6	144.9	6.3	28.61
174	June		271.39	85.65	328.8	344.4	5.6	27.37
175	July		95	81.75	396.4	407	2.2	28.99
176	August		411.62	89.55	261.2	466	1.1	26.6
177	September		501.31	89.55	227.4	482	1.8	25.29
178	October		521.46	87.35	235.2	366.4	1.45	25.44
179	November		241.31	80	328.8	21.4	7	27.29
180	December		36.38	71.3	427.6	193.9	4.23	29.36
181	January	1995	36.71	83.8	469.2	107.65	6.5	29.69
182	February		36.36	81.9	425.4	96.5	7.2	29.34
183	March		36.54	79.35	453.6	216.6	6.3	29.52
184	April		36.67	81.15	456.2	131	7	29.65
185	May		36.18	79.45	417.8	215.5	6.4	29.16
186	June		36.23	82.6	422.4	198.2	6.1	29.21
187	July		221.14	94.6	313.2	324.6	3.2	27.12
188	August		431.87	90.85	276.8	467.7	2.5	26.85
189	September		93.9	87.2	388.6	272.7	4.7	28.88
190	October		36.51	81.25	445.8	155.2	5.4	29.49
191	November		36.48	80.45	443.2	25.9	7.8	29.46
192	December		166.53	86.55	456.2	90.55	6.45	29.65
193	January	1996	93.69	88.75	358.7	44.6	6.3	28.67
194	February		93.779	70.75	368.9	60.3	6.6	28.77
195	March		93.73	80.95	372.5	106.1	5.7	28.71
196	April		93.78	84.6	367.9	71.2	6.1	28.76
197	May		36.71	87.65	469.5	277.2	5.7	29.69
198	June		221.21	91.2	320.8	466.7	4.2	27.19
199	July		221.21	89.2	318.3	232.7	3.7	27.19
200	August		411.71	91.8	266.9	290.6	3.2	26.69
201	September		211.11	88.5	307.4	453.9	3.1	27.09
202	October		241.31	82.6	329.7	219.6	5.9	27.29
203	November		36.21	77.95	418.3	44	7.2	29.19
204	December		271.35	79.5	334.7	131.8	7.2	27.33
205	January	1997	93	79.2	370.6	41.6	6	28
206	February		93.95	72.75	394.1	111.3	5.7	28.93
207	March		271.41	73.65	338.9	114.1	4.6	27.39

Table A1 Continued

208	April		241.31	77.75	326.9	108.5	5.1	27.29
209	May		271.32	80.7	331	280.1	5.4	27.3
210	June		211.03	85.55	302.2	315	4.6	27.01
211	July		241.34	83.15	333.2	161.5	3.5	27.32
212	August		421.74	87.55	274.5	152	2.8	26.73
213	September		271.35	86.65	334	232.1	4.2	27.33
214	October		271.41	85.15	337.9	253.3	6	27.39
215	November		93.91	82.55	390.4	47.8	7	28.89
216	December		36.65	83.9	464	0.9	6.9	29.63
217	January	1998	93.91	63.8	388.1	9.5	5.7	28.89
218	February		40	65.8	431.8	20	7.6	29.3
219	March		36.41	57.85	439.3	50.4	5.7	29.39
220	April		36.11	72.65	410.8	129.8	6.8	29.09
221	May		93.71	75	370.4	143.2	6.2	28.69
222	June		93.82	76.25	381.6	177.5	5.4	28.8
223	July		201	83.3	301	246.6	4	27
224	August		421.81	82.15	276.8	59.9	2.5	26.79
225	September		211.11	84.1	309	499.5	4.8	27.09
226	October		311.51	80.6	345.7	251	7.5	27.49
227	November		36.11	73.3	406.3	28	2.7	29.09
228	December		93.92	66.5	391.7	139.5	7.1	28.9
229	January	1999	93.74	69.05	373	29.8	5.4	28.72
230	February		36.15	68.5	414.6	54.4	6	29.13
231	March		36.44	71.75	443.2	89.1	5.8	29.42
232	April		93.91	74.45	389.6	166.6	5	28.89
233	May		93.71	74.65	370.4	262.1	5.6	28.69
234	June		443	79.7	295	236	4.7	26.99
235	July		401.61	84.05	258.6	241.5	3.3	26.59
236	August		231.71	83.7	269	172.9	3.3	26.69
237	September		471.31	82.65	227.4	399	5.76	25.29
238	October		491.42	82.15	241.2	282.5	5.69	25.4
239	November		441.93	67.15	292.4	23.8	5.83	26.91
240	December		221.21	65.45	318.4	153.15	5.55	27.19
241	January	2000	461.15	63.75	214.7	5.8	6.1	25.13
242	February		441.94	42.95	293.4	11.8	5	26.92
243	March		271.34	61.35	333.5	67.9	7.7	27.32
244	April		441.93	72.5	292	153.1	6.6	26.91
245	May		411.45	73	344.4	92.4	6.8	27.43
246	June		431.91	76.25	287.2	434.9	4.7	26.89
247	July		481.33	82.3	232.6	220.8	3.3	25.31
248	August		421.74	84.8	273.2	241.9	2.1	26.72
249	September		431.85	82.5	284.3	438.8	3.4	26.83
250	October		221.21	76.2	318.4	228.2	3.6	27.19
251	November		93.81	71.6	378.2	16.7	4.9	28.79

Table A1 Continued

252	December		221.21	62	318.4	122.45	4.2	27.19
253	January	2001	93.91	63.85	386.3	11	3.5	28.89
254	February		36.11	53.65	406.8	1	5.5	29.09
255	March		36.31	69.35	425	152.3	4.4	29.29
256	April		93.91	69.4	386	237.7	5.2	28.89
257	May		93.91	73.55	387.3	182.1	4.7	28.89
258	June		365.91	80.05	347	257.9	4	27.49
259	July		443	83.75	297.6	353.2	2.7	26.99
260	August		491.51	85	245.6	139.8	1.3	25.49
261	September		431.91	81.65	289.8	343.3	2.2	26.89
262	October		93.74	76.55	373	114.4	3.7	28.72
263	November		36.31	69	425	18.9	5.2	29.29
264	December		36.11	65.55	406.8	3.9	6.5	29.09
265	January	2002	93.11	52.25	391.2	11.4	4.5	28.9
266	February		36.32	60.75	431.8	27.8	5.3	29.3
267	March		93.11	74.25	391.2	133.6	5.6	28.9
268	April		93.11	74.7	391.2	209.8	6.7	28.9
269	May		93.81	75.3	375.6	201.5	5.6	28.79
270	June		93.63	76.8	362.6	356.6	4.5	28.61
271	July		221.14	83.3	313.2	437.3	1.9	27.12
272	August		411.62	84.4	261.2	308.5	1.1	26.6
273	September		271.62	81.5	336.6	180.9	3	27.39
274	October		93.64	79.75	363.1	237.1	4	28.62
275	November		36.21	71.9	417.2	42.7	6	29.19
276	December		36.13	58.25	412	139.9	6	29.11
277	January	2003	93.91	69.3	385.3	49.3	6	28.89
278	February		36.41	70.9	435.4	26.9	5	29.39
279	March		36.21	69.8	417.2	68.3	4.6	29.19
280	April		95	72.7	399.2	250.8	4.8	28.99
281	May		95	76.85	399	181.2	6.2	28.99
282	June		93.53	77.8	352.2	162.9	4.1	28.51
283	July		241.31	79.6	328.8	155	4.3	27.29
284	August		201	83.05	301.8	170.1	2.3	27
285	September		421.81	85.7	279.4	313.5	1.8	26.79
286	October		93.81	76.65	380.8	293.7	4.3	28.79
287	November		271.41	74.65	338.1	31.3	4.2	27.39
288	December		93.91	66.85	388.6	162.5	6.1	28.89
289	January	2004	93.71	64.55	365.1	35.2	7.2	28.69
290	February		36.11	61.6	406.8	13.5	4.6	29.09
291	March		93.93	61.6	392.2	55.3	1.1	28.91
292	April		93.71	76.15	369.9	106.4	3.6	28.69
293	May		93.86	75.9	384.9	323.4	4.5	28.83
294	June		93.61	76.5	359.6	355.7	4.2	28.59
295	July		441.95	83.4	294.3	214.3	2.7	26.93

Table A1 Continued

296	August		441.95	87.6	294.8	298.6	1.4	26.93
297	September		241.25	80.6	324.6	251.1	2.2	27.23
298	October		93.81	76.6	379.3	247	3.6	28.79
299	November		93.91	73.65	386	28.3	5.4	28.89
300	December		95	69.6	396.1	137.65	5	28.99
301	January	2005	36.71	53.4	469.2	91.45	5.3	29.69
302	February		36.31	66.1	425.4	15.7	6.7	29.29
303	March		36.54	73.45	453.6	167.2	5.5	29.52
304	April		36.61	74.45	456.2	114.4	5.4	29.59
305	May		36.21	75.8	417.8	138.9	4.8	29.19
306	June		36.23	80.7	422.4	292.7	4.7	29.21
307	July		221.14	83.7	313.4	406.8	2.2	27.12
308	August		421.81	83.8	276.8	80.9	1.5	26.79
309	September		93.91	80.05	388.6	177.3	3.8	28.89
310	October		36.51	76.95	445.8	167.2	4.1	29.49
311	November		36.44	68.25	443.2	33.9	6.7	29.42
312	December		36.57	72.45	456.2	100.55	6.3	29.55
313	January	2006	93.71	73.4	361.3	22.5	4.4	28.6
314	February		93.91	70.8	385.8	10.5	4.7	28.89
315	March		93.91	70.95	390.2	61.1	2.9	28.89
316	April		93.83	71.85	382.5	158	4.8	28.81
317	May		271.35	77.65	334.9	246.8	3.7	27.33
318	June		93.83	77.3	382.4	172.5	4.6	28.81
319	July		441.95	82.1	294.4	289	2.3	26.93
320	August		421.72	85.7	271.4	359.9	1.4	26.7
321	September		241.23	83.35	322.6	347.4	2	27.21
322	October		93.81	77.55	375.9	304.5	3.5	28.79
323	November		36.21	67.65	415.2	24.7	5.7	29.19
324	December		36.14	62.9	413.9	164.6	5.8	29.12
325	January	2007	36.04	46.6	403.7	80.2	7.5	29.02
326	February		36.21	65.7	419.9	104.2	7.4	29.19
327	March		411.71	66.1	268.5	56.2	5.8	26.69
328	April		36.04	74.1	403.3	197.7	5.2	29.02
329	May		36.11	75.45	410.6	246.2	6.1	29.09
330	June		93.71	71.5	365.5	380.9	5.6	28.69
331	July		221.12	83.25	311.5	284.7	4.4	27.1
332	August		411.91	89	290.6	171.4	2.7	26.89
333	September		311.51	75.45	349.1	256	3.9	27.49
334	October		93.71	70	367.3	285	5.9	28.69
335	November		93.81	53.7	375.6	37.1	7.5	28.79
336	December		93.91	64.65	390.2	17.1	7.4	28.89
337	January	2008	93.52	55.75	351.5	193.6	4.5	28.5
338	February		36.81	50.9	476.3	27.1	6	29.79
339	March		36.35	69.4	434	95.3	4.7	29.33

Table A1 Continued

340	April		36.04	72.5	403.5	98.3	5	29.02
341	May		36.21	73.9	417.5	137.1	6	29.19
342	June		93.71	75.35	365.8	256.6	5	28.69
343	July		241.31	84.4	327	276.6	3.1	27.29
344	August		311.42	79.7	341.6	313.9	2.3	27.4
345	September		93.61	80.6	356.2	371.7	3.1	28.59
346	October		93.84	74.2	383.7	68.6	5.6	28.82
347	November		95	66.9	397.6	6.9	7.6	28.99
348	December		93.71	66.35	369	23.6	6.1	28.69
349	January	2009	93.61	65.35	356.7	1.6	2	28.59
350	February		93.74	70.5	373.8	134.9	2.2	28.72
351	March		36.35	71.45	434.1	78.3	0.8	29.33
352	April		30.34	74.4	528.4	226.6	3.6	30.29
353	May		93.71	75	366.8	248.6	5	28.69
354	June		311.45	79.5	344.9	207.7	4.2	27.43
355	July		441.94	82.15	293.4	148.7	2.9	26.92
356	August		417.72	84.3	271	254	2.2	26.7
357	September		211.11	83.75	310.7	278.1	3.4	27.09
358	October		441.94	80.45	293.7	192.8	3.8	26.92
359	November		93.81	71.95	377.8	109.4	6.8	28.79
360	December		36.11	67.05	407.8	1.3	7.5	29.09
361	January	2010	93.85	66.45	384.3	55.35	5.2	28.83
362	February		36.04	69	402.5	57.5	5.2	29.02
363	March		311.51	71.7	348.3	38.7	3.3	27.49
364	April		93.83	71.35	382.4	219.9	5.8	28.81
365	May		93.71	75.65	366	125.4	5.3	28.69
366	June		241.3	78.6	327.1	174.6	3.8	27.29
367	July		411.65	84.1	264.9	257.8	2.9	26.63
368	August		491.42	87.7	241.5	455.8	2.4	25.4
369	September		443	83.85	296.3	282.1	3.6	26.99
370	October		241.31	83.1	330.8	373.8	5.5	27.29
371	November		93.71	76.6	368.8	109	6	28.69
372	December		36.25	66.35	424.7	241.4	7.1	29.23
373	January	2011	36.06	53.35	405.3	100.55	5.4	29.04
374	February		93.9	69.9	388	116.2	5.9	28.89
375	March		36.31	72.3	426.2	84.9	6.1	29.29
376	April		93.6	74.85	359	118.3	5.1	28.59
377	May		93.71	77.95	369.2	264	5.6	28.69
378	June		221.2	81.8	320.3	275.2	3.3	27.19
379	July		491.45	82.8	244.5	430.3	2.4	25.43
380	August		471.31	85.9	228.6	277.8	1.6	25.29
381	September		441.95	84	294.9	250.9	2.5	26.93
382	October		36.72	78.95	471.6	240.8	4.4	29.7
383	November		36.11	74.1	409.6	68.8	7	29.09

Table A1 Continued

384	December		93.83	40.65	382	154.8	7.3	28.81
385	January	2012	311.51	69.4	348	65.4	5.4	27.49
386	February		93.81	73.75	375.9	34.4	4.9	28.79
387	March		36.41	68.95	438.3	45.4	5.4	29.39
388	April		93.74	74.55	373.8	162.4	5.7	28.72
389	May		711.41	77.4	36.8	188.8	5.1	23.39
390	June		271.41	81.25	336.4	265.2	3.8	27.39
391	July		491.42	77.6	241.7	396.9	1.7	25.4
392	August		421.75	83.85	274.4	139.9	2.6	26.73
393	September		211.05	81.2	304.8	317.5	3.6	27.03
394	October		93.61	79.8	351.1	178.9	4.7	28.5
395	November		93.71	75.7	367.9	46.9	6.3	28.69
396	December		36.14	75.8	413	112.9	7.6	29.12

Table Keys:

FFB: Fresh Fruit Bunch of Oil Palm Yield.

HUMID: Average Relative Humidity

RADIAT: Solar Radiation

RAIN: Rain Fall

SUN: Mean Sunshine

TEMP: Average Air Temperature