A Principal Component Regression Approach to Determine the Relationship Between Yam Yield and Some Climatic Determinants.

[1]Amokaha, O.A, [2]Apeagee, B. B, [3]Agada, I. O, [4]Akosu, A.M

[1]Department of Computer Science Education, Federal College of Education Technical, Isu, Ebonyi State
[2]Benue State School of Nursing Makurdi.
[3]Department of Physics, Joseph Sarwuan Tarka University Makurdi
[4]Department of Electrical Electronics Technology, Federal Polytechnic Wannune, Benue State.

Abstract— **This work applied the principal component regression approach in modelling the relationship between yam yield and some climatic variables in Makurdi, Benue State, Nigeria. Secondary data were sourced from Benue Agricultural and Rural Development Authority (BNARDA) and Nigerian Meteorological Agency Headquarters, Tactical Air Command, Makurdi – Airport, Benue State. It was established that there was multi-collinearity among the climatic variables. In this study, an attempt has been made to apply the concept of Principal Component Regression as a remedial solution to this problem. After establishing the existence of high collinearity between the independent variables, the concept of Principal Component Regression was applied to mitigate the effect of collinearity and find the best possible linear combinations of variables that can produce large variance with the best entropy. This study results include the fact that four principal components each were obtained for the first, second, and third farming seasons, yielding 81.11 %, 84.15 %, and 89.97 % of the total variability for the first, second, and third phases respectively. The first phase results of Principal Component Regression analysis obtained for eigenvalues were between 2.483233 to 0.063056, Incremental percentage of 35.47 to 0.9, and Condition number of 1.00 to 39.38. The second phase had eigenvalues between 1.976513 to 0.456305, an Incremental percentage of 32.94 to 7.61, and a Condition number of 1.00 to 4.33. And lastly, the third phase had eigenvalues between 1.609677 to 0.501677, an Incremental percentage of 32.19 to 10.03, and a Condition number of 1.00 to 3.21. The study has shown that the application of Principal Component Regression for mitigating the presence of multi-collinearity between the independent variable of climatic data can significantly be reduced and the best possible obtainable linear combination for independent components fit is achieved for estimation and predications of yam yield.**

Keywords: **principal, component, regression, climate, determinants, yields.**

1.0 INTRODUCTION

Climate variability is becoming the most important and severe environmental challenge facing various human activities, especially agriculture. The climate is the most crucial factor which determines the nature of the natural vegetation, the characteristics of the soils, the crops that can be grown, and the type of farming that can be practiced in different regions. Its variability has increased the likelihood of severe, pervasive environmental threats resulting in a variety of impacts on agricultural productivity in Makurdi Metropolis. [1] defined climate change as the change that can be attributed directly or indirectly to human activity that modifies the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable periods.
[2] as statistically significant variations in climate that persist for an extended period, typically decades or longer. Agriculture activities depend largely on climate, precipitation, solar radiation, wind, temperature, relative humidity, and other climatic variables that determine and affect the distribution of crops and livestock as well as their productivity or yields [3].
Although agricultural technologies have improved greatly recently, many researchers believe that the possible variations in climatic determinants still harm crop yields in many regions as climate determines the choice of what plant to cultivate, how to cultivate it as well as the yield of crops [4]. The two major methodologies used by researchers to study the relationship between climate variability and the yield of crops are the crop growth model and regression analysis. Crop growth modeling is a computer-based simulation approach used by many agronomists [5]. However, this study adopted the regression analysis model which is a widely used statistical technique used to estimate models that describe the distribution of crop yield with the help of two or more climatic variables.
The crop development model is quite complex and requires considerable information, such as climate variables, soil, and management options, to replicate the crop-growing process. However, both methods have drawbacks, and such information is typically partial and occasionally unavailable [6]. Multiple regression analysis mainly regards the interpretation of the regression coefficients, in the case of independent coefficients from regression analysis; the least-squares solution gives stable estimates and useful results [6]. In multiple linear regression models, the presence of multicollinearity often causes a huge explanatory problem during analysis. However, both models don't seem to eliminate the problem of collinearity. When multicollinearity is present in the data, ordinary least square estimators are inaccurately estimated [7].

This study aims to use Principal Component Regression Approach to determine the relationship between yam yield and some climatic determinants in Makurdi, Benue State-Nigeria.

Inadequate farm inputs, high cost of labor, and climate change can be identified as some constraints to crop production. Climatic change and its variability are becoming more unpredictable, particularly in Nigeria. Several studies on the impacts of climate variability on crop yield have been carried out in different places by different scholars. Some of these studies have shown significant impacts of climate variability on agricultural activities [8].
The trend in monthly and annual precipitation, minimum and maximum air temperature using the Mann-Kendall test and Sen's slope estimator to evaluate the impact of precipitation and temperature variability on crop yields was carried out using multiple regression analysis [9]. The effect of climatic change on rice production, the trend in rice production, and various factors that affect the output of rice in some Northern parts of Nigeria acknowledge variation in the trend of the climatic factors affecting rice production in the States [10].
A quantitative approach using a multiple linear regression model and descriptive statistics to determine the impact of the dependent variable (yam production data) and independent variables (temperature and rainfall data) was also employed [11]. The result revealed that moderate rainfall and temperature (sunshine and humidity) have a correlation, positive and preponderate effect on yam production. However, extreme rainfall and humidity destroy yam seedlings harming production and leading to a food shortage) [12]. It is a piece of common knowledge that when two independent variables used in a regression model have a high degree of correlation between them it could be said that multicollinearity exists among those variables [13, 14]. To address such a challenge, Principal Component Analysis (PCA) is used to reduce the dimension of high multi-collinearity independent variable components into low-dimensional high entropy principal components that could be used to migrate the effect of multi-collinearity and produce good regression models which are robust) [15].
It has been established that Principal Component Regression (PCR) is another technique for addressing multicollinearity problems in linear regression models [16]. This technique gives the same results similar to the regular method of sampling effects of estimators, however (PCR) has a great advantage for ease of computation. Thus, the estimator of the PCR is more efficient than the OLS estimator [16]. Different scholars used different test statistics to quantify the relationship between or among variables. However, this study employed Spearman's coefficient. It is important to note that when carrying out an analysis of variances test, the observed variables are not affected by a linear transformation, which alters the scale and origin of the variables. Thus, the transformation of data in a manner that does not seriously affect its entropy can also be considered to be a good alternative to the use of multivariate analysis of variance [17]

2.0 METHODOLOGY

2.1 Source and Nature of Data

The secondary data on yam production and the climatic data on rainfall amount (mm), temperature average (°c), sunshine hours, relative humidity (%), radiation, evaporation, and soil temperature for Makurdi metropolis for a period of 31 years (1986 – 2016). These data were sourced from the Benue Agricultural and Rural Development Authority (BNARDA) and Nigerian Meteorological Agency Headquarters, Tactical Air Command, Makurdi – Airport, Benue State respectively.

2.2 Data Transformation

The primary crop of interest used in this study is yam. The climatic determinants considered are; rainfall amount, soil temperature, air temperature, sunshine hours, relative humidity, evaporation, and solar radiation. These determinants were averaged over seven months. (March - September), the eight months (March - October) and the nine months (March - November) growth phases each.
The mean of each climate determinant is denoted by $\overline{X}$, and is the average of each determinant over each growth phase;

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

Where $\overline{X}$ = Arithmetic means of each climatic variable.
x_i = the monthly observations.
n = a total number of monthly observations.

2.3 Correlations of Climate Determinants

If multicollinearity exists between climatic determinants after transformation, a correlation matrix is obtained using Spearman's rank-order correlation. Spearman's correlation coefficient is a statistical measure of the strength of a monotonic relationship between paired data.
Spearman's correlation statistics are given as

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} \quad (2)$$

Where d^2 is the squared difference in ranks for each i observation and n is the sample size. The result of the correlation is presented in Table 2.

2.4 Mathematical Details of the Ordinary Least Square

The Ordinary Least Square is a type of linear least squares method for estimating unknown parameters in linear regression. The assumptions of the ordinary least square regression are stated as follows;
The first assumption is linearity which states that there is a linear relationship between y and X following the functional form;

$$y = X\beta + \epsilon \quad (3)$$

The second assumption is the strict exogeneity which states that the error ϵ terms should be independent of the value of the explanatory variables. X in the equation form;

$$E(\epsilon_i/X) = 0 \qquad E(\epsilon_i) = 0$$

The third assumption states that there is no perfect multicollinearity (columns of X should not be correlated with each other.) in other words, the columns of X are linearly independent. This assumption is known as the identification condition.
The fourth assumption states that the error terms should be homoscedastic, meaning they are evenly distributed around the X values. In equation form;

$E(\epsilon_i^2/X) = \sigma^2$

Consider the multiple regression model from (2.3) given below

$$Y = X\beta + \varepsilon \qquad 4$$

Where Y is the dependent variable, X is an n × m vector matrix of the independent variables, β is an n × 1 vector of the model parameter which B is the regression coefficients to be estimated, ε is an n × 1 vector which represents the errors or residuals and Y is an n × 1 vector of observation.

In ordinary least squares, the regression coefficients can be estimated using the formula which is derived as;

$\varepsilon = Y - X\beta.$

$\varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta).$

Setting equation (2) to Q, we have

$Q = (Y - X\beta)'(Y - X\beta).$

Differentiating Q concerning β and equating to zero.

$\frac{\partial Q}{\partial \beta} = -2X'(Y - X\beta) = 0.$

$-2X'Y + 2X'X\beta = 0.$

$2X'X\beta = 2X'Y.$

$X'X\beta = X'Y.$

$$\beta = (X'X)^{-1}X'Y \qquad 5$$

[18]

We proceed to transform the independent variables X to new independent variables Z called principal components which is a data matrix with a similar structure to X, mathematically written as;

$X'X = PDP' = Z'Z$. See proof in section 2.4.2.

2.4.1 Orthogonal Vectors

Orthogonal vectors are perpendicular such that if $a'a = 1$, the vector a is said to be normalized. The vector a can always be normalized by dividing by its length, $\sqrt{a'a}$. thus $p = \frac{a}{\sqrt{a'a}}$ is normalized so that $p'p = 1$.

A matrix $P = (p_1, p_2, \ldots, p_p)$ whose columns are normalized and mutually orthogonal is called an orthogonal matrix. Since the elements of $p'p$ products of columns of P which have the properties $p'p = 1$, thus $P'P = I$, If P is satisfied (Rencher, 2002) [19].

2.4.2 Singular Value Decomposition

Let X be an $n \times m$ rectangular matrix which can be factored into $U\sum V'$ where;

U is an orthogonal matrix.

$\sum$ is a diagonal matrix, whose diagonal entries are the single value of X.

P is an orthogonal matrix.

We proceed to show that $X'X = (U\sum P')'U\sum P'$

$X'X = P\sum'U'U\sum P'$

$U'U = I$, since U is orthogonal hence,

$X'X = P\sum'\sum P'$

Where $\sum$ is a diagonal matrix D and $\sum'\sum = D$.

Therefore; $X'X = PDP' = Z'Z$

2.4.3 Eigenvalue Analysis of the Correlation Matrix

The Eigenvalues of a correlation matrix can also be used to measure the presence of multicollinearity. If multicollinearity is present in the predictor variables, one or more of the Eigenvalues will be small (near zero).

Let $\lambda_1, \lambda_2, \ldots \lambda_p$ be the eigenvalues of the correlation matrix. The condition number of the correlation matrix is defined as:

$$K = \frac{\lambda_{max}}{\lambda_{min}} \text{ and } K_j = \frac{\lambda_{max}}{\lambda_j} \quad j = 1,2, \ldots P \qquad 6$$

Where λ_{max} is the largest eigenvalue.

λ_{min} is the smallest eigenvalue.

λ_j is the eigenvalue of the jth independent variable.

If the condition number is less than 100, there is no serious problem with multicollinearity and if a condition number is between 100 and 1000 implies a moderate to strong multicollinearity. Also, if the condition number exceeds 1000, severe multicollinearity exits [20].

2.4.4 Variance Inflation Factor

The Variance Inflation Factor (VIF) quantifies the severity of multicollinearity in an ordinary least squares regression analysis. Let R_j 2 denote the coefficient of determination when X_j is regressed on all other predictor variables in the model.

The VIF is given by:

$$VIF = \frac{1}{1 - R_j^2}; \; j = 1,2, \ldots p - 1 \qquad 7$$

The VIF provides an index to measure how much the variance of an estimated regression coefficient is increased because of the multicollinearity. The rule of thumb for interpreting the variance inflation factor states that; values 1 and below are not correlated, values from 1 to 5 are moderately correlated, and values greater than 5 are highly correlated (Kim, 2019) [14]. If any of the VIF values exceeds 10, it is an indication that the associated regression coefficients are poorly estimated because of multicollinearity [20].

2.5 Principal Component Analysis

Most often, when the multicollinearity assumption of the Ordinary Least Square (OLS) fails, principal component regression is a way out. Consider the Principal Component Analysis model below;

$$Z_j = P_{1j}X_1 + P_{2j}X_2 + \cdots + P_{pj}X_p \quad 8$$

Where;

Z is a data matrix (similar in structure to X) made up of the principal components.

P is a vector matrix of constants ($P_{1j}, P_{2j}, .., P_{pj}$) which contains an arbitrary scale factor on which we impose the condition that $P_jP_j = \sum_{j=1}^{p} P_j^2 = 1$.

Suppose that our multiple regression equation from equation (2.4) is written in matrix form below as;

$$\underline{Y} = X\underline{B} + \varepsilon$$

Where Y is the dependent variable, X represents the independent variables, $\underline{B}$ is the vector of regression coefficients to be estimated, and ε represents the errors or residuals. Rewriting the regression equation in terms of the principal component we have;

$$\underline{Y} = Z\underline{A} + \varepsilon \quad 9$$

Where Y is the dependent variable, Z represents the principal components, $\underline{A}$ is the loadings on the principal components, and ε represents the errors or residuals. Therefore, the two sets of regression coefficients $\underline{A}$ and $\underline{B}$ can be related using the formula below;

$$\underline{A} = P'\underline{B}. \quad 10$$

2.6 Use of Statistical Software

The Number Cruncher Statistical System (NCSS) version 12.0 was used to implement the principal component regression analysis for predicting yam yield using seven climatic determinants. Also, the data was analyzed using the Statistical Package for the Social Sciences (SPSS) Version 19.

3.0 RESULTS AND DISCUSSION

This chapter presents the results and discussion of the principal component regression for the nine months, eight months, and seven months growth phases of yam. The descriptive statistics, the correlation matrix of seven climatic determinants, the impact classification using eigenvalues, the regression coefficients classification, principal component models, and model adequacy for the three growth phases of yam yield were analyzed and discussed respectively.

3.1 Descriptive Statistics.

Table 1 presents the descriptive statistics of the three growth phases of yam which are used to present the average mean of the climatic determinants. The total number of counts for each climatic determinant in each of the three phases was 31 years. The mean log to base 10 values translates to the annual mean requirements for each of the climatic determinants for the three growth phases. Observe that in the first growth phase, seven climatic determinants had significant contributions to yam yield; in the second growth phase, six climatic determinants had significant contributions to yam yield; in the third growth phase, five climatic determinants had significant contributions to yam yield.

Table 1: Descriptive Statistics for three growth phases.

Climatic Determinant	Count	Mean	Standard Deviation	Minimum	Maximum

NINE MONTHS (1ST PHASE)					
Temp	31	1.44741	0.00660838	1.427234	1.45805
Sunshine	31	1.36744	0.01120443	1.328153	1.384015
Rainfall	31	2.10301	0.08552045	1.927427	2.254441
Relative H	31	1.84662	0.03106262	1.809934	1.949932
Radiation	31	1.14223	0.02230485	1.100371	1.179296
Evaporatn	31	0.63222	0.04031267	0.5426871	0.7238204
Soil Temp	31	1.46513	0.09682113	1.048355	1.771261
YAM	31	3.31749	0.2379430	2.78831	3.478776
EIGHT MONTHS (2ND PHASE)					
Temp.	31	1.448623	0.00706872	1.426613	1.45883
Rainfall	31	2.148287	0.08794251	1.958922	2.30559
Relative H	31	1.853474	0.03018037	1.820366	1.95000
Radiation	31	1.134702	0.02325693	1.087693	1.16916
Evaporatn	31	0.630294	0.05267583	0.5378191	0.73739
Soil Temp	31	1.469538	0.09346493	1.099508	1.79727
YAM	31	3.31749	0.237943	2.78831	3.47878
SEVEN MONTHS (3RD PHASE)					
Rainfall	31	2.153378	0.09075428	1.951823	2.320235
Relative H	31	1.850965	0.03209681	1.815767	1.950087
Radiation	31	1.13049	0.02384291	1.087376	1.162223
Evaporatn	31	0.6519077	0.05517150	0.551101	0.762357
Soil Temp	31	1.476961	0.09137684	1.157500	1.829579
YAM	31	3.31749	0.237943	2.78831	3.478776

3.2 Correlation Matrix of Climatic Determinant.

The correlation matrix of the climatic determinants' values and level of significance of correlation are shown in Table 2. The main diagonal showed that each climatic determinant perfectly correlates with itself with a 1.000 value. Hence, sunshine and temperature, sunshine and evaporation, and radiation and soil had high positive correlation values of 0.885, 0.465, and 0.492 respectively. Sunshine and temperature correlation were significant at 0.01 and 0.05 levels. Sunshine and evaporation, radiation, and soil correlation were significant at 0.05 level respectively. This indicates that increased sunshine increases temperature and evaporation. Also, an increase in radiation increases soil temperature respectively whereas rainfall and temperature, rainfall and sunshine, rainfall and evaporation, and rainfall and soil temperature had negative correlation values of -0.360, -0.296, -0.408, and -0.264 respectively. In a similar vein, relative humidity and temperature, relative humidity and radiation, relative humidity and evaporation, and relative humidity and soil temperature had values of -0.003, -0.275, -0.240, and -0.356 respectively, indicating an increase in any one paired determinant result in the decrease in the other determinant. Rainfall and temperature, rainfall and evaporation, relative humidity, and soil temperature correlation were not significant at 0.05 levels. It is important to note that the correlation coefficient between pairs of climatic determinants was obtained by correlating the log transform (base 10) of the data on each climatic determinant despite this, seven pairs of the climatic determinant still showed a significant correlation ($p < 0.05$) Hence the need for principal component regression.

Table 2: Correlation matrix of climate determinants

	TEMP.	SUNSHINE	RAINFALL	REL. HUM.	RAD.	EVAPORATN	SOIL TEMP.
TEMP.	1.000	.885**	-.360*	-.003	.167	.465**	.429*
	.	.000	.047	.985	.369	.008	.016
SUNSHINE	.885**	1.000	-.296	.126	.106	.228	.324
	.000	.	.106	.499	.570	.217	.075
RAINFALL	-.360*	-.296	1.000	.054	-.007	-.408*	-.26
	.047	.106	.	.773	.971	.023	.151
REL. HUM.	-.003	.126	.054	1.000	-.275	-.240	-.356*
	.985	.499	.773	.	.134	.193	.049
RAD.	.167	.106	-.007	-.275	1.000	.186	.492**
	.369	.570	.971	.134	.	.316	.005
EVAPORATN	.465**	.228	-.408*	-.240	.186	1.000	.294
	.008	.217	.023	.193	.316	.	.109
SOIL TEMP	.429*	.324	-.264	-.356*	.492**	.294	1.000
	.016	.075	.151	.049	.005	.109	.

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

3.3 Impact Classification using Eigenvalues.

The eigenvalues resulting from the correlation matrix of climatic determinants in Table 2 are given in Table 3 and Table 5. The seven eigenvalues presented in Table 3 are ordered from the highest to the lowest variability by incremental percent and condition percent of the climatic determinants in each growth phase, condition number greater than zero indicates a multicollinearity problem in data. The four principal components presented in Table 5 gives 81.11 %, 84.15 %, and 89.97 % of the total variability for the first, second, and third phase respectively. The main aim of principal component regression is to eliminate multicollinearity while keeping the maximum variability of the original climatic determinants; therefore, our data matrix is best described by four principal components for each phase. Observe also that each of these principal components has a condition number less than 100.

Table 3: Eigenvalues of Correlations for three growth phases

No. of PC	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1ST PHASE				
1	2.483233	35.47	35.47	1.00
2	1.260677	18.01	53.48	1.97
3	1.141986	16.31	69.8	2.17
4	0.791784	11.31	81.11	3.14
5	0.659022	9.41	90.52	3.77
6	0.60024	8.57	99.1	4.14
7	0.063056	0.9	100	39.38
2ND PHASE				
1	1.976513	32.94	32.94	1.00
2	1.178555	19.64	52.58	1.68
3	1.059462	17.66	70.24	1.87
4	0.834203	13.9	84.15	2.37
5	0.494961	8.25	92.39	3.99
6	0.456305	7.61	100	4.33
3RD PHASE				
1	1.609677	32.19	32.19	1.00
2	1.212062	24.24	56.43	1.33
3	0.925959	18.52	74.95	1.74
4	0.750625	15.01	89.97	2.14
5	0.501677	10.03	100	3.21

3.4 Discussion on the principal component regression for three growth phases.

The summary results of each determinant for the three different growth phases and their principal and least square components regression coefficients are presented in Table 4 displaying the change in the value of yam yield corresponding to the unit changes in each climatic determinant. The estimated values of the regression coefficients are a_0, a_1, a_2, a_3, and a_4 representing Intercept, Temperature, Sunshine, Rainfall, and Relative Humidity with coefficient values of 0.3536229, 0.7227507, 0.6267326, -0.0018192, and 3.186196 respectively for the first phase. a_0, a_1, a_3, a_4, and a_5 representing Intercept, Temperature, Rainfall, Relative Humidity, and Radiation with coefficient values 1.155066, -0.3678005, 0.09631784, 3.214812, and 1.769176 respectively for the second phase. a_0, a_3, a_4, a_5, and a_6 showing Intercept, Rainfall, Relative Humidity, Radiation, and Evaporation with coefficient values 4.213791, 0.1267436, 1.82786, -2.784382, and -0.8215092 respectively for the third phase.

The summary of the principal component regression models for the three different phases from Tables 4 and 8 showing the intercept and number of selected climatic determinants with corresponding coefficients for the nine-, eight- and seven-month growth phases of yam were presented as follows;

Model for nine months growth phase (1st PHASE):

$$Y_9 = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4$$

Y = 0.3536229 + 0.7227507 * (Temperature) + 0.6267326 * (Sunshine) -0.001819239 * (Rainfall) + 3.186196 * (Relative Humidity).

Temperature, sunshine, and relative humidity had a positive effect whereas rainfall harmed yam yield respectively.

Model for eight months growth phase (2nd PHASE):

$$Y_8 = a_0 + a_1X_1 + a_3X_3 + a_4X_4 + a_5X_5$$

Y = 1.155066 - 0.3678005 * (Temperature) + 0.09631784 * (Rainfall) + 3.214812 * (Relative Humidity) - 1.769176 * (Radiation).

Rainfall and relative humidity had a positive effect whereas Temperature and radiation harmed yam yield respectively.

Model for seven months growth phase (3rd PHASE):

$$Y_7 = a_0 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6$$

Y = 4.213791 + 0.1267436 * (Rainfall) + 1.82786 * (Relative Humidity) - 2.784382 * (Radiation) - 0.8215092 * (Evaporation).

Rainfall, Relative humidity had a positive effect whereas radiation and evaporation harmed yam yield respectively.
In each of the growth phases, the principal component regression models revealed the magnitude of the impact of each climatic determinant on the yield of yam. This impact is measured by the value of the regression coefficients of the climatic determinant in question. For the first phase, the yield of yam (y) increases by 0.7227507, 0.6267326, and 3.186196 units per unit change in temperature, sunshine, and relative humidity but reduces by 0.001819239 units per unit change in rainfall respectively.
For the second phase, the yield of yam (y) increases by 0.09631784 and 3.214812 units, per unit change in rainfall and relative humidity but reduces by 0.3678005, and 1.769176 units per unit change in temperature and radiation respectively.
For the third phase, the yield of yam (y) increases by 0.1267436 and 1.82786 units, per unit change in rainfall and relative humidity but reduces by 2.784382 and 0.8215092 units per unit change in radiation and evaporation respectively.

Table 4: Regression Coefficient for three growth phases

Climatic Determinant	Principal Component Coef. (*a*)	Regular Least Sq. Coef.(*b*)	Stand'zed Component Coef.	Stand'zed Least Sq. Coef.	Component Standard Error	Least Sq. Standard Error
1ST PHASE						
Intercept	0.3536229	-12.143				
Temperature	0.7227507	17.2883	0.02	0.48	2.314806	14.22467
Sunshine	0.6267326	-7.2185	0.03	-0.34	1.436745	8.219598
Rainfall	-0.0018192	0.40685	-0.001	0.146	0.2350165	0.42068
Relative H	3.186196	2.86644	0.416	0.374	0.8008138	1.088976
Radiation	-2.532375	-3.6879	-0.237	-0.346	0.9531322	1.602387
Evaporation	-1.814676	-2.6198	-0.307	-0.444	0.4994782	0.872027
Soil Temp	-0.5319019	0.01905	-0.216	0.008	0.1695462	0.357308
2ND PHASE						
Intercept	1.155066	-6.548337				
Temp	-0.3678005	6.425197	-0.011	0.191	2.762293	5.827228
Rainfall	0.09631784	0.456716	0.036	0.169	0.246286	0.458297
Relative H	3.214812	2.780651	0.408	0.353	1.010005	1.202464
Radiation	-1.769176	-4.094885	-0.173	-0.4	1.259313	1.616747
Evaporation	-1.078145	-1.670178	-0.239	-0.37	0.3870932	0.753968
Soil Temp	-0.5329729	0.0832215	-0.209	0.033	0.1409174	0.386341
3RD PHASE						
Intercept	4.213791	2.606745				
Rainfall	0.1267436	0.3615403	0.048	0.138	0.2809924	0.4366082
Relative H	1.82786	2.397618	0.247	0.323	0.6024051	1.138163
Radiation	-2.784382	-3.415912	-0.279	-0.342	0.9911239	1.579454
Evaporation	-0.8215092	-1.466064	-0.19	-0.34	0.3072584	0.6933909
Soil Temp	-0.5885525	0.21104	-0.226	0.081	0.1798179	0.4221013

3.5 Model adequacy.

The variance inflation factors (VIFs) and condition numbers from Table 6 and Table 7 were used to check for collinearity among climatic determinants for each growth phase. High VIFs reflect an increase in the variances of estimated regression coefficients due to collinearity among the climatic determinants. Numerical values of VIFs in Table 6 are all less than 1 shows that the climatic determinants are not correlated. Table 7 shows the condition numbers of corresponding eigenvalues for each principal component in each of the three phases. Observe that all the values are less than 100 indicating the absence of multi-collinearity. Condition number 100 above indicates the presence of multi-collinearity. Tables 8, 9, and 10 present the analysis of variance (ANOVA) results summarizing information related to the sources of variation for each of the growth phases.
The results of the study which employed Principal Component Regression to mitigate the effects of multi-collinearity of climatic data for use in the estimation and prediction of yam yield were obtained at the first phase with the highest eigenvalue of 2.483233,

Incremental percentage of 35.47, while the lowest obtained eigenvalue of 0.063056, Incremental percentage of 0.9 as shown in Table 5. F-Ratio of 3.0544 and Prob. level of 0.019993 as shown in Table 8.
Principal Components were obtained at the second phase with the highest eigenvalue of 1.976513, an Incremental percentage of 32.94, while the lowest obtained eigenvalue of 0.456305, an Incremental percentage of 7.61 as shown in Table 4.5. F-Ratio of 2.7215 and Prob. level of 0.03679 as shown in Table 9.
Lastly, Principal Components were also obtained in the third phase with the highest eigenvalue of 1.609677, Incremental percentage of 32.19, and Condition number of 1.00, while the lowest obtained eigenvalue of 0.501677, Incremental percentage of 10 as shown in Table 5. F-Ratio of 2.6529 and Prob. level of 0.046771 as shown in Table 10.
The models for each of the growth phases of yam presented in section 3.4 are significant ($p < 0.05$). Since the probability value of each of the models is less than the significance level.

Table 5: Eigenvalues Components Contributions

No.	Eigenvalue	Percent
1ST PHASE		
1	2.483233	35.47
2	1.260677	53.48
3	1.141986	69.8
4	0.791784	81.11
2ND PHASE		
1	1.976513	32.94
2	1.178555	52.58
3	1.059462	70.24
4	0.834203	84.15
3RD PHASE		
1	1.609677	32.19
2	1.212062	56.43
3	0.925959	74.95
4	0.750625	89.97

Table 6: Variance Inflation Factor for three growth phases

Climatic Determinant	Regression Coefficient	Standard Error	Standardized Regression Coefficient	VIF
1ST PHASE				
Intercept	0.3536229			
Temp	0.7227507	2.314806	0.02	0.183
Sunshine	0.6267326	1.436745	0.03	0.203
Rainfall	-0.001819239	0.2350165	-0.001	0.317
Relative H	3.186196	0.8008138	0.416	0.485
Radiation	-2.532375	0.9531322	-0.237	0.354
Evaporation	-1.814676	0.4994782	-0.307	0.318
Soil Temp	-0.5319019	0.1695462	-0.216	0.211
2ND PHASE				
Intercept	1.155066			
Temp	-0.3678005	2.762293	-0.011	0.272
Rainfall	0.09631784	0.246286	0.036	0.334
Relative H	3.214812	1.010005	0.408	0.662
Radiation	-1.769176	1.259313	-0.173	0.611
Evaporation	-1.078145	0.3870932	-0.239	0.296
Soil Temp	-0.5329729	0.1409174	-0.209	0.124
3RD PHASE				
Intercept	4.213791			
Rainfall	0.1267436	0.2809924	0.048	0.440
Relative H	1.82786	0.6024051	0.247	0.253
Radiation	-2.784382	0.9911239	-0.279	0.377
Evaporation	-0.8215092	0.3072584	-0.19	0.194
Soil Temp	-0.5885525	0.1798179	-0.226	0.182

Table 7: Condition Number of Components for three growth phases

No.	Eigenvalue	Condition Number
1ST PHASE		
1	2.483233	1
2	1.260677	1.97
3	1.141986	2.17
4	0.791784	3.14
5	0.659022	3.77
6	0.60024	4.14
7	0.063056	39.38
2ND PHASE		

1	1.976513	1
2	1.178555	1.68
3	1.059462	1.87
4	0.834203	2.37
5	0.494961	3.99
6	0.456305	4.33
3RD PHASE		
1	1.609677	1
2	1.212062	1.33
3	0.925959	1.74
4	0.750625	2.14
5	0.501677	3.21

All Condition Numbers less than 100 multi-collinearity is NOT a problem

Table 8: Analysis of Variance for the first growth phase model

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob. Level
1ST PHASE					
Intercept	1	341.1779	341.1779		
Model	7	0.8182642	0.1168949	3.0544	0.019993
Error	23	0.8802416	0.03827138		
Total (Adjusted)	30	1.698506	0.05661686		
Mean of Dependent	3.31749				
Root Mean Square Error	0.19563				
R-Squared	0.482				
Coefficient Variation	0.05897				

Table 9: Analysis of Variance for the second growth phase model

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob. Level
		2ND PHASE			
Intercept	1	341.1779	341.1779		
Model	6	0.6877112	0.1146185	2.7215	0.03679
Error	24	1.010795	0.04211644		
Total (Adjusted)	30	1.698506	0.05661686		
Mean of Dependent	3.31749				
Root Mean Square Error	0.2052229				
R-Squared	0.405				
Coefficient Variation	0.0618609				

Table 10: Analysis of Variance for the third growth phase model

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob. Level
		3RD PHASE			
Intercept	1	341.1779	341.1779		
Model	5	0.5887876	0.1177575	2.6529	0.046771
Error	25	1.109718	0.04438873		
Total (Adjusted)	30	1.698506	0.05661686		
Mean of Dependent	3.31749				
Root Mean Square Error	0.2106863				
R-Squared	0.347				
Coefficient Variation	0.06350776				

4.0 CONCLUSION

The results obtained by this study through the application of Principal Component Regression for utilization of climatic data for estimation and prediction of Yam yield is significantly improved as the effect of multi-collinearity among predictor variables is significantly mitigated through analysis of Yam yield for three growth phases.
For the first phase, the yield of yam (y) increases by 0.7227507, 0.6267326, and 3.186196 units per unit change in temperature, sunshine, and relative humidity but reduces by 0.001819239 units per unit change in rainfall respectively.
For the second phase, the yield of yam (y) increases by 0.09631784 and 3.214812 units, per unit change in rainfall and relative humidity but reduces by 0.3678005, and 1.769176 units per unit change in temperature and radiation respectively.
For the third phase, the yield of yam (y) increases by 0.1267436 and 1.82786 units, per unit change in rainfall and relative humidity but reduces by 2.784382 and 0.8215092 units per unit change in radiation and evaporation respectively.
This research work shows that the application of Principal Component Regression for mitigating the presence of multi-collinearity between the independent variable of climatic data can significantly be reduced to the best possible obtainable linear combination for independent components achieved for estimation and predictions of yam yield.

5.0 RECOMMENDATION

The following recommendations were made in the study:

i. The principal component regression approach should be employed in relating the yield of other crops in the phase of multicollinearity.
ii. The study should be extended to build models to assess the impact of climate change on crop yield. Principal component regression models of this nature should be integrated into such climate impact assessment models.

REFERENCES:

1. UNFCCC (1992). The UN Framework Convention on Climate Change (UNFCCC) held its Conference of the Parties (COP27) in Rio de Janeiro in June 1992. https://unfccc.int
2. IPCC (2007). Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland, 104.
3. Ayoade, J.O. (2004). Introduction to Climatology for the Tropics. 2nd Edition, Spectrum Books Limited, Ibadan.
4. Ruohong, C., Jeffrey, D.M., John, C.B., Shurley, W.D and Michael, E.W. (2011). Principal component analysis of crop yield response to Climate Change. Faculty Series 11-01. Retrieved from; ageconsearch.umn.edu/record/103947.
5. Jones, J.W., Hoogenboom, G., Porter C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens P.W., Singh, U., Gijsman, A.J., and Ritchie, J.T. (2003). The DSSAT Cropping System Model. *European Journal of Agronomy 18*(3-4), 235-265.
6. Walker, G.K. (1989). Model for operational forecasting of Western Canada wheat yield. *Agricultural and Forest Meteorology* 44, 339-351
7. Alibuhtto M.C., Peiris, T.S.G. (2015). Principal component regression for solving multicollinearity problem. 5th international symposium
8. Akpan, E.B. & Aye, G.C. (2016). Climate change impacts and adaptations: Does agricultural extension play a role in Jos-South Plateau State Nigeria? *International Journal of Resource Management and Engineering*, 6(12): 5-15
9. Koudahe, K., Koffi, D., Kayode, J.A., Awokola, S.O., and Adebola, A.A., (2018). Impact of climate variability on crop yields in southern Togo. Environmental Pollution and Climate Change, DOI:10.4172/2573-458X.1000148
10. Adedeji, I.A.,Tiku, N. E., Waziri-Ugwu, P. R., and Sanusi, S. O. (2017). The effect of climate change on rice production in Adamawa State, Nigeria. *Agroeconomia 7*(1), 1-13.
11. Uger, F. I (2017). Impact of climate variability on yam production in Benue state: an empirical analysis. *International Journal of innovative research in social sciences and strategic management 4*(2), 14-23
12. Elijah, S. T., Osuafor O. O., Anarah S. E. (2018). Effects of climate change on yam production in cross river state, Nigeria.*International Journal of Agriculture and Forestry 8*(2),104-111.
13. Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *J. Phys.: Conf. Ser.* 949 012009. https://doi.org/10.1088/1742-6596/949/1/012009
14. Kim, J. H (2019). Multicollinearity and misleading statistical results. *Korean J Anesthesiol*. 72(6), 558-569. https://doi.org/10.4097/kja.19087.
15. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 374(2065):20150202.
16. https://doi.org/10.1098/rsta.2015.0202
17. Ayinde, K., Lukman, A. F., Alabi, O. O., & Bello, H. A. (2020). A New Approach of Principal Component Regression Estimator with Applications to Collinear Data. *International Journal of Engineering Research and Technology*,13(7),1616-1622 https://dx.doi.org/10.37624/IJERT/13.7.2020.1616-1622
18. Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin, 83*(4), 579–586. https://doi.org/10.1037/0033-2909.83.4.579
19. Eguda (2006), Statistical Methods, Second Edition, Asofad Printing Limited, Makurdi-Nigeria.
20. Rencher (2002). Methods of multivariate analysis, second edition, a John Wiley and Sons, Inc. Publication.
21. Montgomery, D. C., Peck, E. A., Vining, G. G. (2001). Introduction to linear regression analysis, 3rd edition, Wiley, New York.