Effects of Climate Change on Maize Production in Ghana - A Comparative Study of Parametric and Non-Parametric Regression Models*

[1]A. Buabeng, [1]L. Brew, [1]B. Odoi and [2]F. Obiri-Yeboah
[1]University of Mines and Technology, Tarkwa
[2]Minerals Commission of Ghana, Tarkwa

Buabeng, A., Brew, L., Odoi, B. and Obiri-Yeboah, F. (2021), "Effects of Climate Change on Maize Production in Ghana - A Comparative Study of Parametric and Non-Parametric Regression Models", *Ghana Journal of Technology*, Vol. 6, No. 1, pp. 26 -35

Abstract

The devastating effects of climate change on the production of agricultural commodities have become a source of worry for many developing countries and therefore demand due attention. For these reasons, this paper sought to formulate models for analysing the effect of climate change on maize production in Ghana as there has been an alarming fluctuation in productivity across the country. Agroclimatic data such as wind speed, temperature, humidity, carbon dioxide and precipitate were obtained. First, a Multiple Regression Analysis (MRA) was performed using all the variables that resulted in high multicollinearity levels. Factor Analysis (FA) was employed to transform the dataset into a set of uncorrelated features to remedy the multicollinearity problem and perform a reliable analysis. Thus, the resulting features were used in developing two models based on parametric MRA and non-parametric Multivariate Adaptive Regression Splines (MARS). The results from the analysis indicate that the MARS model based on extracted features achieved a higher prediction accuracy of 76.59% when compared with the MRA's model (73.73%). Moreover, the MARS model produced the least Mean Absolute Percentage Error (MAPE) of 8.32% when compared to MRA's 12.12% during validation.

Keywords: Maize Productivity, Climate Change, Parametric, Non-Parametric, Regression

1 Introduction

One of the biggest challenges facing developing countries such as Ghana is providing food for its rapidly increasing population and therefore demands great attention. Among the staple food crops, maize (zee mays) is the most widely produced and consumed cereal crop in Ghana since 1965 and continues to play an indispensable role in ensuring the nations' food security (Morris *et al*., 1999; Darfour and Rosentrater, 2016). That is, maize has been enormously beneficial to the livelihoods of Ghanaians as it accounts for over 50% of the total cereal production in Ghana. Generally grown in the northern savannah, transitional, forest and coastal savannah zones, it occupies about one million hectares of land distributed all over the country. Out of the total volume of maize produced, 80% is sourced from major producing regions such as the Eastern, Ashanti and Brong Ahafo whiles the three northern regions supply the rest (Angelucci, 2019; Wongnaa *et al*., 2019). Though maize is cultivated by the vast majority of rural households in these producing regions, 85% of its total volume produced is primarily for human consumption while the remaining 15% is used in feeding live stocks (Andam *et al*., 2017).

Over the years, the production of maize has seen a significant increase. However, there has been an alarming fluctuation in its productivity across the country despite its economic benefits. Though these fluctuations could be attributed to the inadequate resources needed for increasing productivity (Darfour and Rosentrater, 2016; Wongnaa *et al*., 2019), irregularities in climate conditions are cited as the primary causes in the continuous reduction in the average yield of maize (Ji *et al*., 2012; Jones and Thornton, 2003; Mati, 2000; Tachie-Obeng *et al*., 2013; Wang *et al*., 2011). For these reasons, a considerable amount of resources has been dedicated to researching the effects of climate change on the productivity of agricultural commodities.

In literature, notable among the research works conducted for exploring the effect of climate change on various crops employed the techniques of parametric modelling (Asantewaa, 2003; De-Graft and Kweku, 2012; Ndamani *et al*., 2016). For instance, Asantewaa (2003) utilised the Error Correction Model and Granger Causality Test to examine the effects of climate on maize supply. De-Graft and Kweku (2012) examined the effects of climatic variables and crop area on the mean and variance of maize yield in Ghana using the Just and Pope stochastic production function based on the Cobb-Douglas functional form. Ndamani *et al*. (2016) in their attempt to determine the determinants of farmers' adaptation to climate change employed the logistic regression model and weighted average index in analysing their data. Concerning other crops, several kinds of research have used

**Manuscript received June 16, 2021*
Revised version accepted September 23, 2021

parametric methods in analysing the effect of climate on cocoa (Anim-Kwapong and Frimpong, 2008; Buabeng *et al*., 2019; Ogunsola and Oyekale, 2013), rice (Felkner *et al*., 2009; Matthews *et al*., 1997), root crops (Sagoe, 2006; Zakari *et al*., 2014). However, these parametric models assume several forms of assumptions (e.g., normality, homogeneity of variance, linearity, independence, stationarity etc.) which is not always achievable with most available data. Thus, limiting the potency of parametric models, and therefore the need for a non-parametric approach that assumes little or no assumption about the phenomenon under study. In view of this, this paper seeks to analyse the effect of climate change on maize production in Ghana using both parametric and non-parametric regression approaches. The paper will reveal new insights in the predictive modelling of agricultural commodities since they are essential frameworks in addressing food security issues.

2 Resources and Methods Used

2.1 Data Source and Description

In order to assess the effect of climate change on maize productivity, secondary data of Ghana's annual agroclimatic variables which span from 1980 to 2019 were acquired. Table 1 shows the agroclimatic variables as well as their source, unit and initials used throughout the paper for simplicity.

Table 1 Variables Considered for the Analysis

Variable	Initials	Unit	Source
Precipitation	PPT	mm	NASA
Specific Humidity at 2 Meters	QV2M	Kg	
Relative Humidity at 2 Meters	RH2M	%	
Temperature at 2 Meters (C)	T2M	℃	
Maximum Temperature at 2 Meters	T2MX	℃	
Minimum Temperature at 2 Meters	T2MN	℃	
Wind Speed at 10 Meters	WS10M	m/s	
Range of Wind Speed at 10 Meters	WS10MR	m/s	
Wind Speed at 2 Meters	WS2M	m/s	
Range of Wind Speed at 2 Meters	WS2MR	m/s	
Wind Speed at 50 Meters	WS50M	m/s	
Range of Wind Speed at 50 Meters	WS50MR	m/s	
Carbon Dioxide Concentration	CO2	g/MJ	World Bank
Maize Production	PROD	Mt	MOFA

2.2 Methods Used

One primary concern when analysis multivariate a multivariate dataset is the occurrence of multicollinearity. The presence of multicollinearity when ignored does not only affect the predictive ability of a model but also affects the estimation of the model parameters and their statistical significance tests (Stamatis, 2016). This subsequently may result in wrong interpretation when making an inference. In literature, among the popular method of addressing the issue of multicollinearity is to replace the independent variables with a fewer number of uncorrelated factors via Factor Analysis (Hoerl and Kennard, 1970).

2.2.1 Factor Analysis

The factor model can be seen as a series of multiple regressions, predicting each of the observable variables X_i from the values of the unobservable common factors f_i as shown in Equation (1)

$$\begin{aligned} X_1 &= \mu_1 + l_{11} f_1 + l_{12} f_2 + \ldots + l_{lm} f_m + \varepsilon_1 \\ X_2 &= \mu_2 + l_{21} f_1 + l_{22} f_2 + \ldots + l_{2m} f_m + \varepsilon_2 \\ &\vdots \qquad\qquad\qquad \vdots \\ X_p &= \mu_p + l_{p1} f_1 + l_{p2} f_2 + \ldots + l_{pm} f_m + \varepsilon_p \end{aligned} \tag{1}$$

The variables mean μ_1 through μ_p can be viewed as the intercept terms for the multiple regression models. The regression coefficients l_{ij} (the partial slopes) for all these multiple regressions are called factor loadings where l_{ij} is the loading of the i^{th} variable on the j^{th} factor. Finally, the errors ε_i are called the specific factors for variable *i*. The basic model is comparable to a regression model where each of the predictors *X* is to be estimated as a linear function of the unobserved common factors $f_{1,} f_2 ..., f_m$. Thus, the explanatory variables are $f_{1,} f_2 ..., f_m$. Therefore, it is assumed that *m* unobserved factors control the variation among the dataset. Generally, Equation (1) is reduced into a matrix notation as shown in Equation (2).

$$\mathrm{X} = \mu + \mathrm{Lf} + \in \tag{2}$$

However, before the Factor procedure is conducted, the suitability of the dataset is tested using the Kaiser-Meyer-Olkin (KMO) test for determining the sampling adequacy (Kaiser, 1974). Also, the Bartlett's sphericity test is carried out for testing the hypothesis that the correlation matrix is an identity matrix, which indicates that a relationship does not exist among the items (Bartlett, 1950). For the KMO, a minimum value of 0.5 is the acceptable threshold to proceed with the factor analysis whiles the Bartlett's test (p < 0.05) to be considered appropriate (Hair *et al*., 2009). For consistency, the number of factors that are needed for retention is decided based on Kaiser's and the scree plot criterion. The Kaiser criterion proposes the retention of all of the factors that are above the eigenvalue of 1. To ensure a well-distributed factor pattern, a Factor Analysis with a varimax rotation is adopted.

2.2.2 Multiple Regression Analysis

Multiple regression analysis is a general parametric statistical technique used to analyse the relationship between a single dependent variable and several independent variables. In the regression model, the dependent variable Y_i is expressed as a linear function of the independent variables $x_i's$ and a random error as shown in Equation (3).

$$Y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \ldots + \beta_q x_{qj} + \varepsilon_j \qquad (3)$$

where the $\beta's$ are the regression coefficients. Due to the parametric nature of the methods' estimation, the basic assumptions accompanying Equation (3), thus, must be achieved to ensure a model's adequacy. The assumptions are as follows:

i. Assumption 1: $E(\varepsilon) = 0$

ii. Assumption 2: $\text{var}(\varepsilon) = \sigma^2$ for all $i = 1, 2..., n$

iii. Assumption 3: $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for all $i = j$

iv. Assumption 4: $\varepsilon_i \sim N(0,1)$

Assumption 1 emphasises the need for the residuals of the model to be linear and signifies that no additional terms are needed to predicts. Assumption 2 requires that the variance of the residuals ε_i be the same or independent of each other. Assumption 3 imposes the condition that the terms be uncorrelated. Assumption 4 enforces the condition that the error terms be normally distributed.

2.2.3 Multivariate Adaptive Regression Splines (MARS)

The MARS is a nonparametric regression technique that works by dividing the variables into regions, producing each region as a least-squares equation (Friedman, 1991). Unlike the classical regression models, MARS assumes no functional relationship between the dependent and the independent variables. Instead, MARS makes this relation from the set of coefficients called basis functions that are exclusively determined from the regression data. MARS can be seen as a stepwise linear regression to improve the performance of a given regression set. MARS creates a separate regression equation for the individual linear region in the model. Each obtained linear region is called a "knot", which highlights MARS as an applicable solution to multivariate regression problems that might cause multidimensionality for other techniques. MARS model estimates predictive variables whose effect on a single predicted variable is being examined in the model as in Equation (4).

$$Y = \beta + \sum_{k=1}^{M} \beta_k h_k(X) \qquad (4)$$

where Y is the independent variable, X is the predictors and β_k is the k^{th} Basis Function (BF) in every linear knot. The estimation of the MARS model is developed in two steps. Initially (the forward step), MARS is estimated with an excessive number of knots in order to get a better estimate of the predictors (Samui and Kim, 2013). Then, the knots that contribute significantly to the overall estimation are retained whiles eliminating the less significant ones (the backward step).

To ensure the goodness of fit, the Generalized Cross-Validation (GCV) (Equation 5) is adopted in removing redundant basis functions (Samui and Kothari, 2012; Yakubu *et al*., 2018).

$$GCV = \frac{1}{N} \frac{\sum_{i}^{N} \left[Y_i - f_M(X_i) \right]^2}{\left[1 - \frac{C(M)}{N} \right]^2} \qquad (5)$$

where N is the number of instances, M denotes the number of BFs in the final model, $C(M) = (h+1) + dH$ is a function for defining complexity which increases M, and d is the penalty for each BF included in Equation (4).

3 Results and Discussion

Table 2 shows the descriptive statistics of the variables considered in the study. As observed, the averages, the deviations from the mean, as well as the kurtosis and the skewness value for each of the variables is shown.

3.1 Preliminary Analysis

First, all the 13 variables were regressed on the production of Maize (PROD). The results are presented in Table 3. As observed, the regression model had two major deficiencies:

(i). About 92% (12 out of 13) of the variables are not significant at (P-Value > 0.05).

(ii). The Variance Inflation Factor (VIF) values of most of the parameters (11 out of 13) were greater than 10, which is an indication of multicollinearity problem in multiple regression analysis involving all the variables (Hair *et al*., 2009).

Henceforth, an application of a direct multiple regression produced inaccurate interpretations (spurious regression). In order to solve the multicollinearity problem and perform a reliable regression analysis, factor analysis is therefore employed to help eliminate or reduce the level of multicollinearity.

3.2 Component Factor Analysis

Preliminary inspection of the sphericity value for the Bartlett's test is 1215.4620 with a significance < 0.0001. Moreover, the overall Measure of Sampling Adequacy (MSA) value of (0.5830 > 0.5) is acceptable (Kaiser, 1974), indicating that the correlation matrix is not an identity matrix and the absence of any correlations between the variables collectively, these tests suggested that dataset is appropriate for factor analysis.

Table 4 contains the eigenvalues and percentage of variance explained regarding the 13 possible factors. Three factors with eigenvalues greater than 1.0 were produced, which is the normal cut-off criterion (latent root criterion) for the determination and extraction of the number of factors. The factors account for 93.15 % of the total explained variance, and the variance that needed explaining is more than 60 % to satisfy the sample adequacy (Brown, 2009; Ye *et al*., 2015). The scree and variance explained plots in Figs. 1 and 2 support the result of the eigenvalue criterion. The scree plot shows the number of extracted factors which indicates a distinct change in gradient in the slope. Hence, three factors are retained for further analysis.

After the extraction technique, the varimax rotation is performed besides the unrotated factor solution in order to improve interpretation since the unrotated factor matrix did not have a completely clean set of factor loadings pattern. Table 5 shows the results factor loadings after the implementation of both the unrotated and varimax-rotated solutions. The results bolden only the highest loadings regarding each factor, while loadings that are less than absolute 0.65 are not considered negligible (Hair *et al*., 2009). In this paper, the varimax rotation technique provides a more favourable result, as it reflects the patterns properly compared with other rotation techniques.

Table 2 Statistical Description

Variable	Mean	Standard Deviation	Kurtosis	Skewness
QV2M	0.0163	0.0009	-0.1744	-0.2021
WS10M	2.8027	0.1017	-0.2861	0.5085
WS10MR	2.8210	0.1164	-0.2470	0.3880
WS2M	1.9094	0.0772	-0.3394	0.4413
WS2MR	2.2920	0.0879	-0.0001	0.4755
WS50M	4.1599	0.1368	-0.2249	0.5653
WS50MR	3.7176	0.1478	-0.1250	0.3800
CO2	0.3626	0.1174	0.4743	1.0236
RH2M	70.2022	2.5670	-0.1045	-0.5553
T2M	26.6261	0.4371	0.9874	-0.3823
T2MX	31.7775	0.6155	0.6092	0.0550
T2MN	22.3399	0.3846	0.0347	-0.5238
PPT	1170.6708	307.6102	9.6626	-2.9244
PROD	1143.6410	534.6135	-0.8980	0.1872

Table 3 Preliminary Results

Variable	Coefficient	Standard Deviation	t	P-value	VIF
Intercept	45286.1201	129495.1231	0.3497	0.7295	
PPT	20.1017	96.3268	0.2087	0.8364	2.1813
QV2M	6787.3082	8954.4866	0.7580	0.4556	38.8870
RH2M	-18775.3265	32205.2028	-0.5830	0.5651	199.0141
T2M	-217478.8043	208509.2622	-1.0430	0.3069	1657.6640
T2MX	117283.0769	90437.8968	1.2968	0.2065	433.5271
T2MN	109408.0656	108565.3880	1.0078	0.3232	494.5028
WS10M	80206.3462	103119.6653	0.7778	0.4440	1942.7500
WS10R	-53209.1120	45459.0275	-1.1705	0.2528	488.9574
WS2M	-42898.9591	57697.5868	-0.7435	0.4641	755.9903
WS2MR	44524.4348	44629.7343	0.9976	0.3280	406.1674
WS50M	-41178.8572	52883.9935	-0.7787	0.4435	418.8882
WS50MR	-1071.6350	14230.6810	-0.0753	0.9406	44.4763
CO2	2209.4601	536.9657	4.1147	0.0004	3.7473

F-value= 14.5562 P-Value= <0.0001 R^2(Adjusted)= 0.8224

Table 3 Eigenvalues of the Reduced Correlation Matrix

Sn	Eigenvalue	Difference	Proportion (%)	Histogram	Cumulative (%)
1	7.3820	4.4265	60.47		60.47
2	2.9554	1.9209	24.21		84.67
3	1.0345	0.6244	8.47		93.15
4	0.4100	0.1693	3.36		96.51
5	0.2406	0.1437	1.97		98.48
6	0.0969	0.0385	0.79		99.27
7	0.0584	0.0247	0.48		99.75
8	0.0338	0.0298	0.28		100.03
9	0.0039	0.0035	0.03		100.06
10	0.0003	0.0009	0.00		100.06
11	-0.0006	0.0015	-0.01		100.06
12	-0.0022	0.0022	-0.02		100.04
13	-0.0044	-	-0.04		100.00
Total	**12.20847**	**-**	**-**	**-**	**-**

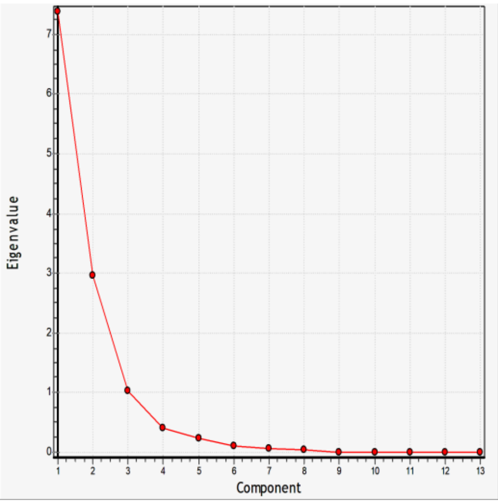

Fig. 1 Scree Plot

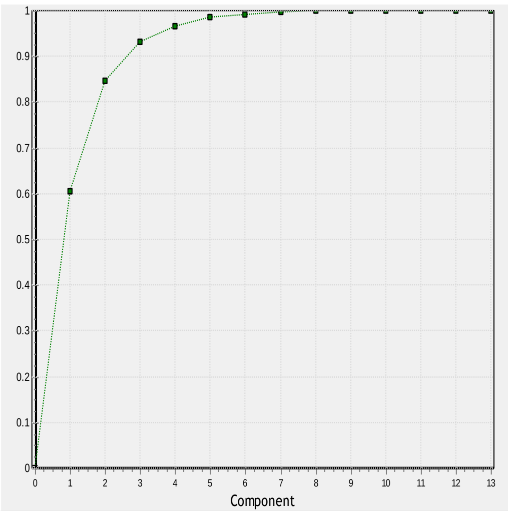

Fig. 2 Variance Explained Plot

3.2.1 Naming the Factors

Since a satisfactory factor solution has been derived, the next attempt is to assign some meaning to the factors. Variables with higher loadings influence to a greater extent the name or label selected to represent a factor. From Table 5 (varimax-rotated factor pattern), each factor is named based on the variables with significant loadings:

(i). Factor 1 consists of eight items that focus primarily on the wind speed and humidity related variables. This component accounts for 53 % of the total variance that was explained among all of the critical components.

(ii). Factor 2 consist of three items that focus primarily on temperature-related variables. This component accounts for 26% of the total variance.

(iii). Factor 3 explains 16% of the total variance and includes two variables; precipitation and carbon dioxide concentration.

The objective of employing factor analysis is to identify and estimate appropriate variables (factor scores) to be used as independent variables for the Multiple Regression Analysis and the Multivariate Adaptive Regression Splines. Factor scores are estimated by multiplying the standardised matrix form of the dataset of the 13 variables with their respective standardised scoring coefficients matrix. The i^{th} factor score of the climate variables is estimated using Equations (6), (7) and (8).

Table 5 Unrotated and Varimax-Rotated Factor Loadings

Variable	Unrotated Factor Loadings			Varimax Rotated Factor Loadings		
	Factor 1	**Factor 2**	**Factor 3**	**Factor 1**	**Factor 2**	**Factor 3**
WS10M	0.93	-0.13	0.25	**0.97**	0.07	0.07
WS50M	0.91	-0.14	0.30	**0.97**	0.04	0.01
WS2M	0.95	-0.11	0.20	**0.96**	0.11	0.10
WS2MR	0.92	-0.29	0.07	**0.92**	-0.02	0.28
QV2M	-0.90	0.04	-0.16	**-0.89**	-0.17	-0.09
WS10MR	0.93	-0.26	-0.06	**0.88**	0.06	0.38
RH2M	-0.92	-0.30	0.07	**-0.77**	-0.56	-0.17
WS50MR	0.85	-0.04	-0.45	**0.64**	0.36	0.62
T2M	0.41	0.89	-0.17	0.17	**0.98**	-0.11
T2MN	-0.01	0.96	-0.10	-0.21	**0.89**	-0.30
T2MX	0.69	0.70	-0.17	0.46	**0.88**	0.04
CO2	-0.19	0.61	0.49	-0.12	0.33	**-0.72**
PPT	0.23	-0.33	-0.55	0.09	-0.05	**0.67**
Variance Explained	**7.38 (60%)**	**2.96 (24%)**	**1.03 (8%)**	**6.53 (53%)**	**3.12 (26%)**	**1.72 (16%)**

$$\begin{aligned} FS1 = & \left(\frac{PPT_i-\mu_{PPT}}{\sigma_{PPT}}\right)*-0.314+\left(\frac{QV2M_i-\mu_{QV2M}}{\sigma_{QV2M}}\right)*-0.231+\left(\frac{RH2M_i-\mu_{RH2M}}{\sigma_{RH2M}}\right)*0.081+\left(\frac{T2M_i-\mu_{T2M}}{\sigma_{T2M}}\right)*0.982 \\ & +\left(\frac{T2MX_i-\mu_{T2MX}}{\sigma_{T2MX}}\right)*-0.427+\left(\frac{T2MN_i-\mu_{T2MN}}{\sigma_{T2MN}}\right)*-0.655+\left(\frac{WS10M_i-\mu_{WS10M}}{\sigma_{WS10M}}\right)*-0.392 \\ & +\left(\frac{WS10MR_i-\mu_{WS10MR}}{\sigma_{WS10MR}}\right)*0.497+\left(\frac{WS2M_i-\mu_{WS2M}}{\sigma_{WS2M}}\right)*0.428+\left(\frac{WS2MR_i-\mu_{WS2MR}}{\sigma_{WS2MR}}\right)*-0.145 \\ & +\left(\frac{WS50M_i-\mu_{WS50M}}{\sigma_{WS50M}}\right)*0.605+\left(\frac{WS50MR_i-\mu_{WS50MR}}{\sigma_{WS50MR}}\right)*-0.263+\left(\frac{CO2_i-\mu_{CO2}}{\sigma_{CO2}}\right)*0.064 \end{aligned} \quad (6)$$

$$\begin{aligned} FS2 = & \left(\frac{PPT_i-\mu_{PPT}}{\sigma_{PPT}}\right)*0.082+\left(\frac{QV2M_i-\mu_{QV2M}}{\sigma_{QV2M}}\right)*-0.101+\left(\frac{RH2M_i-\mu_{RH2M}}{\sigma_{RH2M}}\right)*0.156+\left(\frac{T2M_i-\mu_{T2M}}{\sigma_{T2M}}\right)*1.022 \\ & +\left(\frac{T2MX_i-\mu_{T2MX}}{\sigma_{T2MX}}\right)*0.133+\left(\frac{T2MN_i-\mu_{T2MN}}{\sigma_{T2MN}}\right)*-0.076+\left(\frac{WS10M_i-\mu_{WS10M}}{\sigma_{WS10M}}\right)*-1.207 \\ & +\left(\frac{WS10MR_i-\mu_{WS10MR}}{\sigma_{WS10MR}}\right)*0.160+\left(\frac{WS2M_i-\mu_{WS2M}}{\sigma_{WS2M}}\right)*0.533+\left(\frac{WS2MR_i-\mu_{WS2MR}}{\sigma_{WS2MR}}\right)*-0.213 \\ & +\left(\frac{WS50M_i-\mu_{WS50M}}{\sigma_{WS50M}}\right)*0.448+\left(\frac{WS50MR_i-\mu_{WS50MR}}{\sigma_{WS50MR}}\right)*0.084+\left(\frac{CO2_i-\mu_{CO2}}{\sigma_{CO2}}\right)*-0.046 \end{aligned} \quad (7)$$

$$\begin{aligned} FS3 = & \left(\frac{PPT_i-\mu_{PPT}}{\sigma_{PPT}}\right)*0.131+\left(\frac{QV2M_i-\mu_{QV2M}}{\sigma_{QV2M}}\right)*1.012+\left(\frac{RH2M_i-\mu_{RH2M}}{\sigma_{RH2M}}\right)*-2.011+\left(\frac{T2M-\mu_{T2M}}{\sigma_{T2M}}\right)*-5.873 \\ & +\left(\frac{T2MX_i-\mu_{T2MX}}{\sigma_{T2MX}}\right)*1.818+\left(\frac{T2MN_i-\mu_{T2MN}}{\sigma_{T2MN}}\right)*3.255+\left(\frac{WS10M_i-\mu_{WS10M}}{\sigma_{WS10M}}\right)*5.630 \\ & +\left(\frac{WS10MR_i-\mu_{WS10MR}}{\sigma_{WS10MR}}\right)*-1.429+\left(\frac{WS2M_i-\mu_{WS2M}}{\sigma_{WS2M}}\right)*-2.968+\left(\frac{WS2MR_i-\mu_{WS2MR}}{\sigma_{WS2MR}}\right)*1.106 \\ & +\left(\frac{WS50M_i-\mu_{WS50M}}{\sigma_{WS50M}}\right)*-3.224+\left(\frac{WS50MR_i-\mu_{WS50MR}}{\sigma_{WS50MR}}\right)*1.467+\left(\frac{CO2_i-\mu_{CO2}}{\sigma_{CO2}}\right)*-0.178 \end{aligned} \quad (8)$$

3.3 Development of Parametric and Non-Parametric Models based on Factor Scores

Based on the factor scores generated using Equations (6), (7) and (8), two models (Parametric Multiple Linear Regression and Non-Parametric Multivariate Adaptive Regression Splines) will be fitted to the factor scores data and evaluated to determine the "best" model. The most suitable model is selected based on the magnitude of the adjusted R^2 value (Coefficient of determination), Root Meat Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

3.3.1 Summary of Multiple Regression Analysis (MRA) based on Factor Scores

Multiple regression is applied with the three factors assigned as independent variables, and maize production as the dependent variable. The result is shown in Table 6. As observed, the p-value from the F-test (< 0.0001) shows that the model is statistically significant. The adjusted R-squared value of 73.73% shows the variability of maize production that is accounted for by the model. The regression coefficients of the factors; FS1, 2 and 3 are all highly significant (p-values <0.05). Moreover, the VIF values for all the independent variables is approximately 1, which indicates the absence of multicollinearity.

Regarding the test of assumptions on the parametric MRA model, Table 7 shows the test on the assumptions regarding the regression model in Table 6. As observed, all the assumptions regarding MRA are satisfied. These tests suggest that the MRA model based on factor scores fits very well, hence, is adequate for prediction.

Thus, the Multiple Regression Models based on factor scores is expressed as Equation (9).

$$\begin{aligned} Y &= 1143.641 - 166.6312 * FS1 \\ &\quad + 208.1543 * FS2 - 382.5133 * FS3 \end{aligned} \tag{9}$$

Inference from Equation (9) suggests that FS1 (comprising of wind speed and humidity related variables) and FS3 (precipitation) negatively affects maize production by 166.6312 and 382.5133 Mt respectively. This is because extreme winds can affect maize productivity by toppling plants without firm root systems. Also, prolonged and high humidity levels tend to promote plant rot due to the reduced air circulation needed in the transpiration process (i.e., affects water evaporation and the drawing of nutrients from the soil). Moreover, humid conditions promote the growth of mould and bacteria that cause plants to die as well as the invite of pests whose larva feed on plant roots and thrive in moist soil. In the case of precipitate, although is essential for the smooth growth of maize plants, excessive amounts can injure the plants (root loss), leach vital nitrogen (for photosynthesis) out of the soil compact the soil and may lead to erosion.

On the other hand, Equation (9) suggests that FS2 (comprising of temperature-related variables) contributes positively (208.1543 Mt) to maize production since ambient temperature influences all plant growth processes including photosynthesis, respiration, transpiration, breaking of seed dormancy, and seed germination.

3.3.2 Summary of Multivariate Adaptive Regression Splines (MARS) Model based on Factor Scores

Table 8 shows the result of the MARS model expressed by four Basis Functions in terms of variable structure and the effect of the factors on maize production. The model accounts for 76.59% of the total variation in maize production. Also, all the factors are significant at 95% confidence level. Hence, the Multivariate Adaptive Regression Splines model based on factor scores is expressed as Equation (10).

$$\begin{aligned} Y &= 1479.1847 - 195.5572 * BF1 \\ &\quad + 275.7074 * BF2 - 54.2087 * BF3 \\ &\quad - 70.1248 * BF4 \end{aligned} \tag{10}$$

where the Basis Functions (BF's) in Table 8 are expressed as follows:

BF1 = max(0, FS3 + 3.26047)
BF2 = max(0, FS2 + 2.55013)
BF3 = max(0, FS1 + 1.71568) * BF1
BF4 = max(0, FS3 + 0.348689) * BF2

3.3.2 Validation of MRA and MARS Models

Table 9 compares the validations of the two models; the Multiple Regression Analysis (MRA) model and Multivariate Adaptive Regression Splines (MARS) model. This is achieved by evaluating the performance of models in predicting production values from 2015 to 2019. The predictions made by these two models are presented in Table 9. As observed, the MARS produced the least error (8.32%) when compared to the MRA's 12.29%. Fig. 3 shows the predictions as well as the errors of each year for both MRA and MARS models.

Table 6 Summary of Multiple Regression Models based on Factor Scores

Parameter	Coefficient	Std. Error	t	P-Value	VIF
Constant	1143.6410	43.8797	26.0631	< 0.0001	
FS1	-166.6312	44.0051	-3.7866	0.0006	1.0004
FS2	208.1543	43.9116	4.7403	<0.00001	1.0002
FS3	-382.5133	45.5865	-8.3909	<0.00001	1.0006

F-Value= 36.5447 P-Value= <0.0001 R^2(Adjusted)= 0.7373

Table 7 Test of Assumptions on MRA Model based on Factor Scores

Assumption	Hypothesis	Statistic (P-Value)
Linearity (Runs)	H_0: $E(\varepsilon_i)=0$ H_1: $E(\varepsilon_i)\neq 0$	0.0000
Normality (Shapiro-Wilk)	H_0: Normally distributed H_1: Not normally distributed	0.9818 (0.7677)
Serial Correlation (Box-Pierce)	H_0: True autocorrelation=0 H_1: True autocorrelation $\neq 0$	3.2773 (0.0702)
Homoscedasticity (Breusch-Pagan)	H_0: Constant variance H_1: Non-constant variance	3.3652 (0.3387)

Table 8 Summary of MARS Model based on Factor Scores

Parameter	Coefficient	Std. Error	t	P-Value
Constant	1479.1847	269.4108	5.4904	<0.0001
Basis Function (BS1) 1	-195.5572	71.5270	-2.7340	0.0099
Basis Function (BS2) 2	275.7074	51.3431	5.3699	<0.0001
Basis Function (BS3) 3	-54.2087	12.3633	-4.3847	0.0001
Basis Function (BS4) 4	-70.1248	32.0191	-2.1901	0.0355

F-Value= 32.0870 P-Value= <0.0001 R^2(Adjusted)= 0.7659

Table 9 Cross-Validation of MRA and MARS Models

Year	Observed	Predictions	
		MRA	MARS
2015	1692	1299.57	1394.78
2016	1722	1785.22	1838.02
2017	1990	1752.64	1886.67
2018	2200	2283.44	2060.95
2019	2000	2377.44	2115.24
MAPE		**12.29%**	**8.32%**

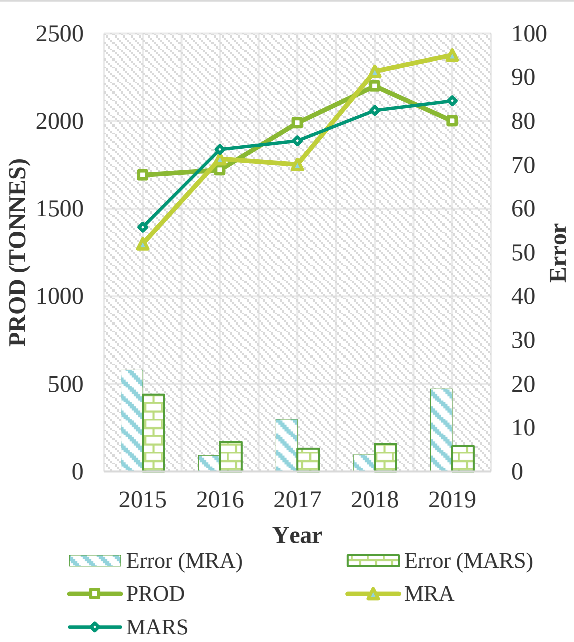

Fig. 3 Validation of Models

4 Conclusion

In this paper, models based on the comparative analysis of parametric Multiple Regression Analysis (MRA) and non-parametric Multivariate Adaptive Regression Splines (MARS) have been developed for predicting maize production in Ghana with cognisance of the devastating effect of climate change on productivity. The results from the factor analysis indicated that three factors (accounting for 93.15% of the total variance in the dataset) were adequate for the model development. The MARS model achieved a higher prediction accuracy of 76.59% as well as the least Mean Absolute Percentage Error (MAPE) of 8.32%. The MRA's model on the other hand achieved 73.73% prediction accuracy with 12.12% MAPE. Deductions from the models indicated that FS1 (comprising of wind speed and humidity related variables) and FS3 (precipitation) negatively affected maize production (166.6312 and 382.5133 Mt respectively) whiles FS2 (comprising of temperature-related variables) contributed positively (208.1543 Mt).

References

Andam, K., Johnson, M., Ragasa, C., Kufoalor, D. and Gupta, D. S. (2017), "A Chicken and Maize Situation the Poultry Feed Sector in Ghana", *Scientia Agricola*, Vol. 73, No. 3, pp. 201-208.

Angelucci, F. (2019), "Analysis of Incentives and Disincentives for Sorghum in Ghana", *Technical Notes Series*, MAFAP, FAO, Rome, 30 pp.

Anim-Kwapong, G. J. and Frimpong, E. B. (2008), "Vulnerability of Agriculture to Climate Change-Impact of Climate Change on Cocoa Production", *Published Annual ICCO Report*, Cocoa Research Institute of Ghana, New Tafo Akim, 44 pp.

Asantewaa, A. (2003), "Effect of Climate Change on Maize Supply in Ghana 1970-2002", *Submitted Master's Thesis*, University of Ghana, Accra, Ghana, 101 pp.

Bartlett, M. S. (1950), "Tests of Significance in Factor Analysis", *British Journal of Psychology*, Vol. 3, pp. 77–85.

Brown, J. D. (2009), "Choosing the Right Type of Rotation in PCA and EFA", *JALT Testing and Evaluation SIG Newsletter*, Vol. 13, No. 3, pp. 20-25.

Buabeng, A., Twumasi-Ankrah, S., Nyantakyi, K. A. and Kumi-Boateng, B. (2019), "Parsimonious Model for Predicting Cocoa Production in Ghana", *Advances and Applications in Statistics*, Vol. 55, No. 2, pp. 198- 220.

Darfour, B. and Rosentrater, K. (2016), "Maize in Ghana: An Overview of Cultivation to Processing", *ASABE Annual International Meeting Presentation Paper*, Orlando, Florida, 16 pp.

De-Graft, A. H. and Kweku, K. C. (2012), "The Effects of Climatic Variables and Crop Area on Maize Yield and Variability in Ghana", *Russian Journal of Agricultural and Socio-Economic Sciences*, Vol. 10, No. 10, pp. 10-13.

Felkner, J., Tazhibayeva, K. and Townsend, R. (2009), "Impact of Climate Change on Rice Production in Thailand", *American Economic Review*, Vol. 99, No. 2, pp. 205-10.

Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines", *The Annals of Statistics*, Vol. 19, No. 1, pp. 1-67.

Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. (2009), *Multivariate Data Analysis*, 7th edition, Pearson, New Jersey, 816 pp.

Hoerl, A. E. and Kennard, R. W. (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems", *Technometrics, Taylor and Francis*, Vol. 12, No. 1, pp. 55-67.

Ji, R. P., Zhang, Y. S., Jiang, L. X., Zhang, S., Feng, R., Chen, P. and Wu, J. (2012), "Effect of Climate Change on Maize Production in Northeast China", *Geographical Research*, Vol. 31, No. 2, pp. 290-298.

Jones, P. G. and Thornton, P. K. (2003), "The Potential Impacts of Climate Change on Maize Production in Africa and Latin America in 2055", *Global Environmental Change, Elsevier*, Vol. 13, No. 1, pp. 51-59.

Kaiser, M. O. (1974), "Kaiser-Meyer-Olkin Measure for Identity Correlation Matrix", *Journal of the Royal Statistical Society*, Vol. 52, pp. 296-298.

Mati, B. M. (2000), "The Influence of Climate Change on Maize Production in the Semi-Humid–Semi-Arid Areas of Kenya", *Journal of Arid Environments, Elsevier*, Vol. 46, No. 4, pp. 333-344.

Matthews, R. B., Kropff, M. J., Horie, T. and Bachelet, D. (1997), "Simulating the Impact of Climate Change on Rice Production in Asia and Evaluating Options for Adaptation", *Agricultural Systems*, Vol. 54, No. 3, pp. 399-425.

Morris, M. L., Tripp, R. B. and Dankyi, A. A. (1999), "Adoption and Impacts of Improved Maize Production Technology: A Case Study of the Ghana Grains Development Project", *Published Economics Program Paper 99-01*, Mexico, D. F., CIMMYT, 46 pp.

Ndamani, F., Watanabe, T., Ndamani, F. and Watanabe, T. (2016), "Determinants of Farmers' Adaptation to Climate Change: A Micro Level Analysis in Ghana", *Scientia Agricola*, Vol. 73, No. 3, pp. 201-208.

Ogunsola, G. O. and Oyekale, A. S. (2013), "Determinants of smallholder Cocoa Farmers' Adaptation to Climate Change in Ile-oluji/Okeigbo Local Government Area of Ondo State, Nigeria", *Asia Life Sciences*, pp. 1-10.

Sagoe, R. (2006), "Climate Change and Root Crop Production in Ghana", *Published Report*, Environmental Protection Agency (EPA), Accra, Ghana, 37 pp.

Samui, P. and Kim, D. (2013), "Least Square Support Vector Machine and Multivariate Adaptive Regression Spline for Modeling Lateral Load Capacity of Piles", *Neural Computing and Applications, Springer*, Vol. 23, No. 3-4, pp. 1123-1127.

Samui, P. and Kothari, D. P. (2012), "A Multivariate Adaptive Regression Spline Approach for Prediction of Maximum Shear Modulus and Minimum Damping Ratio", *Engineering Journal*, Vol. 16, No. 5, pp. 69-78.

Stamatis, D. H. (2016), *Six Sigma and Beyond: The Implementation Process*, 1st edition, Vol. 7, CRC Press, Boca Raton, Florida, 656 pp.

Tachie-Obeng, E., Akponikpe, P. B. I. and Adiku, S. (2013), "Considering effective Adaptation Options to Impacts of Climate Change for Maize Production in Ghana", *Environmental Development, Elsevier*, Vol. 5, pp. 131-145.

Wang, M., Li, Y., Ye, W., Bornman, J. F. and Yan, X. (2011), "Effects of climate change on maize production, and potential adaptation measures: a case study in Jilin Province, China", *Climate Research*, Vol. 46, No. 3, pp. 223-242.

Wongnaa, C. A., Awunyo-Vitor, D., Mensah, A. and Adams, F. (2019), "Profit Efficiency among Maize Farmers and Implications for Poverty Alleviation and Food Security in Ghana", *Scientific African*, Vol. 6, p. e00206.

Yakubu, I., Ziggah, Y. Y. and Peprah, M. S. (2018), "Adjustment of DGPS Data using Artificial Intelligence and classical Least Square Techniques", *Journal of Geomatics*, Vol. 12, No. 1, pp. 13-20.

Ye, G., Jin, Z., Xia, B. and Skitmore, M. (2015), "Analyzing Causes for Reworks in Construction Projects in China", *Journal of Management in*

Engineering, American Society of Civil Engineers, Vol. 31, No. 6, p. 04014097.
Zakari, D. M., Mohammed, A.B., Medugu, N. I. and Sandra, I. (2014), "Impact of Climate Change on Yam Production in Abuja, Nigeria", *International Journal of Science, Environment and Technology*, Vol. 3, No. 2, pp. 458-472.

Authors

A. Buabeng is a Lecturer at the Mathematical Sciences Department of the University of Mines and Technology (UMaT), Tarkwa, Ghana. He holds an MPhil in Mathematics (Statistics). He is currently pursuing his PhD in Mathematics. He is a member of Ghana Mathematics Society, Ghana Statistical Association and the International Association of Engineers (IAENG). His research interests are in Machine Learning, Artificial Neural Networks (ANN), Dimensionality Reduction and Multivariate Prediction for Quality Control.

L. Brew is a Senior Lecturer at the Department of Mathematical Sciences at the University of Mines and Technology (UMaT), Tarkwa, Ghana. He holds a PhD in Mathematics from UMaT. He is a member of the Mathematics Association of Ghana (MAG), Swiss Mathematical Society (SMS) and International Association of Engineers (IAENG). His research interest includes Algebraic topology, Abstract Algebra, Homological Algebra, Statistical Analysis and Time Series Analysis.

A. Odoi is a Lecturer at the Department of Mathematical Sciences at the University of Mines and Technology (UMaT), Tarkwa, Ghana. He holds a PhD in Statistics from UMaT**.** He is a member of the African Institute of Mathematical Sciences (AIMS), Professional Statistical Society of Nigeria (PSSN), Applied Statistics Association of Sri-Lanka (ASASL). His research areas cover Applied Statistics, Bio-Statistics, Financial Statistics, Data Mining and Big Data Analytics.

F. Obiri-Yeboah is a National Service Personnel at the Minerals Commission of Ghana, Tarkwa. He holds a BSc in Mathematics from the University of Mines and Technology (UMaT), Tarkwa, Ghana. His areas of interest include Statistical Analysis, Statistical Modelling, Multivariate and Time Series Analysis.