ISSN 0258-7122 (Print), 2408-8293 (Online)
Bangladesh J. Agril. Res. 48(4): 433-449, December 2023

A MODEL OF ESTIMATION MAIZE YIELD BASED ON WEATHER, AGRONOMICAL AND SATELLITE DATA

N. MOHAMMAD[1], M. A. ISLAM[2], M. M. RAHMAN[3]
I. AHMED[4] AND M. G. MAHBOOB[5]

Abstract

A timely and reliable system of maize yield forecasting well in advance is prime emphasis to farmers and other people who are dependent on cereal crop. The best model was generated using maize field experiment trial which was conducted at Dorbasta union of Gobindagonj upazila, Gaibandha during two consecutive Rabi crops growing season 2018-19 and 2019-20. Randomized Complete Block Design (RCBD) along with five treatments (or varieties) and three replications were considered for maize yield performance. The agronomical and weather parameters and also, satellite data (Landsat 8 OLI) were used for the required maize field experiment. We found that Normalized Difference Vegetation Index (NDVI) was strongly positively correlated with the weather variables in this study. Stepwise regression method was applied for generating best estimated model. Best estimated model (Backward elimination) showed that only five controlled variables which were variety 5 (BHM 13), 1000 grain weight, diameter of cob, plant height and NDVI that were factors to the yield of maize. The developed maize yield forecast model (ideal model) including agronomical, weather and satellite data give the better results of yield estimation at regional level on the basis of best model criterion. Therefore, the ideal model used in specific region including all types of data that gives more precise result on maize yield or production that should be more significant and reliable in national level. So, the researcher, policymaker can use this maize yield prediction model forty to fifty days earlier of harvesting time.

Keywords: Landsat 8, NDVI, BHM13, Multicollinearity, Stepwise regression.

Introduction

The effects of weather change on crop production are global concerns, but they are particularly significant for the sustainable agricultural development of Bangladesh (Hossain et al., 2013). This is a country of variant weather conditions year-round due to its geographic position and physiographic status. Agriculture in Bangladesh is already under pressure, both from huge and increasing demands for food as well as from obstacles related to the degradation of agricultural land and water endowments (Ahmed and Ryosuke, 2000). Bangladesh is one of the countries where crop production especially maize, rice, wheat and potato are most likely to suffer adverse impacts from anthropogenic climate change. Besides, due to the

[1,3,4]Agricultural Statistics & ICT Division, Bangladesh Agricultural Research Institute (BARI), Gazipur, Bangladesh, [2]Department of Agricultural and Applied Statistics, Bangladesh Agricultural University (BAU), Mymensingh, Bangladesh, [5]Forestry Unit, Bangladesh Agricultural Research Council (BARC), Dhaka, Bangladesh.

large share of the agricultural sector in the overall output of the country's economy, the study of weather modification impacts on Bangladesh agriculture has achieved recent attention (Chowdhury and Khan, 2015).

In rice based cropping system, maize is becoming a prominent crop in Bangladesh. Maize is become more popular among the farmers usually due to high yield, higher financial return and numerous uses in recent years. In Bangladesh, maize production and area have grown significantly over the past ten years, and it still growing quickly at an average rate of 20% each year (CIMMYT, 2008; BBS, 2020). It is found that hybrid maize is an emerging high-value crop, having the highest average farm yields in Asia. Economically, hybrid maize is much more profitable than Boro rice, wheat, or most other competing Rabi crops. The hybrid varieties of maize have the highest yield potentiality among cereals. Cultivation of hybrid maize sometimes becomes difficult for many farmers due to unavailability of high quality seed. To fulfill the requirement, a huge amount of hybrid maize seed is now being imported from abroad at a very high cost. Farmers are now growing maize mostly in Rabi season. So, it is necessary to develop maize yield estimation model on the basis of agronomical, weather and satellite parameters in regional level of Bangladesh. The pre harvest forecasting of the crop productivity is a chief precedence to know about the market demand of the crops (Kumar and Shitap, 2020).

The best variable selection/stepwise method is applied to determine the factor(s) that contributed more to crop production in Bangladesh. The factors were the predictor variables and crop production was the dependent variable. Steps were taken to be able to diagnose the identical factors benefaction to the crop production and possibly the combination of factors contribution to crop production. This is in order to establish the factor(s) to give more awareness to increase crop production for lucrative investment in agriculture and food security (Udokang, 2020). The stepwise regression analysis reveals a few variables that have a significant impact on mango and banana yields. The government or policymakers should concentrate a major emphasis on these variables to support the overall development of the cropping system under consideration (Rathod and Mishra 2018). Peiris et al. (2008) predicted coconut production in Sri Lanka using seasonal climate information. Mijinyawa and AkpenPuun (2015) observed in Kwara State, Nigeria that the impact of climate on crop yield was significant for maize and rice yield at 95% probability level while the impact of climate on the yield of millet, sorghum and cowpea was insignificant. So, it is recommended that to escalate the cultivation of crops on which climate had no significant impact on their yield. Kumar and Shitap (2020) used stepwise regression method for forecasting the groundnut productivity in the Junagadh district of Gujarat in India.

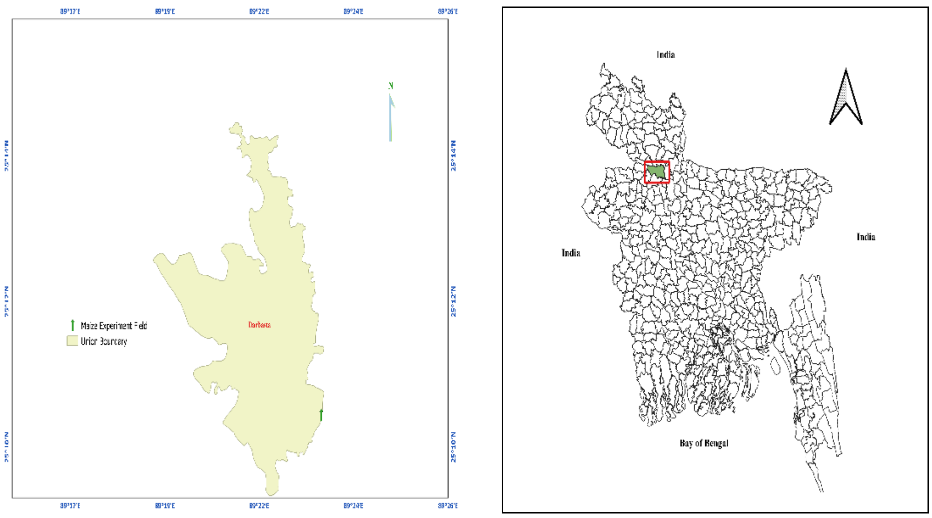
Agarwal et al. (1986), Yang et al. (1992), Dixon et al. (1994), Garde *et al.* (2012), Rathod *et al.* (2012) and Kandiannan *et al.* (2002) proposed methods based on multiple weather-based regression analysis to capture the effect of climate variables on crop yields. Tannura et al. (2008) explored that the explanatory power of the MLR models were much better and they revealed how crop yield and weather conditions were related to one another. Thus, the accuracy of the multiple regression process is much better than the simulation approach and it is also much easier to deal with a multiple regression process as compared to a simulation process. In most of the prior studies linear regression was applied to predict the crop yield, area and production. Therefore, if we consider the number variables, it leads to over fitting of the model and may lead to the problems of multicoliniarity. To overcome these problems stepwise regression is used in this study.

However, the methods of studying the relationship of maize agronomical data, weather parameters as well as remote sensing data were employed in this research. The objectives are to examine the relationship between selected weather or climatic elements and the agronomical parameter of maize crops in the selected location i.e., experiment site and generate the predicted regression model based on the contribution of weather element, remote sensing data as well as yield and yield contributing characters of maize data. All statistical analyses were done using R programme while satellite image processing and NDVI value calculation were done by QGIS software.

Materials and Methods

Study Area

The maize experiment field was set up at Dorbasta union of Gobindagonj upazila, Gaibandha during two consecutive crops growing season 2018-19 and 2019-20. This field was selected through On Farm Research Division (OFRD), Gaibandha. The latitude and longitude of this experiment field were 25° 10′ N and 89° 23′ E respectively which are represented in Figure 1.



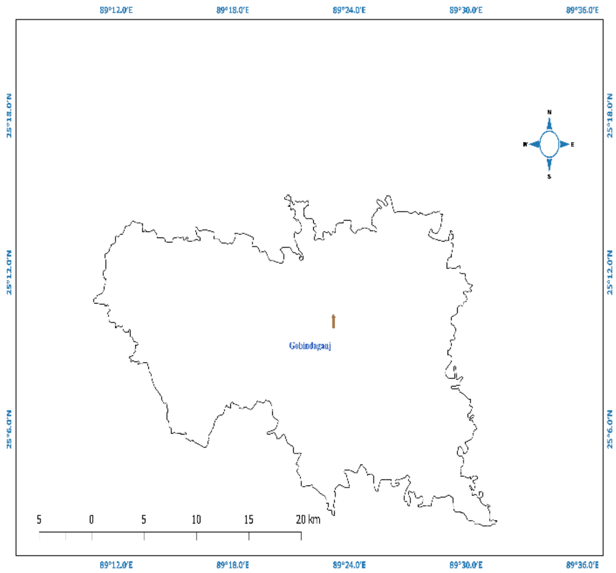


Fig. 1. Study area at Dorbasta union, Gobindagonj, Gaibandha

Field Data

Five treatments (or varieties viz. PS-999, VISHAL-6405, EUREKA, BHM9 and BHM13) were set up at experimental field in Dorbasta union of Gobindagonj upazila, Gaibandha during Rabi season of two consecutive crop growing seasons 2018-19 and 2019-20. Maize seeds were sown on 7th November, 2018 as well as on 7th November, 2019 for two consecutive crop growing seasons. Plot Size was taken 4.2mx5m (7 Lines per Plot). Spacing was 60 × 20 cm between rows and hills, respectively. After seedling emergence one healthy plant per hill was kept. Fertilizers dose and irrigation management were applied as per recommendation. All intercultural operations were done according to the necessity. Agronomical parameter viz. yield and yield contributing parameters of maize viz., days to maturity (DM), plant height (PH), no. of grain/cob (NGC), 1000 grain weight (TGW), length of cob (LOC), diameter of cob (DOC) and grain yield (GY) were collected from maize experiment during two crop growing seasons. Data on plant height (cm), 1000 grain weight (g), length of cob (cm), diameter of cob (cm) and grain yield (t/ha) were recorded from five randomly selected plants. Experimental design was Randomized Complete Block Design (RCBD) with three replications.

Weather Data

The weather parameters viz., maximum temperature (°C), minimum temperature (°C), sunshine hours (h) and relative humidity data (%) are collected from regional weather station at Rangpur of Bangladesh Meteorological Department (BMD). All weather parameters were collected during the maize growing seasons 2018-19 and 2019-20.

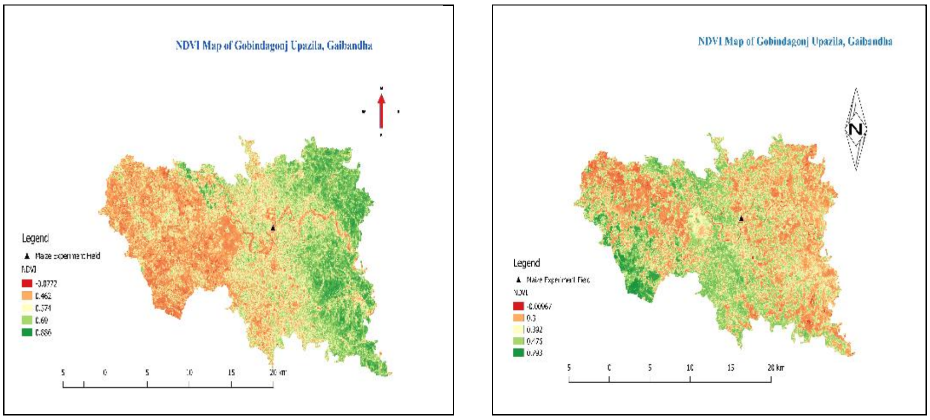
Satellite Data

Landsat 8 images (OLI) were obtained from the United States Geological Survey (USGS) Earth Explorer website (http:// earthexplorer.usgs.gov/). We downloaded total 2 (Two) images, which was maximum cloud free, collected from Landsat 8 OLI satellite data for Dorbasta union of Gobindagonj upazila, Gaibandha during Rabi season of two consecutive crop growing seasons 2018-19 and 2019-20. A total of 2 images were suitable for two experimentation period and used in the consequent analyses (Table 1).

Table 1. Satellite data used in this experiment field

Satellite/Sensor	Acquisition Date (Year-Month-Day)	Spatial Resolution (m)	Path Number	Row Number	Spectral Resolution
Landsat 8 OLI	2019-03-03	30	138	043	9
	2020-02-18				

The single date of image acquisition based on maximum greenness was used for each of the maize experimentation periods i.e. 2018-2019 and 2019-2020 (Figure 2). The maize sowing and harvesting date were considered to be the first week of November and the first week of April for maize experimentation periods i.e. 2018-2019 and 2019-2020 respectively for the entire study site based on the information taken from the location visits.



a b

Fig. 2. Spatial distribution of the NDVI map at Gobindagonj upazila, Gaibandha for Landsat 8 satellite images during growing seasons 2018-19 and 2019-20.[a. 116th days after plantation during 2018-19; b. 101th days after plantation during 2019-20.]

Normalized Difference Vegetation Index (NDVI)

NDVI is a ratio of near-infrared and red band reflectance and is accepted as a surrogate for primary production and has served as a good measure of seasonal vegetation changes, even at coarse scales. The formula for computing NDVI is as follows:

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (1)$$

Where RED (Visible red) and NIR (Near infrared) are reflectance measurements for RED and NIR bands, respectively. Here for Landsat 8/OLI, band 4 and band 5 represented RED and NIR bands. NDVI are related to vegetation amount until saturation at full canopy cover and are therefore related to the biophysically active radiation, efficiencies and productivity (Rondeaux et al., 1996).

Multiple Linear Regressions (MLR) model

The Multiple Linear Regression (MLR) models are applied when two or more independent variables are influencing the dependent variable. It uses a few variables or all variables for prediction as necessary to get a reasonably accurate forecast. The MLR model is presented as follows:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + ... + b_n X_n + e_t \quad (2)$$

Where, b_0 is the intercept, b_i's are the coefficients representing the contribution of the independent variables on the dependent variable Y and e_t is the error at time t is N (i i d) with zero mean and finite variance (Drapper and Smith 1966). R^2 refers the proportion of all variation in the n observed values of the response variable that is expressed by the overall regression model (Bowerman et al., 2005). The higher the R^2, the better the model fits the data (Levin & Rubin, 1994).

Stepwise Regression Analysis

The identification of regressors that are actually influencing the target variable is a crucial aspect of regression modeling. Stepwise regression analysis (SRA) is most often used in exploratory research (Armstrong, 1970). There are numerous methods for selection; stepwise regression analysis is one of the most popular used variable identification algorithms in regression methods. There are two main approaches to analyzing data using stepwise regression (Wang & Jain, 2003). Forward selection and backward elimination are combined in the stepwise regression approach. This is a refinement on forward selection in which at each step all independent variables that have formally been entered into the model are appraised again using their partial F-statistics. An independent variable added at a previous stage may now be superfluous because of the relationships between it and most recent variable entered in the model; backward elimination is a method of stepwise regression where all independent variables begin in the model and subsequent variables are eliminated (Montgomery et al., 2003). By referring Formulas, standard stepwise regression both adds and removes independent variables as needed for each step. According to analysis procedure when all variables not in the model have p-value that are less than the specified Alpha-to-Enter value and when all variables in the model have p-value that are greater than or equal to the specified Alpha-to-Remove value. Alpha-to-Enter is a value that determines if any of the predictors that is not currently in the model should be added to the model. While Alpha-to-Remove is a value that determines if any of the predictors in the model should be removed from the model (Ghani and Ahmad 2010). The predictor variables finally preferred by the stepwise algorithm were incorporated into the final model (Equation 2).

Results and Discussion

Maize yield performance in experiment field

The performance of five hybrids maize varieties at experimental field of two consecutive crops growing seasons 2018-19 and 2019-20, respectively is presented in Table 2. Highly and statistically significant differences were found for all the studied traits during the crop growing season 2018-19 as well as highly and statistically significant differences were found except length of cob during the crop growing season 2019-20. Days to maturity ranged from 148 to 152 and 149 to 152 days during two consecutive crops growing seasons 2018-19 and 2019-20, respectively. The lowest days to maturity (148 days) was found in BARI hybrid maize (BHM9) as well as 149 days for VISHAL-6405 and BHM9 and the highest was in local hybrid variety PS-999 (152 days) in two consecutive crops growing season 2018-19 and 2019-20, respectively.

Plant height was the lowest (176.53 cm) in BHM9 and the highest (181.40 cm) in EUREKA for season of 2018-19 and that was the lowest (177.97 cm) in BHM9 and the highest (182.73 cm) in EUREKA for season of 2019-20. Length of cob is maximum (19.83 cm) for variety VISHAL-6405 and minimum (19.10 cm) for BHM13 for season of 2018-19 and highest (19.57 cm) and lowest (18.70 cm) for variety for VISHAL-6405 for season of 2019-20. The lowest thousand grain weight were found 393 g and 395 g in BHM9 and the highest were found 480 g and 482 g in PS-999 during crops growing season 2018-19 and 2019-20, respectively. The hybrid variety VISHAL-6405 showed the highest yield which was (12.43 t/ha) and (12.52 t/ha) followed by BHM 13 which were (10.83 t/ha) and (10.82 t/ha) during crops growing seasons 2018-19 and 2019-20, respectively.

Table 2. Performance of maize hybrids for different traits under the maize experimental field at Dorbasta union of Gobindagonj upazila, Gaibandha during the season of 2018-19 to 2019-20

Maize Growing Season 2018-2019								
Sl. No.	Variety	DM	NGC	PH (cm)	TGW(g)	LOC (cm)	DOC (cm)	GY(t/ha)
1	PS-999	152	453	178.63	480	19.17	17.03	11.84
2	VISHAL-6405	149	632	177.43	461	19.83	17.33	12.43
3	EUREKA	151	468	181.40	418	19.37	17.03	11.81
4	BHM9	148	495	178.77	393	18.57	16.70	11.22
5	BHM13	150	505	176.53	443	19.10	17.27	10.83
	Min	148	453	176.53	393	19.10	16.70	10.83
	Max	152	632	181.40	480	19.83	17.33	12.43
	LSD (0.05)	1.19	7.60	0.99	4.25	0.52	0.27	0.82
	CV (%)	0.422	0.79	0.29	0.51	1.45	0.84	3.73
	F-test	***	***	***	***	**	**	**
Maize Growing Season 2019-2020								
Sl. No.	Variety	DM	NGC	PH (cm)	TGW(g)	LOC (cm)	DOC (cm)	GY(t/ha)
1	PS-999	152	461	180.03	482	19.50	16.97	12.03
2	VISHAL-6405	149	632	179.80	453	19.57	17.47	12.52
3	EUREKA	151	478	182.73	417	19.23	16.67	12.24
4	BHM9	149	496	179.23	395	18.70	17.30	11.62
5	BHM13	150	509	177.97	442	19.10	17.27	10.82
	Min	149	461	177.97	395	18.70	16.67	10.82
	Max	152	632	182.73	482	19.57	17.47	12.52
	LSD(0.05)	1.68	7.55	0.93	10.68	0.63	0.29	0.59
	CV (%)	0.59	0.78	0.27	1.29	1.75	0.89	2.66
	F-test	**	***	***	***	NS	**	**

***<0.001, **<0.01, *<0.05; NS= Non Significant; Note: DM = Days to maturity, NGC = No. of grain/cob, PH = Plant height, TGW = 1000 grain weight, LOC= length of cob, DOC = Diameter of cob and GY = Grain yield

Variable classification of maize experiment in regression analysis

Regression analysis has been carried out to know the factors influencing yield of maize at experiment field in Dorbasta union of Gobindagonj upazila, Gaibandha during two maize growing seasons viz. 2018-19 and 2019-20, respectively. Regression model was fitted for yield of maize. The response variables in the study are yield of maize, whereas independent variables include agronomical variables, weather variables and satellite parameter like NDVI value listed in Table 3. Here explanatory variables variety and year were treated as dummy. The NDVI values were taken from single date of maximum NDVI distribution on high resolution spatial satellite image like Landsat 8 OLI during two years maize experiment. Here, the NDVI values in maize experimental location were 0.66 and 0.54 after 116th and 101th days after plantation during two maize growing years. The weather data was marked on this date when NDVI values are taken from this date during two years' maize experiment viz. 2018-19 and 2019-20, respectively.

Table 3: Variables considered for classical regression analysis

Notation	Variables	Units
Y	Yield of maize (GY)	Ton per hectare (T ha^{-1})
X_1	Plant height (PH)	Centimeter (cm)
X_2	Days to maturity (DM)	Numbers
X_3	Number of grain/cob (NGC)	Numbers
X_4	1000 grain weight (TGW)	Gram (g)
X_5	Length of cob (LOC)	Centimeter (cm)
X_6	Diameter of cob (DOC)	Centimeter (cm)
X_7	NDVI	
X_8	Maximum temperature (MAT)	Degree Celsius (°C)
X_9	Minimum temperature (MIT)	Degree Celsius (°C)
X_{10}	Sunshine hour (SSH)	Hour (h)
X_{11}	Relative humidity (HD)	Percentage (%)
	Variety (Dummy)	
X_{12}	V2 (VISHAL-6405)	
X_{13}	V3 (EUREKA)	
X_{14}	V4 (BHM9)	
X_{15}	V5 (BHM13)	
X_{16}	Year (Dummy)	

Note: Variety V_1 (PS-999) is considered as reference in all dummy for varieties (X_{12} to X_{15}).

Correlation analysis and Multicollinearity

A correlation analysis is the method in regression that is used to look the strength of the relationship between two variables. Pearson linear correlation was applied to check the existence of relationship between two variables in this research. The

Pearson correlation coefficient (r) between the agronomical parameters, weather parameters, satellite data and the selected maize yields were computed.

From the correlation matrix (Figure 3), we saw that weather parameter like maximum temperature, minimum temperature, sunshine hour and relative humidity were perfectly positively correlated (r =1) with NDVI and other agronomical parameters as well as the three dummy of variety like V2 and V3 and year were also positively correlated with other agronomical parameters. Excluding these variables (7 variables) which are highly positively or negatively correlated and the variables which are insignificant in this correlation matrix. VIF (variation inflation factor) value for those variables was very high or undefined in the regression model. So, it is clearly seen that the multicollinearity problem is strongly present among the independent variables. Multicollinearity refers a situation that has significant degree of correlation between the predictor variables. Any analysis can detect multicollinearity by examining the variance inflation factor's value (VIF). Multicollinearity is not a severe issue when the value of VIF is less than 5. While multicollinearity is noticeable if VIF is more than 5. When the value of VIF exceeds 10, multicollinearity becomes more problematic.

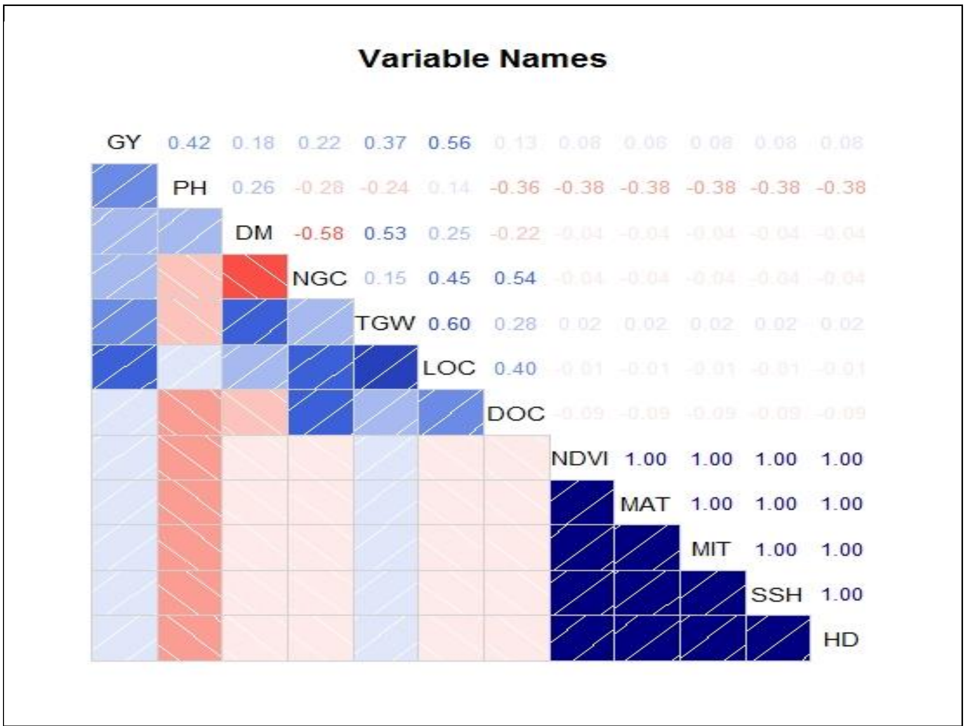


Fig. 3. Correlogram of correlation matrix

Regression analysis of maize yield time series

A total number of 9 independent variables were used in this study. Firstly the multiple linear regression (MLR) analysis was carried out by considering all the explanatory variables. The R^2 of MLR model obtained in Table 4 shows that all predictor variables considered in the study explains 79.90 percent of the variation in response variable. Though the R^2 of MLR model is very high but all of the variables in the model are non-significant and variance inflation factor (VIF) is also high in majority of the variable. This clearly denotes the multicollinearity problem among the explanatory variables. To overcome the same, one of the measures is to remove unimportant variables which are explaining less variation in dependent variables in the model (Gujarati et al., 2013). The dropping and adding of variable can be done using the stepwise regression analysis.

Table 4. Multiple linear regression analysis of maize yield

Variable	Coefficient	Std. Error	t-test	P-value	VIF
Constant	-38.341836	28.840282	-1.329	0.1987	
X_1	0.193471	0.119094	1.625	0.1199	9.974900
X_2	-0.020304	0.110233	-0.184	0.8557	5.836325
X_3	0.001151	0.002904	0.396	0.6962	6.972129
X_4	0.010471	0.006708	1.561	0.1342	8.917560
X_5	0.091040	0.303554	0.300	0.7673	4.000540
X_6	0.542524	0.301391	1.800	0.0870	1.999086
X_7	3.406770	1.817878	1.874	0.0756	2.525861
X_{14}	0.115507	0.505741	0.228	0.8217	8.688687
X_{15}	-0.803040	0.417701	-1.923	0.0689	5.926925
R^2			0.799		
F Statistic			8.83		
P-value			0.0003		

Note: X_1 = Plant height (PH), X_2 = Days to maturity (DM), X_3 = Number of grain/cob (NGC), X_4 = 1000 grain weight (TGW), X_5 = Length of cob (LOC), X_6 = Diameter of cob (DOC), X_7 = NDVI, X_{14} = V4 (BHM9) and X_{15} = V5 (BHM13)

Stepwise regression analysis: Generating maize yield model

The statistical significance of each predictor variable in a linear regression model is iteratively examined through stepwise regression. The forward selection procedure starts from zero and gradually integrates each additional variable while evaluating for statistical significance. The backward elimination method begins with a full model that is loaded with many variables, and then it eliminates one variable to determine its significance in regard to the overall outputs. Stepwise regression is a technique for fitting data into a model to generate the desired output, but it has drawbacks as well.

Hence, stepwise regression analysis was carried out to fit the model. The detailed summary of stepwise regression for backward elimination process is given in Table 5.

Table 5. Stepwise regression (Backward elimination) analysis of maize yield

Variable	Coefficient	Std. Error	t-test	P-value	VIF
Constant	-36.099913	11.785032	-3.063	0.005338	
X_{15}	-0.962188	0.202358	-4.755	0.000078	1.569513
X_4	0.009194	0.002241	4.102	0.000407	1.123281
X_6	0.730373	0.228328	3.199	0.003852	1.294541
X_1	0.163468	0.052317	3.125	0.004607	2.171904
X_7	3.129348	1.259195	2.485	0.020312	1.367392

Note: X_1 = Plant height (PH), X_4 = 1000 grain weight (TGW), X_6 = Diameter of cob (DOC), X_7 = NDVI and X_{15} = V5 (BHM13).

The stepwise regression for maize yield data was completed in five steps, the maximum R^2 was increased in each step and it was obtained 78.6 percent in final steps. The unexplained or non-significant variables are removed from the model so that we can get maximum error degrees of freedom. In this stepwise regression analysis i.e., backward elimination process, we obtained total five significant independent variables in Table 5 as compare to MLR model in Table 4 in which no one of the variables were significant. Here the result from the Table 5 shows that multicollinearity is not present among the desired predicted variables i.e., ($1 < VIF < 5$). The results of stepwise regression for backward elimination illustrate that the variable like variety 5 (X_{15}) i.e., BHM13 significantly influence the changes in the maize yield which explains highly significant in this model. Based on this result we can say that as the variety BHM13 significantly contributes the changes in maize yield (t/ha). The variable like 1000 grain weight (X_4), diameter of cob (X_6), plant height (X_1) and NDVI values (X_7) also significantly influence the changes in maize yield and also production.

Table 6. Stepwise regression (Forward selection) analysis of maize yield

Variable	Coefficient	Std. Error	t-test	P-value	VIF
Constant	-26.473339	11.553187	-2.291	0.031015	
X_4	0.010329	0.002367	4.364	0.000209	1.082183
X_{15}	-0.783426	0.229129	-3.419	0.002249	1.738747
X_1	0.168722	0.058789	2.870	0.008434	2.369708
X_3	0.002939	0.001244	2.362	0.026628	1.247862
X_7	2.933192	1.355513	2.164	0.040646	1.369191

Note: X_1 = Plant height (PH), X_3 = Number of grain/cob (NGC), X_4 = 1000 grain weight (TGW), X_7 = NDVI and X_{15} = V5 (BHM13)

Analogously, the precise summary of stepwise regression for forward selection approach is given in Table 6. The forward selection for maize yield data was executed in ten steps, the maximum R^2 obtained was 75.26 percent and it was presented in last step. The unimportant or insignificant variables are dropped from the model so that we can get maximum error degrees of freedom. In this forward selection approach, we obtained total five significant explanatory variables (Table 6). Here the output show that multicollinearity is not present among the desired controlled variables i.e., (1 < VIF < 5). Outputs of stepwise regression for forward selection demonstrate that the variable like 1000 grain weight (X_4) significantly contributes in increasing the maize yield which explains highly significant in this model. Based on this output we can say that as 1000 grain weight (TGW) increases, the maize yield (t/ha) also increases. The variable like variety 5 (X_{15}), the plant height (X_1), number of grain/cob (X_3) and NDVI values (X_7) also significantly influence the changes in maize yield as well as production.

According to Table 5 and Table 6, we can develop two model using MLR equation based on the stepwise regression including backward elimination and forward selection process which are given bellow:

Model 1 (Backward elimination):

$$\hat{Y} = -36.099913 - 0.962188\,X_{15} + 0.009194\,X_4 + 0.730373\,X_6 + 0.163468\,X_1 + 3.129348\,X_7$$

Model 2 (Forward selection):

$$\hat{Y} = -26.473339 + 0.010329\,X_4 - 0.783426\,X_{15} + 0.168722\,X_1 + 0.002939\,X_3 + 2.933192\,X_7$$

Based on the Model 1, the coefficients for variety 5, 1000 grain weight, diameter of cob, plant height and NDVI values were ($P<0.05$) significant. This explained that these coefficients are factor to the maize yield for backward elimination. While in Model 2, the coefficients for 1000 grain weight, variety 5, the plant height, number of grain/cob and NDVI values were ($P<0.05$) significant. This explained that these coefficients are factor to the maize yield for forward selection.

Model simulation and best model assessment

According to Table 5 and Table 6, it presents the result of stepwise regression using open source program R. There are two approaches that used to select the independent variables. In the approach 1, X_{15} indicates smallest p-value than Alpha-to-Enter (p=.000<.05). Therefore X_{15} is the first variable that enters into the model (backward elimination). In the approach 2, the p-value for X_4 shows 0.000209 which is smallest than 0.05. Therefore X_4 is the first variable that enters into the model (forward selection). Model 1 and model 2 both includes 5 predicted variables after running step wise regression. After that there are no controlled variables that enter and remove from those models. The best model is selected after model simulation from model 1 and model 2 based on model accuracy criteria which are presented in Table 7.

Table 7. Best model selection based on model selection criteria

Model	R^2	$\overline{R^2}$	Cp	NRMSE	MAE	RMSE	MAPE
Model 1	0.79	0.74	6	2.77	2.57	0.317	0.022
Model 2	0.75	0.70	9.78	2.98	2.64	0.341	0.023

Note: AIC- MAE- Mean Absolute Error; NRMSE- Normalized Root Mean Square Error; RMSE- Root Mean Square Error; MAPE- Mean Absolute Percentage Error; Cp- Mallows Cp.

For the results of maize yield in Table 7, the value of R^2 and adjusted R^2 (79% and 74%) higher for model 1 than model 2 (R^2=75% and adjusted R^2=70%). MAE, NRMSE, RMSE and MAPE values of model 1 are smaller than model and Mallows Cp in model 1 is closer to number of controlled variables than model 2. Hence, these statistics indicates model 1 which containing controlled variables variety 5 (X_{15}), 1000 grain weight (X_4), diameter of cob (X_6), plant height (X_1) and NDVI (X_7) values are provided better fits to the data. Weather variables are not significantly effect in this model. The final model (best model) shows only five controlled variables by using stepwise regression (backward elimination) and produced regression model such as follows:

$$\hat{Y} = -\,36.099913 - 0.962188\,X_{15} + 0.009194\,X_4 + 0.730373\,X_6 + 0.163468\,X_1 + 3.129348\,X_7$$

However, depending on the availability of information, on necessary variables (X_3, X_6) alter model (Model 2) may be used in practice. The stepwise regression analysis suggested two competing models both of which including X_{15} (Variety BHM13). However, despite being the best possible models these two lack the prediction efficiency for the entire country /region where different varieties are cultivated.

Conclusions

The crop yield forecast models have been improved by considering experimental data on the maize crop at Dorbasta union of Gobindagonj upazila, Gaibandha in this study during two consecutive crops growing season 2018-19 and 2019-20, respectively. Five hybrids maize varieties like viz. PS-999, VISHAL-6405, EUREKA, BHM9 and BHM13 are performed almost better during the maize growing period. Highly and statistically significant differences were found for all the studied variables during the crop growing season 2018-19 and except length of cob during the crop growing season 2019-20.

Local or regional based crop yield forecasting conveys the maximum emphasis than national level crop yield estimation as NDVI varies field to field. We found that NDVI was perfectly positive correlated (r=1) with the weather variables in this research. And also weather variables were not significant in this model. Actually, NDVI is a cumulative contribution of weather parameters (Hao et al., 2012; Pei et al., 2019). In this study multiple linear regressions (MLR) like the stepwise regression analysis was used for maize yield forecasting. Stepwise regression viz., forward selection and backward elimination were generated to build forecasting model or ideal model for maize yield. Backward elimination (Model 1) obtained totally five significant predictor variables and similarly for Forward selection (Model 2). Multicollinearity was not happened strongly among the desired predicted variables for both. By comparing two models, Backward elimination (suggest as a best model) is a better model fits to the data than Forward selection. Therefore, the best model or the forecast developed maize yield model used in specific region including all types of data that gives more precise result on maize yield or production that should be more significant and reliable in national level. This best model which can be used by the researcher, policymaker for maize yield prediction.

References

Agrawal, R., R.C. Jain, and M.P. Jha. 1986. Models for studying rice crop weather relationship. Mausam, **37(1):** 67-70.

Ahmed, A., S. Ryosuke. 2000. Climate Change and Agricultural Food Production of Bangladesh:An Impact Assessment Using GIS-Based Biophysical Crop. Simulation Model; Center for Spatial Information Science, University of Tokyo: Komaba, Japan, 2000.

Armstrong, J. 1970. How to avoid exploratory research. *Journal of Advertising Research*, **10(4):** 27–30.

BBS, 2020. Yearbook of Agricultural Statistics, 2020. 32nd series, published: 21th May, 2021. Bangladesh Bureau of Statistics (BBS), Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka, Bangladesh.

Bowerman, B.L., R.T. O'Connell and A.B. Koehler. 2005. Forecasting, time series and regression. 4th ed. United States of America: Brooks/Cole Thomson Learning Inc.

Chowdhury, I.U.A. and M.A.E. Khan. 2015. The Impact of Climate Change on Rice Yield in Bangladesh: a Time Series Analysis. *RJOAS*, 4(40), April 2015.

CIMMYT Bangladesh, 2008. Maize whole family training. In: Food Security in Bangladesh: Improving Wheat, Maize and Papaya Production and Impacts of Arsenic Contamination. Project Annual Report for 2007-2008, pp.31-44.

Dixon, B.L., S.E. Hollinger, P. Garcia and V. Tirupattur. 1994. Estimating Corn Yield Response Models to Predict Impact of Climate Change, *Journal of Agricultural and Resource Economics*, **19(1):** 58-68.

Draper, N.R. and H. Smith. 1998. *Applied Regression Analysis*, 3rd Edition, John Wiley & Sons, New York.

Garde, Y.A., S. Singh, G.C. Mishra and T. Singh. 2012. Weather Based Pre-Harvest Forecasting of Wheat at Ghazipur (U.P.), *International Journal of Agricultural Sciences*, **8(2):** 325-328.

Ghani, I.M.M. and S. Ahmad. 2010. Stepwise Multiple Regression Method to Forecast Fish Landing. *Procedia Social and Behavioral Sciences* 8 (2010) 549–554.

Gujarati, D.N., D.C. Porter and S. Gunasekar. 2013. Basic Econometrics (Fifth Edition), Tata McGraw-Hill Education Pvt. Ltd, ISBN 10: 0071333452 / ISBN 13: 9780071333450

Hao, F., Z. Xuan, O. Wei, K. S. Andrew and A. G.Toxopeus. 2012. Vegetation NDVI linked to temperature and precipitation in the upper catchments of yellow river. *Environ model assess* (2012) **17:** 389-398.

Hossain, A., Teixeira da Silva, J.A. Phenology.2013. Growth and Yield of Three Wheat (*Triticum.*aestivum L.) Varieties As Affected by High Temperature Stress. *Not. Sci. Biol.* **2013**, 4, 97–106.

Kandiannan, K., R. Karithikeyan, R. Krishnan, C. Kailasam and T.N. Balasubramanian. 2002. A Crop-Weather Model for Prediction of Rice Yield Using an Empirical-Statistical Technique, *Journal of Agronomy and Crop Science*, **188:** 59-62.

Kumar, K. S. and M. Shitap. 2020. Statistical evaluation of stepwise regression method and autoregressive integrated moving average method for forecasting of groundnut (arachis hypogaea l.) productivity in Junagadh district of Gujarat. *Int.J.Curr.Microbiol.App.Sci* (2020) **9(11):** 84-93.

Levin R.I. and D.S. Rubin. 1994. Statistics for management. 6th ed. Englewood Cliffs, N.J.: Prentice Hall.

Mijinyawa, Y. and T. D. Akpenpuun. 2015. Climate change and its effect on grain crops yields in the middle belt in Nigeria. *African Journal of Environmental Science and Technology.* Vol. 9(7), pp. 641-645, July, 2015.

Montgomery, D. C., E.A. Peck and G. Vining. 2003. *Introduction to Linear Regression Analysis.* 3rd Edition, John Wiley and Sons (Asia) Pte. Ltd.

Pei, Z., F. Shibo, Y. Wunian, W. Lei, W. Mingyan, Z. Qifei, H. Wei and N. K. Dao. 2019. The relationship between NDVI and climate factors at different monthly time scales: A case study of grasslands in inner Mongolia, China (1982-2015). *Sustainability* 2019,11,7243.

Peiris, T. S. G., J. W. Hansen and L. Zubair. 2008. Use of Seasonal Climate Information to Predict Coconut Production in Sri Lanka. *Int. J. Climatol*., 28: 103–110. doi: 10.1002/joc.1517.

Rathod, S. and G. C. Mishra. 2018. Statistical Models for Forecasting Mango and Banana Yield of Karnataka, India. *J. Agr. Sci. Tech*. (2018) Vol.20: 803-816.

Rathod, S., H.S. Surendra, R. Munirajappa, and K.B. Murthy. 2012. Statistical Assessment on the Factors Influencing Agriculture Diversification in Different Districts of Karnataka. *Environment and Ecology*, 30 (3A): 790-794.

Rondeaux, G., M. Steven & F. Baret. 1996. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, ***55*****(2):** 95-107.

Tannura, M.A., S.H. Irwin and D.L. Good. 2008. Weather technology and corn and soybean yield in the U.S. Corn Belt, Marketing and Outlook Research Report 2008–2011, Department of Agricultural and ConsumerEconomics, University of Illionis, Illionis.

Udokang, A. E. 2020. A stepwise approach of factors affecting agricultural production in Nigeria. *Science Journal of Business and Management* 2020; 8(3): 149-153 http://www.sciencepublishinggroup.com/j/sjbm doi: 10.11648/j.sjbm.20200803.17.

Wang, G. C., & C. L. Jain. 2003. Regression analysis: Modeling & forecasting. Institute of Business Forec.

Yang, S.R., W.W. Koo and W.W. Wilson. 1992. Heteroscedasticity in crop yield models, *Agricultural and Resource Economics*, **17(1):** 103-109.