Development of Rice yield forecasting model using different statistical technique & artificial neural network approach for Jalpaiguri, West Bengal

[1]Shyamal Mahadani, [2]Dr. Abhishek James

[1]Student, [2]Assistant Professor
Department of Environmental Sciences & NRM
Sam Higginbottom of Agriculture, Technology and Sciences, Prayagraj, India

Abstract : Crop yield forecasting under the present climate change scenario needs an effective model and its parameter that how crop respond to the weather variable. By incorporating regression models and Artificial Neural Networks (ANN), one can make highly satisfactory predictions of rice crop yield. Crop yield forecasting using regression and artificial neural network is developed through SPSS package software which will help agronomists and farmers to get accurate crop yield reports from various agricultural sources. In this study, the focus is on the development of regression techniques and artificial neural network in agricultural field. Historical rice crop yield data and weather data of past fifteen years (2005-2019) were collected for Jalpaiguri District of West Bengal. Analysis was carried out by fixing 15 years of dataset for calibration (80%) and remaining dataset for the validation (20%). Stepwise linear regression technique and neural network used for predicting the yield of rice crop. Performance analysis is done with least Normalized Root Mean Squared Error (nRMSE) values for the study. The present study concluded that the performance of weighted and unweighted model with respect to nRMSE has least error as compared to all other models which is 0.014.

Index Terms: Yield forecasting, weather variable, ANN, nRMSE, kharif rice.

I. INTRODUCTION

Rice is the staple food of over half the world's population. It is the seed of the grass species *Oryza sativa* (Asian rice) or less commonly *Oryza glaberrima* (African rice). Agriculture has always been one of the vital occupations that serve mankind, both in terms of livelihood and employment. Due to the substantial increase in the population, the nutritional status of the poor is growing bad, which must be improved. The major effect of population increase has been prominently shown on the environment, the damage of which is increasing rapidly, which ultimately hinders agricultural production. Studies show that the modern techniques used in agriculture have not been environment-friendly, though they are technologically advanced than the primitive techniques. Yield prediction is one of the most critical issues faced in the agricultural sector. Farmer's lack of knowledge about harvest glut, uncertainties in the weather conditions and seasonal rainfall policies, depletion of nutrition level of soils, fertilizer availability and cost, pest control, post–harvest loss and other factors leads to decrease in the production of the crops. Regression Analysis can be defined as a structured approach which stresses on the analysis of data for the research purpose on decision making and problem solving. There are problems/situations that require simultaneous analysis of multiple variables or objects for efficient decision making.

As the weather parameters plays a vital role in crop production, an attempt has been made to develop the statistical equations using modified **Hendrick and Scholl** method to investigate the feasibility of estimating the yield of *kharif* rice and jute crops in advance based on weather variables using past weather and crop yield records of Jalpaiguri district.

Study area

The study area, Jalpaiguri district extends between 26°15'47" to 26°59'34" N latitude and 88°23'2" to 89°73'0" E longitude comprising an area of 3044.00 km2. Jalpaiguri district is situated in the northern part of West Bengal and is bordered by Bhutan in the north, Darjeeling district in the west and north-west, Koch Bihar and Bangladesh in the south and Alipurduar district in the east.

II. MATERIALS AND METHODS

Data Collection

The present experiment "Developing regression model to forecast the rice yield at Jalpaiguri region." was conducted during the Kharif season of 2022. Yearly Production (Kg) and area (ha) under Rice crop in Jalpaiguri District for the period 2005 – 2019 were collected from the (Directorate of Economics and Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare). For each year, the total production of the district was divided by the total acreage to calculate the rice productivity. Delay data of maximum & minimum Temperature (°C), maximum & minimum Relative Humidity (%) and Rainfall (mm) number of rainy days for the period 2005 – 2019 were collected from the meteorological observatory located in (Regional Meteorological Centre), Kolkata. SPSS places constraints on internal file structure, data type, data processing and matching files, which together consider as the simplified programming. SPSS software was used for the statistical analysis and to develop a multiple regression model for analyzing drought conditions. The Parameters which were taken in the SPSS for the development of model were weather parameters of Kharif Season. Daily data of Tmax, Tmin, MaxRH, MinRH, Average Rainfall had been converted into its weekly average values using Weather cock software. Weighted and Unweighted parameter have considered as independent variable and Kharif Productivity index (Yield) has been taken as dependent variable. The co-efficient of determination (R^2) and correlation coefficient (r) were also calculated with the help of SPSS software.

Development of Yield Prediction Model Using Different Technique

Simple and weighted weather indices are developed for Jalpaiguri District. Summation of individual weather variable or interaction of two weather variable at a time were used for generating simple weather indices, sum product of individual weather variable or interaction of weather variables and its correlation with adjusted crop yield were resulted with weighted weather indices. Computation of simple and weighted weather indices were based on following formula.

Simple weather indices

$$Zij = \sum_{w=1}^{m} Xiw \qquad Zii'j = \sum_{w=1}^{m} Xiw\ Xi'w$$

Weighted weather indices

$$Zij = \sum_{w=1}^{m} r^{j}iw\ Xiw \qquad Zii'j = \sum_{w=1}^{m} r^{j}ii'w\ Xiw\ Xi'w$$

Where,

$XiwXii'w$ = value of *i'th/i'th* weather variable under study in weather week, $r^{j}iw/r^{j}ii'w$ = correlation coefficient of yield with its weather variable or product of *i'*th/*i'*th weather variable in the week, m = week at which forecast done,

Table 1: Simple and weighted weather indices used for developing model

Weather Parameter	Simple weather indices					Weighted weather indices				
	Tmax	**Tmin**	**R/F**	**RH(I)**	**RH(II)**	**Tmax**	**Tmin**	**R/F**	**RH(I)**	**RH(II)**
Tmax	Z10					Z11				
Tmin	Z60	Z20				Z61	Z21			
R/F	Z70	Z110	Z30			Z71	Z101	Z31		
RH (I)	Z80	Z120	Z140	Z40		Z81	Z111	Z131	Z41	
RH (II)	Z90	Z130	Z150	Z150	Z50	Z91	Z121	Z141	Z151	Z51

P= number of variables. Now, Simple and weighted weather indices used for developing model are given in Table 1.

Multiple linear regression analysis

Multiple linear regression models involved more than one independent variable and one dependent variable. It is of the form. Stepwise regression procedure was adopted for selection of the best regression variable among many independent variables **(Singh et al. 2014).**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

Where, β_0 is constant, β_i's are coefficients of X_i's, X_i's are the independent variables also known as predictors and Y is the dependent variable and ε is the error.

Principal component analysis

PCA, is a statistical procedure that allows you to summarize the information content in large data tables by means of a smaller set of "summary indices" that can be more easily visualized and analyzed.

Model Performance

For testing the performance of developed statistical forecasting models, R^2, Root Mean Square Error (RMSE) and Normalized Root Mean Square Error (nRMSE) were calculated using the following formula:

$$R^2 = \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(M_i - \bar{M})(O_i - \bar{O})}{\sigma M \sigma O} \right)^2 \qquad \text{nRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)} \times \frac{100}{\bar{O}} \qquad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)^2}$$

$$\text{APE} = \frac{O_i - P_i}{P_i} \times 100$$

Where ***RMSE*** is Root Mean Square Error, ***nRMSE*** is Normalized Root Mean Square Error, $\boldsymbol{P_i}$ is the predicted value, $\boldsymbol{O_i}$ is the observed value, ***n*** is the number of observations and ***M*** is the mean of observed value.

Model performs excellent having nRMSE value > 10%, good having nRMSE value between 10-20%, fair having nRMSE value between 20-30%

Artificial Neural Network

ANN models are the extreme simplification of human neural systems. An ANN comprises of computational units analogous to that of the neurons of the biological nervous system known as artificial neurons. Mainly, the ANN model constitutes of three layers, viz., input, hidden, and output. Each neuron in the nth layer is interconnected with the neurons of the (n + 1) th layer by some signal. Each connection is assigned a weight. The output may be calculated after multiplying each input with its corresponding weight.

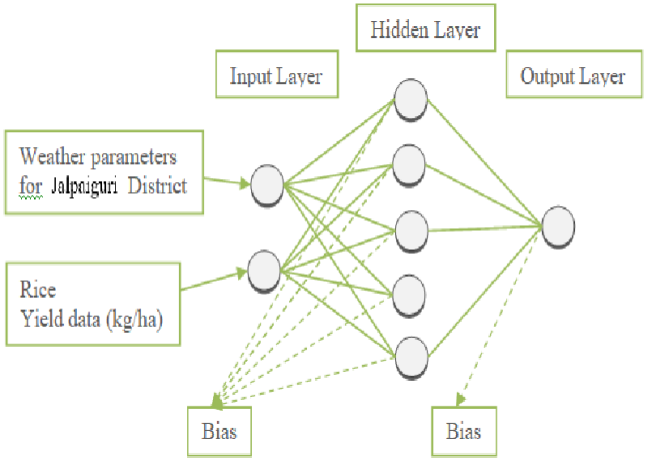


Fig.1. Different layers and different network structural representation of the ANN used in the study.

Multilayer perceptron (MLP) technique is one of the popular neural network types. This network interpreted as a form input-output model, with weights and threshold (biases) as free parameters of the model. By learning process, it attains optimized weighted value of variables, and it tries to produce the output based on the corresponding input provided. The main objective of the neural network is to produce its own output having reduced discrepancies with target output value, which will help to transform the input into meaningful output.

III. RESULTS AND DISCUSSION

Rice yield prediction using stepwise regression method based on different parameters.

The yield variations explained by model together with standard error are shown in Table.2. Coefficient determination (R^2) has been significantly above at 5% probability level for rice in Jalpaiguri. The R^2 was ranged between 85% (Simple Weather-based Model) to 99% (Simple Weather-based Model). However, the percent standard error was ranged between 12.84 (Simple Weather-based Model) and 180.08 (Weighted and Unweighted).

Table.2. Yield Forecast Models of Rice

METHODS	Sl. No	Yield Forecast Model (s)	R^2	SE
Simple Weather-based Model	1.	Y= -6821.004 + 307.409(RH_E37) + 39.086(RH_E44)	0.85	146.97
	2.	Y= -4822.056 + 336.451(RH_E37) + 36.128(RH_E44) – 71.862(tmax_37)	0.97	70.17
	3.	Y= -5124.547 + 406.773(RH_E37) + 45.532(RH_E44) – 72.186(tmax 37) – 91.230(RH_E38)	0.99	36.16
	4.	Y= -5404.583 + 386.657(RH_E37) + 42.047(RH_E44) – 69.525(tmax_37) – 73.949(RH_E38) + 15.615(tmax_40)	0.99	12.84
Weighted and Unweighted Model	1.	Y= 1498.919 + 88.542(T)	0.83	180.08
	2.	Y= 1979.590 + 86.140(T) + 36.752(Z11)	0.95	80.38
	3.	Y= -329.777 + 81.485(T) + 43.146(Z11) + 5.958(Z20)	0.97	61.28
	1.	Y= 1498.915 + 88.542(T)	0.83	148.08
	2.	Y= 1499.204 + 87.185(T) + 38.932(P_5)	0.92	103.32
	3.	Y= 1500.075 + 86.524(T) + 39.552(P_5) – 20.786(P_4)	0.96	75.29

Regression analysis for conceptual model

Principal component analysis	4.	Y= 1540.933 + 81.703(T) + 39.817(P_5) + 20.476(P_4) + 14.683(P_2)	0.98	57.05
Artificial Neural Network	1.	Y= 64.231x - 126987	0.66	-

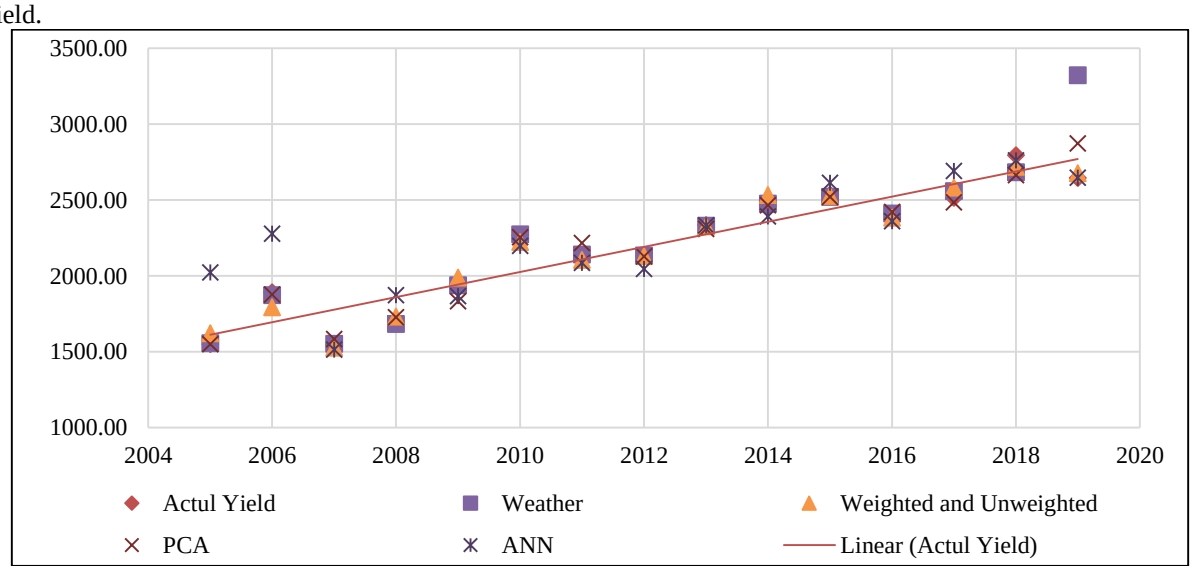
Linear regression analysis was conducted to assess whether the weather parameters (Tmax, Tmin, rainfall, relative humidity-I relative humidity-II) significantly predicted the yield of the rice crop of Jalpaiguri district. The 'Stepwise' variable selection method was used to develop the linear regression model. In the stepwise technique, at each step predictor variables get added and the significance of the model was tested at each step.

Methods		**2017**			**2018**			**2019**		
		Observed Yield	**Predicted Yield**	**Error %**	**Observed Yield**	**Predicted Yield**	**Error %**	**Observed Yield**	**Predicted Yield**	**Error %**
Simple Weather-based Model	**Model 1**	2514	2587.46	2.89	2794	2440.49	12.66	2645	3039	**14.90**
	Model 2	2514	1961.98	21.98	2794	2067.28	26.02	2645	2601.69	**1.64**
	Model 3	2514	2586.83	2.87	2794	2703.34	3.26	2645	3419.72	**29.28**
	Model 4	2514	2559.80	1.79	2794	2682.57	4.00	2645	3321.95	**25.58**
Weighted and Unweighted Model	**Model 1**	2514	2649.97	5.38	2794	2738.51	2.00	2645	2827.05	**6.88**
	Model 2	2514	2566.68	2.06	2794	2770.36	0.86	2645	2676.41	**1.18**
	Model 3	2514	2581.88	2.67	2794	2718.79	2.70	2645	2676.61	**1.19**
PCA Model	**Model 1**	2514	2649.96	5.38	2794	2738.50	2.00	2645	2827.05	**6.88**
	Model 2	2514	2522.75	0.32	2794	2706.83	3.13	2645	2826.96	**6.87**
	Model 3	2514	2506.78	0.32	2794	2657.10	4.91	2645	2826.49	**6.85**
	Model 4	2514	2485.83	1.15	2794	2665.38	4.61	2645	2873.26	**8.62**
Artificial Neural Network	**Model 1**	2514	2691.67	7.03	2794	2761.58	1.17	2645	2647.98	**0.11**

Models had less than ± 10% error in rice yield prediction for all methods during the years. This has indicated that the model can be used for prediction of rice yield through these methods. The result revealed that agrometeorological yield model explained the yield variability due to variations in Temperatures, Rainfall, Relative humidity during the different stages (tillering, panicle initiation, booting and physiological maturity).

Graphical representation of fitted models using different technique

Various methods have been taken into the context for the graphical representations of best performing models in accordance with their higher R^2 value and standard error. The graph is plotted between actual yield and predicted yield using x-axis as year were y-axis as yield.



- Simple Weather-based method, Model-4 which shows best performance in accordance with R^2 (Table. 2). The Model-4 Shows R^2 of 0.999 & standard error of 12.846.
- Weighted & Unweighted method, Model-4 which shows best performance in accordance with R^2 (Table. 2). The Model-3 Shows R^2 of 0.978 & standard error of 61.280.

- PCA method, Model-4 which shows best performance in accordance with R^2 (Table. 2). The Model-4 Shows R^2 of 0.983 & standard error of 57.053.
- ANN method, Model-4 which shows best performance in accordance with R^2 (Table. 2). The Model-4 Shows R^2 of 0.664.

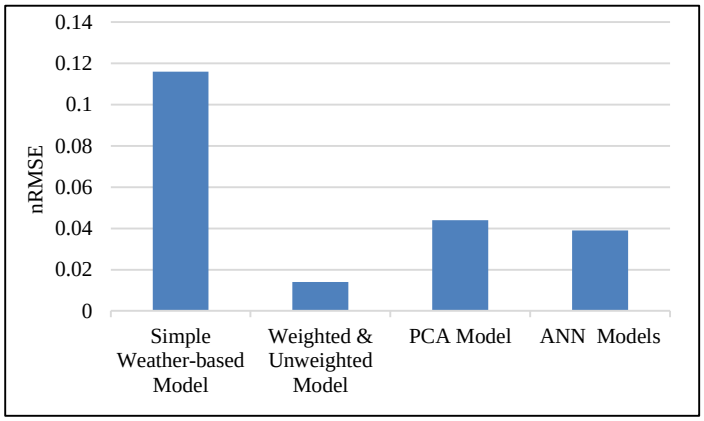
Performance Analysis of Different Methods

Among all the methods the best fitted models are taken for the forecast of rice crop yield is supported by Normalized Root Mean Squared Error as shown in Table.4. According to the performance of each method best fitted models has been chosen.

Different Methods	nRMSE
1. Simple Weather-based Model	0.116
2. Weighted & Unweighted Model	0.014
3. PCA Model	0.044
4. ANN Model	0.039

Table.4. Performance analysis of various method

Resulting as Simple Weather-based shows the best performance with least nRMSE value of 0.014 followed by ANN which is 0.039 and so on. Simple Weather-based is considered as the most accurate and reliable forecasting methods from other three methods used in the forecasting of rice crop yield.



IV. CONCLUSION

From all the above studies of developing yield prediction model of rice crop using various statistical methods and machine learning technique, and from all the above result, from the Table 4 the best fitted models for each technique can be easily conclude that Weighted and Unweighted Model with nRMSE value of 0.014 is best among the all models, were the second best suited model is ANN Model with nRMSE value of 0.039 , the third best suited model is PCA Model with nRMSE value of 0.044 & Weighted and Simple Weather-based Model secure the last position as nRMSE value is 0.116.

V. REFERENCES

1. Banakara, K. B. and Popat R. C., Amaresh and Pandya H. R. (2018). Pre-harvest forecast of Kharif Rice using weather parameters in Bharuch district of Gujarat state. Int. J. Chem. Stud., 6(6), 1111- 1116.
2. Bates, D., Mächler, M., Bolker, B. and Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.
3. Chauhan, V.S., Shekh, A.M., Dixit, S.K., Mishra, A.P. and Kumar,S. (2009). Yield prediction model of rice in Bulsar district of Gujarat. J. Agrometeorology., 11 (2): 162-168.
4. Diwan, U. K., Puranik, H. V., Das, G. K. and Chaudhary, J. L. (2018). Yield prediction of wheat at preharvest stage using regression based statistical model for 8 districts of Chhattisgarh, India. Int. J. Curr.Microbiol. Appl. Sci.,7(1):2180-2183,
5. Field, A. (2013). Discovering statistics using SPSS (4th ed.). Thousand Oaks, CA: Sage.
6. Ghosh, K., Balasubramanian, R., Bandopadhyay, S., Chattopadhyay, N., Singh, K. K. and Rathore, L. S. (2014). Development of crop yield forecast models under FASAL-a case study of kharif rice in West Bengal. J. Agrometeorol, 16(1) 1:8.
7. Kalubarme, M.H., and Ahuja, S.C. (1996). Agrometeorological rice yield modelling in Karnal. J. Indian Soc. Remote Sens., 24:125– 132.
8. Kumar N., Pisal R., Shukla, S. P. and Pandey, K. K. (2014). Crop yield forecasting of paddy, sugarcane and wheat through linear regression technique for south Gujarat. Mausam. 65: 361-364.
9. Latwal, A., Nain, A. S., Jha, A. and Ranjan, R. (2017). Wheat yield prediction using satellite data and meteorological parameters for Udham Singh Nagar district of Uttarakhand. Int. J. Basic Appl. Agril. Res., 15(3): 130-132.
10. Osborne, J. and Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. Pract. Assess. Res. Evaluation, 8(2): 1-9.