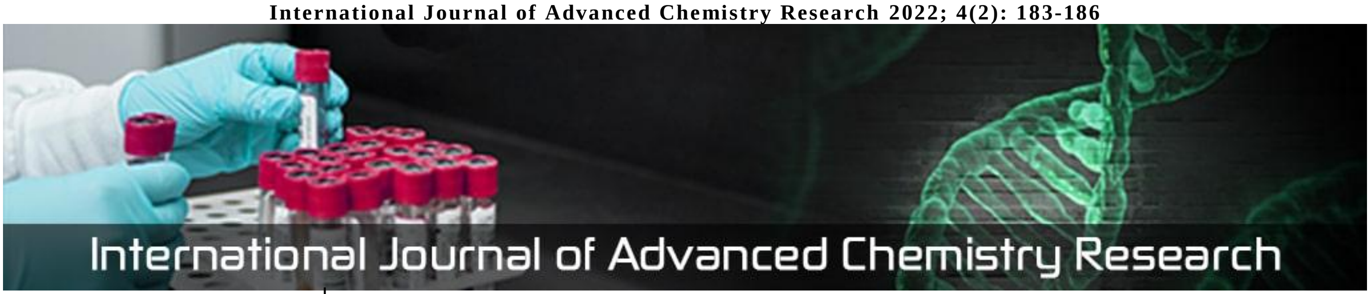

ISSN Print: 2664-6781
ISSN Online: 2664-679X
Impact Factor: RJIF 5.32
IJACR 2022; 4(2): 183-186
www.chemistryjournals.net
Received: 18-08-2022
Accepted: 23-09-2022

Verma G
Department of Agril Statistics, Indira Gandhi Agriculture University, Raipur, Chhattisgarh, India

Pandey KK
Department of Agril Statistics, Indira Gandhi Agriculture University, Raipur, Chhattisgarh, India

Corresponding Author:
Verma G
Department of Agril Statistics, Indira Gandhi Agriculture University, Raipur, Chhattisgarh, India

Development of yield forecast models with raw data through stepwise regression analysis (forward method) and discriminant function analysis for pigeon pea crop on Chhattisgarh plain zone

Verma G and Pandey KK

Abstract
Under this investigation we studied about minimization of weekly data through stratification and Simple Randon Sample technique. Individual effect of weather parameters has been considered under the study. Forecasting model developed on two statistical technique i.e. Stepwise Regression Analysis (Forward Method) technique and Discriminat Function Analysis technique. Both the techniques are found very suitable for the district wise and zone level pre-harvest forecast for the pigeonpea crop on Chhattisgarh. Models fitted with the Stepwise Regression Analysis on 35 variables and Time Trend (T) for Zone. The developed model is showing best fit on the basis of very high value of significance and maximum R^2 (84%) for the plain Zone, highly significant at 0.1% level of significance and models fitted with DFA then 2 Discriminant score (ds1 and ds2) and Time Trend (T) the models found highly significant and R^2 value 69% for plain zone for Pigeonpea crop. The developed model has been validated by the error parameters *viz.* Minimum MAE, Minimum MSE, Minimum RMSE, Minimum PE and Minimum PD along with maximum R^2. These developed models are useful to farmers of Chhattisgarh to decide their future prospects and possible course of action in advance. This was very challenging task for the researchers to develop a precise & accurate and best fit model for the future forecast. May this study will clear the fog through new methodology and also will create the new way on the direction of forecast.

Keywords: Pigeonpea, coefficient of determination, stepwise regression analysis and discriminant function analysis

1. Introduction

Forecasting is a technique that uses historical data as inputs to make informed estimates that are predictive in determining the direction of future trends. Forecast of the crop production at suitable stages of crop period before the harvest are vital for rural economy. Forecasts are used by the government & industry to protect life and property. The weather and climatic information plays a major role before and during the cropping season and if provided in advance it can be helpful in stirring the farmer to form and use their own resources in order to gather the benefits. The advance knowledge of weather parameters in a particular region is advantageous in effective planning.

The final estimates based on objective crop-cutting experiments are of limited utility as those become available quite later after the crop harvest. The Statistical techniques employed for forecasting purposes should be able to provide objective, consistent and comprehensible forecasts of crop yield with reasonable precisions well in advance before the harvest. Several studies have been carried out to forecast crop yield using weather parameters etc.

In the current situation of India faces increasing population and industrial development which are adversely distressing the crop yield in India. Keeping in mind early crop yield forecast will help farmer to formulate the cropping pattern, agricultural practices which will results in to the increase yield of the farmers. Therefore main objective of the present study was to develop different approaches for forecasting the rice yield before harvesting with help of weather parameters. The forecasting of crop yield may be done by using three major objective methods (i) biometrical characteristics (ii) weather variables and (iii) agricultural inputs (Agrawal, *et al.* 2001).

2. Material and Methods

2.1 Data on Yield

The yield data on Pigeonpea crop for 30 years (1990-91 to 2019-20) for Raipur, Bilaspur, Durg and Rajnadgaon Districts. 22 year (1998-99 to 2019-20) for Dhamtari, Mahasamund, Kawardha, Korba and Janjgir have been procured from the published booklets and official website (http://agridept.cg.gov.in/index.htm) of the Directorate of Agricultural, Govt. of Chhattisgarh to development of the statistical models at zone level.

2.2 Weather data

The weekly weather data (1990-91 to 2019-20) procured from the department of Agro meteorology, Indira Gandhi Agriculture University, Raipur (CG.) and used for the Chhattisgarh plain Zone for Pigeonpea crop. The data of 7 weather variables have been procured for 30^{th} weeks of the Pigeonpea crop cultivation period. Which include 27^{th} SMW from first year to 4^{th} SMW from 2^{nd} year. The variables are Maximum Temperature (Tmax) °C, Minimum Temperature (Tmin) °C, Morning Relative Humidity (RH-I)%, Afternoon Relative Humidity (RH-II)%, Bright Sunshine hours (SS) hours/day, rainfall (mm), Wind Velocity (WV) respectively.

2.3 Statistical Methodology

Crop yield is affected by technological change and weather variability. Technological change includes the impact of increased fertilizer applications, improved management practices, pest and diseases control, improved genetic qualities of seed and other technological factors design to increase yield.

However, in order to study the individual effect of weather variable and joint effect of pair of weather variables, the procedures laid down Agrawal *et.al.* 1986 have been applied. They expressed the effect of change in weather variables on yield in w^{th} week as a linear function of respective correlation coefficients between yield and weather variables. The trend effect is generally expected to consider on time series yield data, but its effect to be removed from yield known as trend effect because calculation of correlation coefficients of yield with weather variables to be used as weights.

2.3.1 Procedure of Stratification to minimize the weeks

Divide the population of N units into strata. Let the i^{th} stratum have N_i. I = 1, 2......., k number of units. The strata are constructed such that they are no overlapping and homogeneous with respect to the characteristic under study. Such that

$$\sum_{i=1}^{k} Ni = N$$

Draw a sample of size n from i^{th} (i=1, 2,......,k) stratum using simple random sampling independently from each stratum. All the sample units drawn from each stratum with constituent a stratified sample of size.

$$n = \sum_{i=1}^{k} ni$$

2.3.2 Effect of Individual Weather Variable

Total five weather variable selected from stratums. Now all five selected variables along with T has been used for individual effect. The model for individual effect is given as follows.

$$Y = bo + b_1S_1P_1w_k + b_2S_2P_2w_k + \ldots\ldots + b_5 S_i P_j w_k + b_6 T$$

Where;
Y = Crop Yield
Bo = Intercept of the model.
B = Regression coefficient.
S_i ($i=1, 2..... 5$) = Number of stratification.
P_j ($i=1, 2..... 7$) = Weather parameters.
w_k ($k=1,2,.....n$) = Number of week for each weather parameter.
T = trend variable (crop-years under consideration).

2.3.3 Forecast Model based on Weather Variables through Stepwise Regression Analysis (Forward Method)

This is new proposed method with all five variables selected for each weather parameters. Total number of selected variables for one district is 35.

$$Y = a + \beta_i S_i P_j w_k$$

Where;
Y = Crop yield.
a = Intercept of the model.
β_i ($i=0,1,2.....n$) = Regression coefficient selected week of respective parameters from particular stratum.
S_i ($i=1, 2..... 5$) = Number of stratification.
P_j ($i=1, 2..... 7$) = Weather parameter
w_k ($j=1, 2..... n$) = Number of week for each weather parameter.

2.3.4 Procedure to develop forecasting model through Discriminant Function Analysis

The data is divided in to three groups *viz.* Congenial, Average and Adverse on the basis of productivity of Pigeonpea for the given time series data. In this procedure discriminant scores have been computed using the un weighted and weighted averages (weather indices) for the first weather variable (Here, discriminating factors will be only two). This process is continued up to seventh weather variables, and finally we get two discriminant scores ds1, and ds2. Using crop-yield as dependent variable and discriminant scores ds1, ds2 and the time trend T as independent variable, the following model is fitted for the development of forecast model.

Model: $y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T$

Where;
y = Crop yield.
βo = Intercept of the model.
β_i ($i=1,2,3$) = The regression coefficients.
ds_1 and ds_2 are the two discriminant scores.
T = trend variable ($T= 1, 2\ldots.n$); n is crop-years under consideration.

2.3.4 Validation and comparison of the models

Following procedures have been used for the comparison and the validation of the developed models. These procedures are given bellow.

3. Percent Deviation

The formula for computation of Percent Deviation of forecast yield from actual yield is given by Md. Azfar *et al.* (2015) [2]. This measures the deviation (in percentage) of forecast from the actual yield data. The formula for calculating the percent deviation of forecast is given below:

$$\text{Percent deviation} = \frac{(\text{Actual Yield} - \text{Forecasted Yield})}{(\text{Actual yield})} x100$$

Root Mean – Square Error (RMSE): Root Mean Square Error (RMSE):- RMSE indicate the magnitude of the average error, but provide no information on the relative size of the average difference between (F) and (O). This method is most popular and important for the validation of the model. The formula for calculating the RMSE is given by Varshneya *et al. (*2010) of forecast is given below:

$$RMSE = \left[\frac{\sum_{i=1}^{N}(F_i - O_i)^2}{n}\right]^{1/2}$$

Where;
I = 1, 2, 3...N
N = Sample size.
F_i = Actual yield.
O_i = Predicted yield.

Mean Absolute Error (MAE)

Mean absolute error represents the difference between the original and predicted values extracted by averaged the absolute difference over the data set Kotz *et al.* (2006).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x|$$

Where;
N =the number of error;
$|x_i - x|$ = the absolute error

Mean Squared Error (MSE)

Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and predicted values. When a model has no error, the MSE equals zero. Error increase, its value increases. The mean squared error also known as the mean squared deviation Kotz *et al.* (2006).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2$$

Where;
N = number of data point
Y_i = observed value
$\hat{Y}$ = predicted value

Percent Error (PE): PE is defined as ratio of RMSE to mean observed value expressed as percentage. This is another important tool of the model validation. The formula for calculating the Percent Error is given by Varshneya *et.al.* (2010) of forecast is given below:

$$PE = \frac{RMSE}{\bar{o}} \times 100$$

Where;
PE = Percent error
RMSE = Root mean square error
$\bar{o}$ = Mean of observed value

Coefficient of determination (R^2): The models were validated on the basis of (R^2) which can be computed from the formula given by Draper and Smith (1988). It is in general used for checking the adequacy of the model. R^2 is given by the following formula

$$R^2 = 1 - \frac{ss_{res}}{ss_t}$$

Where ss_{res} and ss_t are the residual sum of square and the total sum of square respectively.

R^2 never decreases when a regressed is added to the model, regardless of the value of the contribution of the variable in the model. Therefore, it is difficult to judge whether an increase in R^2 is really important.

4. Results and Discussion

Model developed for Zone on Pigeon pea through Stepwise Regression Analysis

The multiple Regression model has been obtained:

$Y = 1517.82 - 44.61\ S_3P_4w_{18} - 14.94\ S_3P_4w_{17} - 27.52\ S_4P_6w_{26} - 34.60T$

Table 1: Model developed for zone on pigeon pea through stepwise regression analysis for zone

District	Constant	Entered Variable	Coefficient	P-Value	Std. Error	R^2	Significance
Zone	1517.82	$S_3P_5w_{18}$	-44.61**	0.009	15.32	0.84***	0.001
		$S_3P_6w_{17}$	-14.94***	0.001	3.31		
		$S_4P_7w_{26}$	-27.52**	0.005	8.64		
		T	-34.60***	0.001	3.63		

*** p<0.001, ** p<0.01,* p<0.05, + p<0.1

Table-1 shows The Stepwise Regression Analysis technique has been used for the selection of significant variables. Only $S_3P_5w_{18}$, $S_3P_6w_{17}$, $S_4P_7w_{26}$, along with T (Time trend) only four variable entered in this model for Chhattisgarh plain zone. The coefficients of entered variable are given in table; showing significant. The value of R^2 found i.e. 84% which is highly significant at 0.1% level of significance and the model is also found highly significant.

Table 2: Validation and comparison of developed model by stepwise regression analysis for zone

	Year	Actual Yield	Predicted Yield	MAE	MSE	RMSE	PE	PD	R^2	Significance
Zone	2015	1029.43	387.03	706.14	519347	720.66	37.52	62.40	0.84***	0.001
	2016	1291.31	463.37					64.12		
	2017	872.09	268.47					69.22		
	2018	952.93	415.81					56.36		
	2019	1305.59	385.96					70.44		

***$p < 0.001$, ** $p < 0.01$,* $p < 0.05$, + $p < 0.1$

The value of R^2 found i.e. 84% which is highly significant at 0.1% level of significance. The value of MAE (706.14), PE (37.52), MSE (519347), RMSE (720.66). The validation of model minimum PD show for 2018 years followed by 2015 and maximum 2019 followed by 2017, the predicted yield of different years have been shown on table-2.

Model Developed for Zone on Pigeon pea through Discriminant Function Analysis

The multiple regression equation fitted two discriminant score and time trend T is

$$Y = 948.65 + 120.06\ ds1 + 35.55\ ds2 - 14.77\ T$$

Table 3: Model developed for zone on pigeon pea by discriminant function analysis

	Variable	Coefficient	Standard Error	P-Value	R^2	Significance
Zone	Intercept	948.65	126.60		0.69 ***	0.001
	ds1	120.06***	30.45	0.001		
	ds2	35.55	36.88	0.344		
	T	-14.77^{+}	7.68	0.066		

***$p<0.001$, **$p<0.01$,*$p<0.05$, + $p<0.1$

Table 4: Validation and comparison of developed model by discriminant function analysis for zone

	Year	Actual Yield	Predicted Yield	MAE	MSE	RMSE	PE	PD	R^2	Significance
Zone	2015	1029.43	1174.12	183.38	57649.85	240.10	4.94	-14.06	0.69***	0.001
	2016	1291.31	1013.09					21.55		
	2017	872.09	892.74					-2.37		
	2018	952.92	913.08					4.18		
	2019	1305.59	872.13					33.20		

***$p<0.001$, **$p<0.01$,*$p<0.05$, + $p<0.1$

The above model has been developed through discriminant scores for Chhattisgarh plain zone. Coefficient of the score ds2 is showing non-significant and score ds1 & T are showing significant at 0.1% & 10% level respectively. The value of R^2 is 69% at 0.1% level (Table-3) and the model is also found highly significant.

The result of the validation of model (Table-4) developed by Discriminant Function Analysis for Zone. The values given in table are showing the status of model the value of Percent Deviation (PD) is well enough to say the model is very efficient for forecast. The value of PE is 4.94 indicates that the model is most efficient. The value of R^2 (0.69) is highly significant at 0.1% level of significance.

5. References

1. Anderson TW. An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York; c1984.
2. Azfar Mohd, Sisodia BVS, Rai VN, Devi Monika. Pre-harvest forecasting models for rapeseed and mustard yield using principle component analysis of weather variable. Mausam; c2015. p. 761-766.
3. Agrawal R, Jain RC, Jha MP. Models for studying rice crop weather relationship. Mausam. 1986;37(1):67-7.
4. Draper NR, Smith H. Applied Regression Analysis. 3rd edition, John Wiley & Sons Inc.; c1988.
5. Agrawal R, Jain RC, Mehta SC. Yield forecast based on weather variables and agricultural input on agrioclimatic zone basis. Ind. J. Agric. Sci. 2001;71(7):487-90.
6. Pandey KK, Rai VN, Sisodia BVS, Bharti AK, Gairola KC. Pre-harvest forecast models based on weather variable and weather indices for eastern U.P. Advances in Bioresearch; c2013. p. 118-122.
7. Pandey KK, Rai VN, Sisodia BVS, Singh SK. Effect of weather variables on rice crop in eastern Uttar Pradesh, india. Plant archives; c2015. p. 575-579
8. Kumar Neeraj, Pisal RR, Shukla SP, Pandey KK. Crop yield forecasting of paddy and sugarcane through modified hendrick and scholl technique for south Gujarat. Mausam; c2016. p. 405-410.
9. Pandey KK, D Maurya, Gupta G, Mishra SV. Yield forecasting models based on weather parameters for eastern U.P. Vegetos, An International Journal of Plant Research; c2016.
10. Pandey KK, Rai VN, Sisodia BVS, Pandey R, Pandey V. Development of forecasting model on weather parameters for eastern U.P.. Environ. Life Sci. Rel. Sci; c2016. p. 1370-1372.
11. Pandey KK, Chandrakar CK, Singh SK, Maurya D, Gupta G. Identification of most important weeks on minimum temperature for wheat crop, International Journal of Current Microbiology and Applied Sciences; c2017. p. 788-794.
12. Wei B, Varshney A, Patwari N, Hu W, Voigt T, Chou CT. April. DRTI: Directional radio tomographic imaging. In Proceedings of the 14th International Conference on Information Processing in Sensor Networks; c2015. p. 166-177.
13. Abu-Ellail FF, Hussein EM, El-Bakry A. Integrated selection criteria in sugarcane breeding programs using discriminant function analysis. Bulletin of the National Research Centre. 2020;44(1):1-14.
14. Smith AM, Blackshaw RE. Weed–crop discrimination using remote sensing: A detached leaf experiment. Weed Technology. 2003;17(4):811-820.