International Journal of Statistics and Applied Mathematics 2023; SP-8(5): 654-664

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(5): 654-664

https://www.mathsjournal.com
Received: 05-06-2023
Accepted: 08-07-2023

Neeraj Singh
Assistant Professor, Agricultural Statistics, Kamla Nehru Institute of Physical and Social Sciences, Sultanpur, Uttar Pradesh, India

Boya Venkatesu
Assistant Professor, School of Business, Woxsen University, Hyderabad, Telangana, India

Anushree Shukla
Assistant Professor, Department of Mathematics and Statistics, SHUATS, Naini, Prayagraj, Uttar Pradesh, India

Ume Kulsum
Research Scholar, Division of Agricultural Statistics, SKUAST – Kashmir, Jammu and Kashmir, India

Corresponding Author:
Neeraj Singh
Assistant Professor, Agricultural Statistics, Kamla Nehru Institute of Physical and Social Sciences, Sultanpur, Uttar Pradesh, India

Meteorological parameters-based pre-harvest forecasting of wheat crop yield by using discriminant function analysis

Neeraj Singh, Boya Venkatesu, Anushree Shukla and Ume Kulsum

Abstract
In the present paper, an application of discriminant function analysis on meteorological parameters for developing suitable statistical models to forecast pre-harvest wheat yield in Azamagrh district of Eastern Uttar Pradesh has been demonstrated. Time series data on wheat yield for 18 years (2000-01 to 2017-18) have been divided into three groups, *viz*. congenial, normal, and adverse based on de-trended yield distribution. Considering these groups as three populations, discriminant function analysis using weekly data of crop season on five meteorological parameters has been carried out. The discriminant scores obtained from this have been used as regressor variables along with time trend in development of statistical models. In all six procedures using weekly weather data have been proposed. The models developed have been used to forecast the wheat yield for the year 2015-16 and 2017-18 which were not included in the development of the models. It has been found that most of the models provide reliable forecast of the wheat yield about two months before the harvest. However, the model -D_5 has been found to be the most suitable among all the models developed.

Keywords: Meteorological parameters, Crop yield, Discriminant function analysis, Pre-harvest Forecast model

Introduction
Forecast of the crop production at suitable stages of crop period before the harvest are vital for rural economy and important for advance planning, formulation and its implementation in regards to crop procurement, distribution, price structure and import/export decisions etc. It is useful to farmers to decide in advance their future prospects and course of action. Various research workers have made efforts in the past to develop statistical models based on time series data on crop-yield and weather variables for pre-harvest forecasting of crop yield. Rai and Chandrahas (2000) [8] made use of discriminant function analysis of weather variables to develop statistical models for pre-harvest forecasting of rice-yield in Raipur district of Chhattisgarh have recently developed forecast models for wheat yield in Kanpur district (U.P.) using discriminant functions analysis of weakly data on weather variables. Since the discriminant function analysis discriminates best between sets of observations from two or more groups and classify the future observations into one of the previously defined groups, an attempt has been made in the present paper to develop suitable statistical models for forecasting of pre-harvest wheat crop yield in Azamgarh district of Uttar Pradesh using discriminant functions analysis of weekly data on weather variables.

Materials and Methods
The present study is related to Azamgarh district (Eastern Uttar Pradesh, India) which falls under middle gangetic agro-climatic region and 8th eastern plain agro-climatic zone in the eastern part of Uttar Pradesh at a distance of about 270 km from Lucknow. Azamgarh district lies between latitude 26° 03’ N and longitude 83° 13’ E and it is bounded by the districts of Mau in east, Gorakhpur in the north, Ghazipur in south-east, Jaunpur in the south-west, Sultanpur in the west and Ambedkar Nagar in the north-west. The district consists of 22 developmental blocks located between Tamsa, Choti Saryu and Ghaghara rivers.

The total geographical area of the district is 4234 sq. km. covering 110 sq. km forest area supporting a population of 46.13 lakhs with a density of 1138 persons per sq km. The net sown area of the district is 3.03 lakh ha with total cropped area of 4.24 lakh ha. The area under cultivation is 71% with cropping intensity at 167%. Maximum temperature goes up to 43.5 °C in summer and minimum temp. 6.8 °C goes in winter. The soil is sandy loam new and deep.

The average annual rainfall is 1031 mm. Rice and wheat are the major crops of the district; the other crops grown in the district are maize, barley, pulses, oilseeds and potato. Sugarcane is the main cash crop of the district. The Sathiaon Sugar Mill was established in 1975 and situated at a distance of about 13 km. from district headquarter Azamgarh on the road connecting to Azamgarh-Mau. The study has been conducted for rice and wheat crop which is the principal food grain crop of the kharif and rabi season in Azamgarh district.

The objective is to develop pre-harvest forecast model for wheat yield. The time series data on yield for rice and wheat crop of Azamgarh district of eastern Uttar Pradesh pertaining for the period from 2000-01 to 2017-18 have been procured from the website http://updes.up.nic.in/spatrika/spatrika.htm by Economics and Statistics Division, Planning Department, Government of Uttar Pradesh.

Weekly weather variables data for rice and wheat crop in the district of Azamgarh, Eastern Uttar Pradesh have been obtained from the National Data Centre, India Meteorological Department, Pune for the study period 2000-01 to 2017-18. The data on five weather variables *viz*. Maximum Temperature, Minimum Temperature, Rainfall, wind-velocity and Sun-shine hours have been used in the study.

Statistical Methodology

The technique of discriminant function analysis is used to identify an appropriate function that discriminates best between sets of observations from two or more groups and classifying the future observations into one of the previously defined groups. Consider that observations are classified into k non-overlapping groups on the basis p variables. The technique identifies linear functions where the coefficients of the variables are determined in such a way that the variation between the groups gets maximized relative to the variation within the groups. The maximum number of discriminant functions that can be obtained is equal to minimum of (k-1) and p. These functions are used to calculate discriminant scores, which are used to classify the observations into different groups developed forecast models for wheat yield in Kanpur district of Uttar Pradesh using discriminant function analysis technique that provided reliable yield forecast about two months before harvest. This paper applies the technique used by them along with a few modifications for the development of suitable models for pre-harvest forecast of wheat crop yield in Azamgarh district of Uttar Pradesh.

In order to apply discriminant function analysis for modeling yield using weather variables, crop years under consideration have been divided into three groups, namely adverse, normal and congenial on the basis of crop yield adjusted for trend effect. Data on weather variables in these three groups were used to develop linear discriminant functions and the discriminant scores were obtained for each year. These discriminant scores were used along with year index (trend variable) as regressors and crop yield as regressand in developing the forecast models. In the present study the number of groups is three and number of weather variables is five, therefore only two discriminant functions can be obtained which are sufficient for discriminating a crop year into either of the three groups.

Three groups of crop years, *viz*. adverse, normal and congenial have been obtained as follows: Let $\bar{y}$ and s be the mean and standard deviation of the adjusted crop yields of n years. The adjusted crop yields less than or equal to $\bar{y}-s$ would form adverse group, the adjusted crop yields between $\bar{y}-s$ and $\bar{y}+s$ would form normal group and adjusted crop yields above or equal to $\bar{y}+s$ would form congenial group. The adjusted crop yields were assigned codes 1, 2 and 3 if they belong to adverse, normal and congenial groups, respectively.

It is, however, known that weather variables affect the crop differently during different phases of crop development. Its effect depends not only on its magnitude but also on its distribution pattern over the crop season. Therefore, using weekly weather data as such in developing the model poses a problem as number of independent variables in the regression model would increase enormously. To solve this problem, following weather indices have been developed using the procedure of Agrawal *et al*. (1983, 1986) [2, 3].

$$Z_{ij} = \sum_{w=1}^{n} r_{iw}^{j} X_{iw} \Big/ \sum_{w=1}^{n} r_{iw}^{j}$$

$$Z_{ii',j} = \sum_{w=1}^{n} r_{ii'w}^{j} X_{iw} X_{i'w} \Big/ \sum_{w=1}^{n} r_{ii'w}^{j},$$

j=0,1 and i=1,2,…,p. (3.3.3.5)

where

Z_{ij} is un-weighted (for j = 0) and weighted (for j = 1) weather indices for i^{th} weather variable and $Z_{ii',j}$ is the un-weighted (for j = 0) and weighted (for j = 1) weather indices for interaction between i^{th} and i'^{th} weather variables. X_{iw} is the value of the i^{th} weather variable in w^{th} week, $r_{iw}/r_{ii'w}$ is correlation coefficient of yield adjusted for trend effect with i^{th} weather variable/product of i^{th} and i'^{th} weather variable in w^{th} week, n is the number of weeks considered in developing the indices and p is number of weather variables. Standard Metrological Weeks (SMW) data have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions. In all 30 indices (15 weighted and 15 un-weighted) consisting of 5 weighted and 10 weighted interaction weather indices and 5 un-weighted and 10 un-weighted interaction weather indices have been constructed. Here only the first 15 years data from 2000-01 to 2014-15 have been utilized for model fitting and remaining three years were left for the validation of the model. For quantitative forecasting, linear regression models are fitted by taking the discriminant scores and the trend variable as the regressors and crop yield as the regressand. The following models are considered.

Model-D_1

In this procedure, 5 un-weighted weather indices have been used as discriminating variables. Now, based on these 5 indices, the discriminant function analysis has been done and two sets of scores have been obtained. On the basis of these two sets of scores, the regression model has been fitted taking the yield as the regressand and the two sets of scores and the trend variable (T) as the regressors. The model fitted here is

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

Where
y = crop yield
βo = intercept of the model
$\beta_{i's}$ $(i=1,2,3)$ = the regression coefficients
ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2, 3, …. n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D$_2$
In this procedure, 5 weighted weather indices have been used as discriminating variables. Now, based on these 5 indices, the discriminant function analysis has been done and two sets of scores have been obtained. On the basis of these two sets of scores, the regression model has been fitted taking the yield as the regressand and the two sets of scores and the trend variable (T) as the regressors. The model fitted here is

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

Where
y = crop yield
βo = intercept of the model
$\beta_{i's}$ $(i=1,2,3)$ = the regression coefficients
ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3, ……, n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D$_3$
In this procedure, all 30 (weighted and un-weighted including interaction indices) have been used as discriminating variables in discriminant function analysis and two sets of discriminant scores from two discriminant functions have been obtained. Forecasting model has been fitted taking un-trended yield as the regressand variable and the two sets of discriminant scores and the trend variable (T) as the regressor variables. The form of the model fitted is as follows:

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

Where
y = crop yield
βo = intercept of the model
$\beta_{i's}$ $(i=1,2,3)$ = the regression coefficients
ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3, ……, n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D$_4$
In this procedure, 5 weighted and 5 un-weighted weather indices have been used as discriminating variables. Now, based on these 10 indices, the discriminant function analysis has been done and two sets of scores have been obtained. On the basis of these two sets of scores, the regression model has been fitted taking the yield as the regressand and the two sets of scores and the trend variable (T) as the regressors. The model fitted here is

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

Where
y= crop yield
βo = intercept of the model
$\beta_{i's}$ $(i=1,2,3)$ = the regression coefficients
ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3, ……, n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D$_5$
In this procedure, 5 un- weighted and 10 un-weighted interaction weather indices have been used as discriminating variables. Now, based on these 15 indices, the discriminant function analysis has been done and two sets of scores have been obtained. On the basis of these two sets of scores, the regression model has been fitted taking the yield as the regressand and the two sets of scores and the trend variable (T) as the regressors. The model fitted here is

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

Where
y= crop yield
βo = intercept of the model
$\beta_{i's}$ $(i=1,2,3)$ = the regression coefficients
ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3, ……, n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Model-D$_6$
In this procedure, 5 weighted and 10 weighted weather indices have been used as discriminating variables. Now, based on these 15 indices, the discriminant function analysis has been done and two sets of scores have been obtained. On the basis of these two sets of scores, the regression model has been fitted taking the yield as the regressand and the two sets of scores and the trend variable (T) as the regressors. The model fitted here is

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

Where
y= crop yield
βo = intercept of the model
$\beta_{i's}$ $(i=1,2,3)$ = the regression coefficients
ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3, ……, n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Comparison and validation of forecast models
Different procedures have been used in the present study for the comparison and the validation of the models developed. These procedures are given below. The six models were compared on the basis of adjusted coefficient of determination

(R_{adj}^2), the percent deviation of forecast from actual, and percent standard error (SE).

Results and Discussion

In order to carry out discriminant function analysis, the wheat yields are adjusted for trend effect. The crop years have been divided into three groups namely adverse, normal and congenial. The actual wheat yields, adjusted rice yield and the groups indicated by 1, 2, and 3 as adverse, normal and congenial, respectively, are given in the Table 1.

Table 1: Actual and adjusted yield of wheat crop

Year	Actual yield (Q/ha)	Adjusted yield (Q/ha)	Groups
2000-01	22.36	21.80	2
2001-02	23.86	22.73	2
2002-03	20.03	18.34	2
2003-04	24.89	22.63	2
2004-05	21.18	18.36	2
2005-06	23.98	20.59	2
2006-07	24.56	20.61	2
2007-08	25.94	21.43	2
2008-09	27.24	22.16	2
2009-10	25.84	20.20	2
2010-11	28.00	21.79	2
2011-12	28.90	22.13	2
2012-13	29.42	22.08	2
2013-14	29.72	21.82	2
2014-15	21.38	12.92	1
2015-16	27.27	18.24	1
2016-17	33.54	23.95	3
2017-18	34.40	24.24	3

The models have been developed by utilizing 15 years data of wheat yield (2000-01 to 2014-15) and remaining three years were left for the validation of the model. The models developed are described below:

Model-D_1

The discriminant function analysis was carried out to find out the discriminant functions and discriminant scores, this results two discriminant functions and two set discriminant scores. The estimated discriminant functions, discriminant scores and other relevant results are given in Appendix – (a). The classification results based on estimated discriminant functions are given in Table 2.

Table 2: Classification Results

		CAT	Predicted Group Membership			Total
			1	2	3	
Original	Count	1.00	1	1	0	2
		2.00	0	14	0	14
		3.00	0	0	2	2
	%	1.00	50.0	50.0	.0	100.0
		2.00	.0	100.0	.0	100.0
		3.00	.0	.0	100.0	100.0

94.4% of original grouped cases correctly classified.

The discriminant score ds_1 and ds_2 and time trend (T) were considered as regressor variables and actual wheat yield as regressand for fitting the regression model (Model-D_1). The results of the fitted model are presented in the Table 3.

Table 3: Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	R^2 %
1	Constant	20.632	1.837	42.8
2	ds_1	0.625	0.753	
3	ds_2	- 0.476	0.689	
4	Trend	0.542	0.196	

* $p < 0.05$, ** $p < 0.1$, DB Statistic: 1.894

The Durbin-Watson result is non-significant according to tabulated value for n=15 and k=3.

Forecast model- D_1

$$Y = 20.632 + 0.625\, ds_1 - 0.476\, ds_2 + 0.542\, T$$

The model is validated by forecasting the wheat yield for the year 2015-16 to 2017-18. The results of validation are given below in the Table 4.

Table 4: Validation of the model – D_1

Year	Actual yield	Predicted yield	R^2 %	Percent deviation	RMSE
2015-16	27.27	28.70	42.8	5.24	4.13
2016-17	33.54	28.46		- 15.15	
2017-18	34.40	29.56		- 14.07	

** $p < 0.1$

Model-D_2

Following the procedure in Model-D_2 as described, the discriminant function analysis was carried out. The estimated discriminant functions, discriminant scores and other relevant results are given in Appendix – (b). The classification results based on estimated discriminant function are presented in the Table 5.

Tale 5: Classification Results

		CAT	Predicted Group Membership			Total
			1	2	3	
Original	Count	1.00	2	0	0	2
		2.00	1	13	0	14
		3.00	0	1	1	2
	%	1.00	100.0	.0	.0	100.0
		2.00	7.1	92.9	.0	100.0
		3.00	.0	50.0	50.0	100.0

88.9% of original grouped cases correctly classified.

The discriminant scores ds_1 and ds_2 and time trend (T) were considered as regressor variables and actual wheat yield as regress and for fitting the forecast model. The results of the fitted model are presented in the Table 6.

Table 6: Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	R^2 %
1	Constant	20.474	1.097	70.90
2	ds_1	1.502	0.421	
3	ds_2	0.012	0.489	
4	Trend	0.559	0.119	

* $p < 0.05$, ** $p < 0.1$, DB Statistic: 1.848

The Durbin-Watson result is non-significant according to tabulated value for n=15 and k=3.

Forecast model- D_2

$Y = 20.474 + 1.502\ ds_1 + 0.012\ ds_2 + 0.559\ T$

The wheat yields for the year 2015-16 to 2017-18 were forecasted by the fitted model for its validation. The results are presented in the Table 7.

Table 7: Validation of the model- D_2

Year	Actual yield	Predicted yield	R^2 %	Percent deviation	RMSE
2015-16	27.27	25.72	70.90	- 5.68	2.94
2016-17	33.54	30.45		- 9.21	
2017-18	34.40	30.65		- 10.90	

** $p<0.1$

Model-D_3

Following the procedure in Model-D_3 as described, discriminant function analysis was carried out to find out discriminant functions and discriminant scores for each year under consideration. The estimated two discriminant functions, two sets of discriminant scores (ds_1 and ds_2) and other relevant results are given in Appendix – (c). The classification results based on the estimated discriminant functions are presented in the Table 8.

Table 8: Classification Results

		CAT	Predicted Group Membership			Total
			1	2	3	
Original	Count	1.00	2	0	0	2
		2.00	0	14	0	14
		3.00	0	0	2	2
	%	1.00	100.0	.0	.0	100.0
		2.00	.0	100.0	.0	100.0
		3.00	.0	.0	100.0	100.0

100.0% of original grouped cases correctly classified.

Forecast model was obtained by fitting regression model where discriminant scores ds_1 and ds_2 and time trend (T) were considered as regressor variables and actual wheat yield as regressand. The results of the fitted model are presented in the Table 9.

Table 9: Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	R^2 %
1	Constant	19.822	1.238	80.1
2	ds_1	0.330	0.250	
3	ds_2	0.719	0.492	
4	Trend	0.597	0.102	

* $p< 0.05$, ** $p<0.1$, DB Statistic: 1.844

The Durbin-Watson result is non-significant according to tabulated value for n=15 and k=3

Forecast model- D_3

$Y = 19.822 + 0.330\ ds_1 - 0.719\ ds_2 + 0.597\ T$

The Model-D_3 has been validated by forecasting the wheat yield for the year 2015-16 to 2017-18. The results are presented in the Table 10.

Table 10: Validation of the model- D_3

Year	Actual yield	Predicted yield	R^2 %	Percent deviation	RMSE
2015-16	27.27	30.86	80.1	13.16	9.33
2016-17	33.54	21.22		- 36.73	
2017-18	34.40	24.57		- 28.58	

** $p<0.1$

Model-D_4

Discriminant function analysis was carried out following the Model-D_4 described. Two discriminant functions were obtained and using these functions two sets of discriminant score were computed for each year under consideration (2000-01 to 2014-2015). The estimated discriminant functions, discriminant scores and other relevant results are given in Appendix – (d). The classification results obtained by estimated discriminant function are presented in the Table 11.

Table 11: Classification Results

		CAT	Predicted Group Membership			Total
			1	2	3	
Original	Count	1.00	2	0	0	2
		2.00	0	14	0	14
		3.00	0	0	2	2
	%	1.00	100.0	.0	.0	100.0
		2.00	.0	100.0	.0	100.0
		3.00	.0	.0	100.0	100.0

100% of original grouped cases correctly classified.

The forecast model was obtained by fitting regression model using discriminant scores ds_1 and ds_2 and time trend as regressor variables and actual wheat yield as regressand. The results of the fitted model are presented in the Table 12.

Table 12: Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	R^2 %
1	Constant	19.817	1.071	79.4
2	ds_1	1.299	0.310	
3	ds_2	- 0.516	0.431	
4	Trend	0.627	0.112	

* $p< 0.05$, ** $p<0.1$, DB Statistic: 1.553

The Durbin-Watson result is inconcludable according to tabulated value for n=15 and k=3.

Forecast model- D_4

$Y = 19.817 + 1.299\ ds_1 – 0.516\ ds_2 + 0.112\ T$

In order to validate the model, wheat yields for the year 2000-01 to 2014-2015 were forecasted using the Model-D_4. The results are presented in the Table 13.

Table 13: Validation of the model- D_4

Year	Actual yield	Predicted yield	R^2 %	Percent deviation	RMSE
2015-16	27.27	17.06	79.4	- 37.44	11.66
2016-17	33.54	21.64		- 35.48	
2017-18	34.40	21.66		- 37.03	

** $p<0.1$

Model-D_5

The discriminant functions analysis has been carried out using the procedure in Model-D_5 as described. Two estimated discriminant functions, two sets of discriminant scores and other relevant results are given in Appendix – (e). The classification results based on the estimated discriminant functions are presented in Table 14.

Table 14: Classification Results

		CAT	Predicted Group Membership			Total
			1	2	3	
Original	**Count**	1.00	2	0	0	2
		2.00	0	14	0	14
		3.00	0	0	2	2
	%	1.00	100.0	.0	.0	100.0
		2.00	.0	100.0	.0	100.0
		3.00	.0	.0	100.0	100.0

100.0% of original grouped cases correctly classified.

The regression model was fitted with discriminant scores ds_1 & ds_2 and time trend (T) as regressor variable and actual wheat yield as regressand. The results of the fitted model are presented in the Table 15.

Table 15: Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	R^2 %
1	Constant	20.060	1.167	80.9
2	ds_1	0.372	0.076	
3	ds_2	0.455	0.496	
4	Trend	0.606	0.113	

* *p*< 0.05, ** *p*<0.1, DB Statistic: 1.482

The Durbin-Watson result is inconcludable according to tabulated value for n=15 and k=3.

Forecast model- D_5

$$Y = 20.060 + 0.372\, ds_1 + 0.455\, ds_2 + 0.606\, T$$

The results of validation of the Model –D_5 are presented in the Table 16.

Table 16: Validation of the model- D_5

Year	Actual yield	Predicted yield	R^2 %	Percent deviation	RMSE
2015-16	27.27	21.56		- 20.94	
2016-17	33.54	33.59	80.9	0.15	3.51
2017-18	34.40	32.32		- 6.05	

** *p*<0.1

Model-D_6

Two discriminant functions and two sets of discriminant scores (ds_1 & ds_2) have been obtained by carrying out discriminant function analysis using the procedure in Model-D_6 as described. The estimated discriminant functions, discriminant scores and other relevant results are given in Appendix – (f). The classification results based on the estimated discriminant functions are presented in the Table 17.

Table 17: Classification Results

		CAT	Predicted Group Membership			Total
			1	2	3	
Original	Count	1.00	2	0	0	2
		2.00	0	14	0	14
		3.00	0	0	2	2
	%	1.00	100.0	.0	.0	100.0
		2.00	.0	100.0	.0	100.0
		3.00	.0	.0	100.0	100.0

100.0% of original grouped cases correctly classified.

The forecast model was obtained by fitting the regression model where discriminant scores ds_1 & ds_2 and time trend (T) were considered as regressor variables and actual wheat yield as regressed. The result of the fitted model is presented in the Table 18.

Table 18: Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	R^2 %
1	Constant	19.657	1.082	76.1
2	ds_1	0.573	0.164	
3	ds_2	0.116	0.500	
4	Trend	0.606	0.112	

* *p*< 0.05, ** *p*<0.1, DB Statistic: 1.610

The Durbin-Watson result is inconcludable according to tabulated value for n=19 and k=3.

Forecast model- D_6

$$Y = 19.657 + 0.573\, ds_1 + 0.116\, ds_2 + 0.606\, T$$

The forecast Model-D_6 has been validated by forecasting the wheat yield for the year 2015-16 to 2017-18. The results are presented in the Table 19.

Table 19: Validation of the model-D_6

Year	Actual yield	Predicted yield	R^2 %	Percent deviation	RMSE
2015-16	27.27	23.34		- 14.41	
2016-17	33.54	27.70	76.1	- 17.41	5.07
2017-18	34.40	29.16		- 15.23	

** *p*<0.1

Conclusion

The best three models for wheat crop obtained by the application of discriminant and principal component analysis of weekly weather data have been given below:

Forecast model - D_3

$$Y = 19.822 + 0.330\, ds_1 - 0.719 ds_2 + 0.597\, T$$

Forecast model - D_4

$$Y = 19.817 + 1.299 ds_1 - 0.516 ds_2 + 0.112 T$$

Forecast model - D_5

$$Y = 20.060 + 0.372 ds_1 + 0.455 ds_2 + 0.606 T$$

The forecast yields for the years 2015-16 to 2017-18 obtained from the aforesaid models, actual yield and various statistical measures for validation and comparison of the models are presented in the Table 20.

The Table 20 represents the forecast yields by proposed technique from the three best models for the year 2015-16 to 2017-18 along with corresponding actual yield of wheat crop. It may be pointed out here that the Model-D_6 was found best for wheat yield in Faizabad district of Uttar Pradesh (Sisodia *et al.*, 2014) [10].

Finally, the model-D_5 and D_3 based on discriminant function analysis can be recommend for pre-harvest forecasting of wheat crop yield one and half months before the harvest.

Table 20: Details of best three forecast models for wheat crop

Methods based on	Model	Year	Actual yield (Q/ha)	Predicted yield (Q/ha)	Percent deviation	R^2 (%)	RMSE
Discriminant Function Analysis	D_3	2015-16	27.27	30.86	13.15	80.1	9.33
		2016-17	33.54	21.22	36.72		
		2017-18	34.40	24.57	28.57		
	D_4	2015-16	27.27	17.06	37.44	79.4	11.66
		2016-17	33.54	21.64	35.49		
		2017-18	34.40	21.66	37.03		
	D_5	2015-16	27.27	21.56	20.94	80.9	3.51
		2016-17	33.54	33.59	0.13		
		2017-18	34.40	32.32	6.05		

The aforesaid developed models can be used for pre-harvest forecast of wheat crop yield in neighbouring district of Azamgarh district provided they have almost similar agro-climatic conditions.

References

1. Agrawal R, Jain RC, Jha MP. Joint effects of weather variables on rice yields. Mausam. 1983;32(2):177-81.
2. Agrawal R, Jain RC, Jha MP. Models for studying rice crop weather relationship. Mausam. 1986;37(1):67-70.
3. Chandrahas AR, Aditya K. Use of discriminant function analysis for forecasting crop yield. Mausam. 2012;36(3): 455-458.
4. Anderson TW. An introduction to applied Multivariate Statistical Analysis, John Wiley & Sons Inc. New York; c1984.
5. Draper NR, Smith H. Applied Regression Analysis. 3rd edition, John Wiley & Sons Inc. New York; c1998.
6. Hair Joseph F, Anderson Rolph E, Tatham Ronald L, Black William C. Multivariate Data Analysis with Readings, Prentice Hall, Inc New Jersey; c1995.
7. Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. Third Edition, Prentice Hall of India Private Limited, New Delhi; c2001.
8. Chandrahas TR. Use of discriminant function of weather parameters for developing forecast model of rice crop. Publication of IASRI, New Delhi; c2000.
9. Sharma S. Applied Multivariate Techniques, John Wiley & Sons Inc. New York; c1996.
10. Sisodia. Forecasting of pre-harvest crop yield using discriminant function analysis of meteorological parameters. Journal of Agrometeorology. 2014;16(1):121-125
11. Neeraj. Effect of weather variables on the yield of wheat crop in District Azamgarh of Eastern Uttar Pradesh, India. The Pharma Innovation. 2021;10(0):1015-1017.
12. Neeraj. Effect of weather variables on the yield of rice crop in district Azamgarh of eastern Uttar Pradesh, India. International Journal of Chemical Studies. 2021;9(2):259-261.

Appendix – (a)

Eigenvalues				
Function	**Eigenvalue**	**% of Variance**	**Cumulative %**	**Canonical Correlation**
1	.924[a]	87.0	87.0	.693
2	.138[a]	13.0	100.0	.348
a. First 2 canonical discriminant functions were used in the analysis.				

Standardized Canonical Discriminant Function Coefficients		
	Function	
	1	2
z10	.798	.096
Z20	-.686	.438
Z30	.867	.936
Z40	-.201	.341
Z50	.337	-.072

Functions at Group Centroids		
WHEAT YIELD	Function	
	1	2
1.00	-.979	.881
2.00	.446	-.057
3.00	-2.140	-.485
Unstandardized canonical discriminant functions evaluated at group means		

Discriminant scores

DS1	DS2
1.4148	-1.37386
1.67105	1.88587
-0.04289	-0.01302
0.34213	-0.91335
0.8743	0.5555
2.43202	-1.63187
0.32563	0.82524
0.02433	-0.7642
-1.22516	-0.39877

0.22641	-0.60637
-0.25025	-0.6229
-0.69152	-0.24965
1.74053	1.34715
-0.60391	1.16819
-1.33102	1.31189
-0.62604	0.45111
-2.55525	-0.44776
-1.72516	-0.52319

Appendix – (b)

Eigenvalues				
Function	**Eigenvalue**	**% of Variance**	**Cumulative %**	**Canonical Correlation**
1	.922[a]	67.8	67.8	.693
2	.438[a]	32.2	100.0	.552
a. First 2 canonical discriminant functions were used in the analysis.				

Standardized Canonical Discriminant Function Coefficients		
	Function	
	1	2
Z11	-.953	1.429
Z21	-.271	1.019
Z31	.219	-.257
Z41	.161	.416
Z51	.910	.585

Functions at Group Centroids		
WHEAT YIELD	Function	
	1	2
1.00	-2.477	.071
2.00	.324	.233
3.00	.208	-1.702
Unstandardized canonical discriminant functions evaluated at group means		

Discriminant scores

DS1	**DS2**
1.54317	-1.51492
0.5574	1.17655
-0.17587	1.55689
1.1008	1.32759
-2.36119	-1.10161
0.91276	0.39386
1.35694	0.16518
0.4042	0.98969
-0.08283	-1.38504
-0.94389	1.54049
1.18771	-0.68255
0.52711	-0.19142
1.08611	0.43357
-0.57446	0.55424
-2.48963	-0.14997
-2.46348	0.29163
0.32353	-1.44462
0.09165	-1.95955

Appendix – (c)

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	38.120[a]	83.1	83.1	.987
2	7.754[a]	16.9	100.0	.941
a. First 2 canonical discriminant functions were used in the analysis.				

Standardized Canonical Discriminant Function Coefficients		
	Function	
	1	2
Z10	5.999	2.730
Z20	1.446	11.462
Z30	26.572	-.682
Z40	-2.051	1.146
Z50	-.171	.381
Z11	2.348	1.933
Z21	1.908	-.925
Z31	-1.277	-1.949
Z41	-2.197	-.535
Z51	-2.441	.327
Z120	-3.768	-11.122
Z130	-19.139	20.167
Z230	-2.015	-20.270
Z340	-1.117	-.212
Z451	6.508	1.376

Functions at Group Centroids		
WHEAT YIELD	Function	
	1	2
1.00	-9.272	-5.849
2.00	2.997	.138
3.00	-11.707	4.880
Unstandardized canonical discriminant functions evaluated at group means		

Discriminant scores

DS1	DS2
3.50384	0.32158
3.1883	-0.45549
2.23902	-0.64572
2.39934	0.44427
3.37214	-0.22682
2.49093	0.18505
3.19457	-0.18734
5.10441	0.39388
0.37664	0.59129
2.85694	-0.02643
4.22882	2.07723
3.20587	-1.36633
2.6028	0.97226
3.19417	-0.13976
-9.25128	-5.37081
-9.29181	-6.32634
-12.0649	6.62812
-11.3498	3.13138

Appendix – (d)

Eigenvalues				
Function	**Eigenvalue**	**% of Variance**	**Cumulative %**	**Canonical Correlation**
1	2.241[a]	65.2	65.2	.832
2	1.194[a]	34.8	100.0	.738
a. First 2 canonical discriminant functions were used in the analysis.				

Standardized Canonical Discriminant Function Coefficients		
	Function	
	1	2
Z10	.970	.543
Z20	-.386	-.429
Z30	-.039	.873
Z40	-.692	.066
Z50	-1.202	.720
Z11	-.149	1.190
Z21	.236	.902
Z31	.466	.109
Z41	1.064	.086
Z51	1.459	-.462

Functions at Group Centroids		
WHEAT YIELD	Function	
	1	2
1.00	-3.510	1.181
2.00	.668	.216
3.00	-1.166	-2.690
Unstandardized canonical discriminant functions evaluated at group means		

Discriminant scores

DS1	DS2
1.19907	-1.29413
0.75847	1.86217
-0.58011	0.90281
2.89152	-0.28198
-0.00514	1.27495
1.1899	1.5211
1.49965	-0.03705
1.4795	0.40999
-0.31092	-1.29642
-0.36923	1.11962
0.67467	-1.05096
-0.66352	-0.09368
1.63148	0.69053
-0.04209	-0.70841
-4.17839	0.70184
-2.8423	1.6603
-1.17226	-2.78922
-1.16028	-2.59144

Appendix – (e)

Eigenvalues				
Function	**Eigenvalue**	**% of Variance**	**Cumulative %**	**Canonical Correlation**
1	64.461[a]	95.5	95.5	.992
2	3.040[a]	4.5	100.0	.867
a. First 2 canonical discriminant functions were used in the analysis.				

Standardized Canonical Discriminant Function Coefficients		
	Function	
	1	2
Z10	.211	1.367
Z20	5.995	4.469
Z30	9.444	-6.575
Z40	-2.197	.696
Z50	-13.763	.060
Z120	-5.201	-6.351
Z130	-11.333	9.952
Z230	6.943	-4.552
Z140	11.011	-3.619
Z240	-9.423	4.479
Z340	-4.616	-.006
Z150	14.913	.092

Functions at Group Centroids		
WHEAT YIELD	Function	
	1	2
1.00	-20.638	-.424
2.00	2.856	-.582
3.00	.644	4.500
Unstandardized canonical discriminant functions evaluated at group means		

Discriminant scores

DS1	DS2
3.76689	-1.7554
2.60379	-1.33042
2.62628	-1.57028
3.08013	-0.1298
3.75553	-0.69383

3.39155	-1.23197
3.34617	-1.07836
-0.18197	0.29442
3.17953	0.95082
2.83199	-0.30398
3.50774	-0.8534
3.14849	1.23807
2.52463	-0.07089
2.4073	-1.6175
-20.1071	-0.14319
-21.1691	-0.70425
1.63088	5.75064
-0.34274	3.24931

Appendix – (f)

Eigenvalues				
Function	**Eigenvalue**	**% of Variance**	**Cumulative %**	**Canonical Correlation**
1	21.416[a]	87.9	87.9	.977
2	2.936[a]	12.1	100.0	.864
a. First 2 canonical discriminant functions were used in the analysis.				

Standardized Canonical Discriminant Function Coefficients		
	Function	
	1	2
Z11	.423	-3.334
Z21	-15.552	7.282
Z31	-10.340	-16.353
Z41	-.121	3.061
Z51	2.557	4.092
Z121	15.542	-8.553
Z131	22.326	5.685
Z231	-10.931	12.647
Z141	-15.802	5.964
Z241	16.065	-6.337
Z151	1.768	6.780
Z251	-4.308	-5.195
Z451	2.579	-3.823

Discriminant scores

DS1	DS2
1.41617	1.9602
0.84071	-1.60774
1.12302	-0.82863
2.05838	-0.88578
1.51758	-0.13546
4.0251	-0.52552
3.81313	0.53275
3.22454	0.24641
2.35143	-0.64525
2.46299	-0.98275
1.80263	-1.53104
0.94992	0.72925
2.00189	-0.01902
1.78982	-0.63121
-11.0423	-2.84424
-10.2672	-1.16046
-4.96486	5.09254
-3.10293	3.23593