Scope
Volume 14 Number 02 June 2024

"Use of Discriminant Function Analysis for Forecasting Crop Rice Yield in District Jaunpur, Eastern Uttar Pradesh, India"

Piyush Kumar Singh[1]* Ramesh Pratap[2] Singh , Vishva Deepak Chaturvedi[3] & Prabhas Kumar Shukla[2]

[1]Department of Agricultural Statistics, Acharya Narendra Dev University of Agriculture and Technology, Kumarganj – Ayodhya (UP), India
[2]Department of Biochemistry, Acharya Narendra Dev University of Agriculture and Technology, Kumarganj – Ayodhya (UP), India
[3]Department of Genetics & Plant Breeding, CSK HPKV Palampur, India

Correspondence Author: **Piyush Kumar Singh**

Abstract: This research aims to demonstrate how discriminant function analysis may be used to create a rice production forecasting model for Jaunpur (India). Discriminant function analysis is a method of creating a linear/quadratic function that best discriminates different populations and so provides a qualitative evaluation of the likely yield. Time series data from 18 years (2000-2019) have been divided into three categories: congenial, normal, and adverse, based on yield distribution. Taking these three groups as three populations, discriminant function analysis has been carried out. The regressors in the model were discriminant scores obtained from this. The use of weekly weather data has been proposed in a variety of ways. The models were used to forecast yields for the three years following 2015-16: 2015-17: 2017-18. (which were not included in model development). About two months before harvest, the method offered a reliable yield prediction.

Keywords – Weather variables, Weather indices, Discriminant function analysis, Crop yield forecast modeling.

1. Introduction

Crop production projections must be accurate and timely in order for an agrarian economy to function. Crop yield estimates are critical for long-term planning, policy creation, and execution in areas such as food procurement, distribution, pricing, and import-export decisions, among others. These are also useful to farmers to decide in advance their future prospects and course of action. Thus, reliable and timely pre-harvest forecasts of crop yield are very important. For this purpose, researchers have tried weather-based models using different statistical approaches. In the present paper, the use of discriminant function analysis has been explored for forecasting crop yield. Data on rice yield in Uttar Pradesh's Jaunpur area was used to show the process. Approximately one week after the pre-sowing stage, Crown root initiation happens 20-25 days after seeding, or three weeks after germination. After crown root initiation, the tillering phase lasts around 2-3 weeks. The Jointing and Reproductive phase, which begins after the flowering stage, is the most active period of plant growth.

Because weather during the pre-sowing period is critical for crop establishment, data from two weeks prior to sowing have been incorporated into the model development. Furthermore, because the forecast is needed long ahead of harvest, weather data from two months prior to harvest has been taken into account. As a result, information on five meteorological variables, including maximum and minimum temperatures, rainfall, wind velocity, and sunshine hour, was collected.

2. Data

The time series data on yield for rice crop of Jaunpur district of eastern Uttar Pradesh pertaining for the period from 2000-01 to 2017-18 have been procured from the website *updes.up.nic.in* by **Economics and Statistics Division, Planning Department,** Government of Uttar Pradesh. Weekly weather variables data for rice crops in the Jaunpur, Eastern Uttar Pradesh district 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 Sunshine hours have been used in the study. Data from 2015-16 to 2017-18 were used for validation of the models.

Rice harvesting in the Jaunpur district usually begins in June. As a result, 16 weeks of weather data (from the 23rd to the 38th SMW) were used to create statistical models, and discriminant scores were calculated for each year. These scores were used along with the year as regressors in developing the forecast models. In the present study, the number of groups is three, and the number of weather variables is five. Therefore, only two discriminant functions are sufficient for discriminating a crop year into either of the three groups.

However, using weather variables in developing the model poses a problem. Weather variables affect the crop differently during different phases of development. Thus, the extent of weather influence on crop yield depends not only on the magnitude of weather variables but also on the distribution pattern of weather over the crop season, which, as such, calls for the necessity of dividing the whole crop season into finer intervals. However, doing so will increase the number of variables in the model, and in turn, a large number of parameters will have to be evaluated from the data, and a sufficient number of observations may only be available for a partial estimation of these parameters. This gives rise to the problem of a number of variables under study that are more than a number of observations. To solve this problem, suitable strategies have been suggested, and the following six models were

3. Statistical methodology

Discriminant function analysis is a multivariate technique discussed in many books, to mention a few: Anderson (1984), Hair *et al.* (1995), Sharma (1996), Johnson and Wichern (2006), etc. This technique is used to identify appropriate functions that discriminate best between a set of observations from two or more groups and classify future observations into one of the previously defined groups.

Suppose observations are to be classified into k groups on the basis of p variables. The technique involves identifying linear/quadratic function(s) where the coefficients 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 a minimum of k-1 and p. These functions are used to calculate discriminant scores, which are used to classify the observations into different groups.

Rai and Chandrahas (2000) developed forecast models for rice in the Raipur district using the discriminant function technique and provided reliable yield forecasts about two months before harvest. This paper applies the technique used by them, along with a few modifications.

To apply discriminant function analysis for modeling yield using weather variables, crop years have been divided into three groups, namely congenial, normal, and adverse, 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 scores were used along with the year as regressors in developing the forecast models. In the present study, the number of groups is three, and the number of weather variables is five. Therefore, only two discriminant functions are sufficient for discriminating a crop year into either of the three groups.

However, using weather variables in developing the model poses a problem. Weather variables affect the crop differently during different phases of development. Thus, the extent of weather influence on crop yield depends not only on the magnitude of weather variables but also on the distribution pattern of weather over the crop season, which, as such, calls for the necessity of dividing the whole crop season into finer intervals. But, doing so will increase the number of variables in the model, and in turn, a large number of parameters will have to be evaluated from the data, and a sufficient number of observations may not be available for precise estimation of these parameters. This gives rise to the problem of a number of variables under study that are more than a number of observations. To solve this problem, suitable strategies have been suggested, and the following six models were proposed:

Model-D_1:

In this procedure, five unweighted weather indices have been used as discriminating variables. Now, based on these five 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 the trend variable (T=1,2,3,……,n) and ε is an error term assumed to follow an independently normal distribution with mean 0 and variance σ^2 .

Model-D_2:

In this procedure, five weighted weather indices have been used as discriminating variables. Now, based on these five 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 the trend variable (T=1,2,3,……,n) and ε is an error term assumed to follow an 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) were used as discriminating variables in discriminant function analysis, and two sets of discriminant scores from two discriminant functions were obtained. The forecasting model has been fitted, taking un-trended yield as the regressand variable, 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 the trend variable (T=1,2,3,……,n) and ε is an error term assumed to follow an independently normal distribution with mean 0 and variance σ^2 .

Model-D_4:

In this procedure, five weighted and five unweighted weather indices have been used as discriminating variables. Now, based on these ten 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

ds₁ and *ds₂* are the two discriminant scores. T is the trend variable (T=1,2,3,……,n) and ε is an error term assumed to follow an independently normal distribution with mean 0 and variance σ^2.

Model-D$_5$:

In this procedure, five un-weighted and ten un-weighted interaction weather indices were 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 the trend variable (T=1,2,3,……,n) and ε is an error term assumed to follow an independently normal distribution with mean 0 and variance σ^2.

Model-D$_6$:

In this procedure, five weighted and ten 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. Based on these two sets of scores, the regression model has been fitted, taking the yield as the regress 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 *ds₁* and *ds₂* are the two discriminant scores. T is the trend variable (T=1,2,3,……,n) and ε is an error term assumed to follow an independently normal distribution with mean 0 and variance σ^2 .

3.5. Comparison and validation of models

The six models were compared on the basis of the adjusted coefficient of determination of R^2_{adj}, which is as follows:

$$R^2_{adj} = 1 - \frac{ss_{res}/(n-p)}{ss_t/(n-1)}$$

where ssres/(*n*-*p*) is the residual mean square and *sst*/(*n*-1) is the total mean square.
From the fitted models, Rice yield forecasts for the years 2015-16 to 2017-18 were obtained, and forecasts were compared on the basis of Root Mean Square Error (RMSE).

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

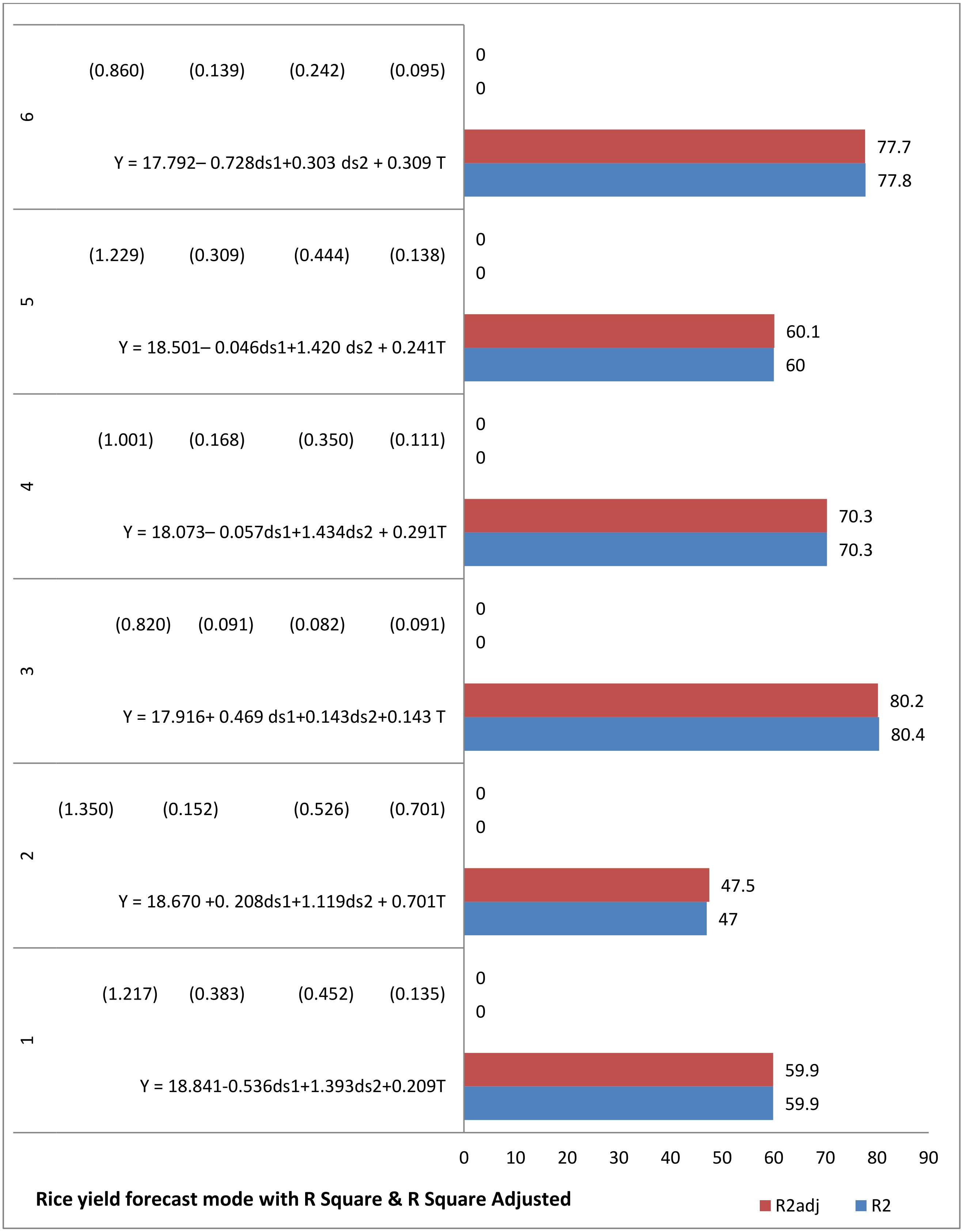
Where O_i and E_i are the observed and forecasted values of the crop yield, respectively, and n is the number of years for which forecasting has been done.

Table 1 Rice yield forecast models

model	Forecast regression equation	R²	R²adj
1.	**$Y = 18.841 - 0.536ds_1 + 1.393ds_2 + 0.209T$** (1.217) (0.383) (0.452) (0.135)	59.9	59.9
2.	**$Y = 18.670 + 0.208ds_1 + 1.119ds_2 + 0.701T$** (1.350) (0.152) (0.526) (0.701)	47.0	47.5

3.	**$Y = 17.916+ 0.469\ ds_1+0.143ds_2+0.143\ T$** (0.820) (0.091) (0.082) (0.091)	80.4	80.2
4.	**$Y = 18.073- 0.057ds_1+1.434ds_2 + 0.291T$** (1.001) (0.168) (0.350) (0.111)	70.3	70.3
5.	**$Y = 18.501- 0.046ds_1+1.420\ ds_2 + 0.241T$** (1.229) (0.309) (0.444) (0.138)	60.0	60.1
6.	**$Y = 17.792- 0.728ds_1+0.303\ ds_2 + 0.309\ T$** (0.860) (0.139) (0.242) (0.095)	77.8	77.7

Note: figures in brackets denote the Standard Error of regression coefficients



Rice yield forecast mode with R Square & R Square Adjusted

Table 2
Actual & forecast yield of rice (Q/ha)

Models	Prediction period	Actual yield(Q/ha)	Forecasted yield(Q/ha)	Percent deviation	RMSE
D_1	2015-16	17.6	19.46	-10.54	1.50
	2016-17	23.4	23.34	0.35	
	2017-18	23.1	24.91	-8.03	
D_2	2015-16	17.6	28.94	-64.44	8.73
	2016-17	23.4	31.78	-35.71	
	2017-18	23.1	28.56	-23.86	
D_3	2015-16	17.6	13.48	23.38	2.94
	2016-17	23.4	21.10	9.92	
	2017-18	23.1	21.18	8.15	
D_4	2015-16	17.6	17.54	0.33	1.83
	2016-17	23.4	26.45	-12.95	
	2017-18	23.1	23.99	-4.03	
D_5	2015-16	17.6	19.16	-8.88	1.08
	2016-17	23.4	24.27	-3.63	
	2017-18	23.1	23.67	-2.64	
D_6	2015-16	17.6	28.12	-59.74	6.17
	2016-17	23.4	21.52	8.12	
	2017-18	23.1	22.88	0.79	

4. Results and conclusions

Table 1 shows the forecast models developed using the six techniques (described in section 3) as well as the adjusted coefficient of determination Radj 2. The trend variable T was found to be significant in all of the models. Apart from trend T, two important variables were discovered in the model: discriminant scores ds1 and ds2.The adjusted coefficient of determination $R_{adj}{}^{2}$ varied between 47.0 to 80.4 in different models, the maximum (80.4) being in model 3. RMSE was computed on the basis of yield forecasts for the years 2015-16 to 2017-18. The results (Table2)revealed that the percent deviation of forecast varied from 0.35 to 10.54 in model-1,23.71to, 64.44 in model-2, 8.15 to 23.38 in model-3, 0.33 to 12.95 in model-4, 2.64 to 8.88 in model-5 and 0.79 to 59.74 in model-6 over the three years. The RMSE varied from a minimum of 1.08 in model 5 to a maximum of 6.17 in model 6. Thus, it is concluded that model 3 is the most suitable model among the models considered for forecasting Rice yield for the Jaunpur district of Uttar Pradesh. The model provides a reliable forecast around two months before harvest.

References

1. Kumar, N., R. R. Pisal, S. P. Shukla and K. K. Pandey (2014).Crop yield forecasting of paddy, sugarcane, and wheat through linear regression technique for south Gujarat.*Mausam*, **65(3):** 361-364.
2. Mehta, S.C., R. Agrawal & V. P. N. Singh (2000). Strategies for composite forecast. *Jr Indian Soc. Ag. Stat*., **53(3):**262-72.
3. Pandey, K.K., R. P. Kaushal, A. N. Mishra and V. N. Rai (2009).A study on the Impact of Weather Variables with different distributions. *Ind. J. Agricult. Stat. Sci*., **5(1):** 139-53.
4. Pandey, K. K., V. N. Rai and B. V. S. Sisodia (2014). Weather Variables-Based Rice Yield Forecasting Models For
5. Rai, T. and Chandrahas (2000). Use of discriminant function of weather parameters for developing forecast model of rice crop. (*IASRI Publication*).
6. Ramasubramanian, V. and R. C. Jain (1999). Use of growth indices in Markov chain model for crop yield forecasting.*Biometrical Journal*, **41(1):** 99-109.
7. Robertson, G. W. (1974). Wheat yields for 50 years at SwiftCurrent Saskatchewan in relation to weather. *J. Plant Sci*.,**54**: 625-50.

8. Saksena, A., R. C. Jain and R. L. Yadav (2001). Development of early warning and yield assessment models for rainfed crops based on agro-meteorological indices. (*IASRIPublication*).

9. Srivastava, A. K., P. K. Bajpai, R. L. Yadav and S. S. Hasan(2007). Weather-Based Sugarcane Yield Prediction Model for the State of Uttar Pradesh. *Journal of Indian Society of Agricultural Statistics*, **61(3)**: 313-327.

10. Varshneya, M. C., S. S. Chinchorkar, V. B. Vaidya and V. Pandey(2010). Forecasting model for seasonal rainfall for different regions of Gujarat. *J. Agrometeorology*, **12(2)**: 202-207

11. Shankar U and Gupta BRD (1987). Forecasting the yield of paddy at Chisurah in West Bengal using multiple regression techniques. Mausam 38: pp. 415–418.

12. Agrawal R, Chandrahas, Kumar A (2012) Used discriminant function analysis For forecasting crop yield. Mausam 63: 455-458.

13. Agrawal R, Jain RC, Jha MP (1986). Models for studying rice crop weather relationship. Mausam 37: 67-70