International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2019; 7(3): 1897-1900

Received: 19-03-2019
Accepted: 21-04-2019

Forecasting of pre-harvest rapeseed and mustard yield using discriminant function analysis of meteorological parameters

Sarvesh Kumar, VN Rai, KK Mourya, Annu and Ravi Prakash Gupta

Sarvesh Kumar
Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, Uttar Pradesh, India

VN Rai
Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, Uttar Pradesh, India

KK Mourya
Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, Uttar Pradesh, India

Annu
Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, Uttar Pradesh, India

Ravi Prakash Gupta
Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, Uttar Pradesh, India

Correspondence
Sarvesh Kumar
Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, Uttar Pradesh, India

Abstract
An application of discriminant function analysis of meteorological parameters for developing suitable statistical models to forecast Rapeseed & Mustard yield in Sultanpur district of Eastern Utter Pradesh has been demonstrated. Time series data on Rapeseed & Mustard yield for 25 years (1990-91 to 2004-15) 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 six 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 yield for the year 2012-13, 2013-14 and 2014-15 which were not included in the development of the models. It has been found that most of the models provide reliable forecast of the Rapeseed & Mustard yield about one and half months before the harvest. However, the model-5 has been found to be the most suitable among all the models developed.

Keywords: Meteorological parameters, crop yield, discriminant function analysis, forecast model

Introduction
In India Rapeseed-mustard (Brassica spp.) is an important crop among oilseed crops in the world and second largest cultivator after China. Although there has been a significant increase in oilseed production since 1960s, the demand for oilseeds production in the future is likely to go up due to population increase and their income. Rapeseed and Mustard is much sensitive to climatic variables and hence climate could have significant effect on its production. Rapeseed & mustard is one of major oil seeds crops of Uttar Pradesh in general and particularly in Eastern Uttar Pradesh. There are very limited studies to assess the impact of climate change on oilseed crops as compared to cereals. The individual effect of weather factors on rice and wheat yields have been studied by various research workers in the past. Notably among them are Fisher (1924) [7], Hendricks and scholl (1943) [8], Huda *et al.* (1975) [9], Jain *et al.* (1980) [10], Agrawal *et al.* (1980, 1986) [1, 3], Annu *et al*. (2017) etc. Yield forecast models for rice and wheat based on weather factors have also been developed by Agrawal *et al.* (1980, 1983, 1986) [1, 2, 3], and Jain *et al.* (1980) [10]. Recent work of Agrawal *et al.* (2012) [4], Sisodia *et al.* (2014) [11, 12] and Yadav *et al.* (2014) [11, 12] have made application of discriminant function analysis of weekly weather data to develop forecast model for rice and wheat yield.
Application of discriminant analysis is a multivariate technique concerned with separating distinct sets of objects (or sets of observations) and allocating new objects (or observations) to the previously defined groups. As an example, a medical researcher is interested in differentiating between the patients who have had a heart attack and those who have not yet had a heart attack. The medical researcher then wants to use some identified factors to predict whether a patient is likely to have a heart attack in the future. In a similar way forecasting of crop yield on the basis of the weather variables can be done using discriminant analysis. In this context the crop year can be categorized as congenial, normal or adverse using discriminant function analysis on the basis of data on weather variables.

Development of statistical forecast models
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 six, therefore only two discriminant functions can be obtained which are sufficient for discriminating a crop years 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].

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

$$Z_{ii',j} = \sum_{w=1}^{n} r_{ii'w}^{j} X_{iw} X_{i'w} \Big/ \sum_{w=1}^{n} r_{ii'w}^{j} \quad ,j=0,1 \text{and } i=1,2,\ldots,p.$$

where Z_{ij} is un-weighted (for j=0) and weighted (for j=1) weather indices for ith 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. $\boldsymbol{X_{iw}}$ 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. Here, p=6 and n=21, i.e. 21 weeks data from 40th week to 52nd week of a year and 1st week to 8th week of the next year have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions. Here only the first 25 years data from 1990-91 to 2014-15 have been utilized for model fitting and remaining three years were left for the validation of the model. The growth process of the crop has various phases and weeks within phases. In the development of pre-harvest model based on discriminant function analysis the entire 21 weeks data from 40th SMW to 52nd SMW of a year and 1st SMW to 8th SMW of the next year have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions. In all 42 indices (21 weighted and 21 un-weighted) consisting of 6 weighted and 15 weighted interaction weather indices and 6 un-weighted and 15 un-weighted interaction weather indices have been constructed. 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$_1$

This model is the 2nd model of Agrawal *et al.* (2012) [4]. This model utilizes the complete data over 21 weeks and also considers relative importance of weather variables in different weeks. Using six weighted weather indices of six weather variables as discriminating variables, discriminant function analysis has been carried out and two discriminant functions have been obtained. Two sets of discriminant scores for the years under consideration from these two discriminant functions were obtained. For developing forecast model, these two sets of discriminant scores along with the trend variable were utilized as the regressors and the yield as the regressand. The form of model considered is as follows:

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

Model-D$_2$

This model is 4th model of Agrawal *et al.* (2012) [4]. Two discriminant functions and therefrom two sets of discriminant scores have been obtained using the first week data (40th SMW) on six weather variables. Next, two sets of discriminant scores obtained from first week and the second week (41st SMW) data on six weather variables data have been used as discriminating variables, so in all there were 8 discriminating variables, and based on these 8 discriminating variables the discriminant function analysis has been done and, therefore, two sets of discriminant scores have been obtained. This process was repeated upto the last week till the time of forecast (8th SMW or 21st week) and finally two sets of discriminant scores have been obtained. Based on these two sets of discriminant scores, the forecasting model taking yield as the regressand and the discriminant scores and the trend variable as the regressor variables has been fitted. The form of model considered is as follows:

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

Model-D$_3$

In this procedure, all 42 (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:

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

Model-D$_4$

In this procedure, 6 weighted and 6 un-weighted weather indices have been used as discriminating variables. Now, based on these 12 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

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

Model-D5

In this procedure, discriminant function analysis has been carried out using the un-weighted and weighted averages (weather indices) for the first weather variable (here, discriminating factors will be only two). Using the two sets of discriminant scores on the basis of first weather variable, and un-weighted and weighted averages (weather indices) for the second weather variable, discriminant function analysis has been further carried out (here, the discriminating factors will be 4). This process is continued up to sixth weather variables, and finally we get two sets of discriminant scores ds_1 and ds_2. Using crop-yield as regressand and discriminant scores ds_1 & ds_2 and the time trend T as regressor variables, the following model is fitted for the development of forecast model.

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

Model-D6

In this procedure, discriminant function analysis have been carried out using weekly data of the first weather variable spread over 21 weeks as discriminating variable. Using two sets of discriminant scores obtained from two estimated discriminant functions based on data of the first weather variable and 21 weeks data of the second variable, discriminant function analysis has been again performed and two sets of discriminant scores are obtained (here discriminating variable will now become 23). Using these two sets of discriminant scores and 21 weeks data of third weather variable have been again used to carry out discriminant function analysis and subsequently two sets of discriminant scores have been obtained. This process is continued up to sixth weather variables, and ultimately we get two sets of discriminant scores ds_1 and ds_2. These two sets of scores and the trend variable (T) as the regressor variables and crop-yield as the regressand were utilized to develop forecast model by fitting the following model.

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

Comparison of the model

Based on these two forecast models, the forecast yields for the 2012-13, 2013-14 and 2014-15 have been computed and result are presented in Table-The values of adj R^2, percent deviation of forecast from actual yield, RMSE and %SE (CV) have also been computed for best three mode are also presented in the Table-1. The result of the Table – 1 have been graphically presented in Fig – 1 to high light the forecasting power of the models.

Table 1: Comparison between actual and forecasted yield of different years of Sultanpur District

Model	Year	Actual yield	Predicted yield	Percent deviation)	PSE	R^2	Adj R^2	RMSE
I	2012-13	7.32	8.03	9.74	7.25	79.5	76.0	0.74
	2013-14	4.99	5.21	4.47	20.28			
	2014-15	4.45	5.50	23.71	18.79			
II	2012-13	7.32	8.89	14.95	6.19	84.8	82.3	1.37
	2013-14	4.99	6.65	21.53	11.19			
	2014-15	4.45	5.09	33.42	17.48			
III	2012-13	7.32	7.52	2.78	7.42	79.4	75.9	0.64
	2013-14	4.99	5.17	3.75	23.18			
	2014-15	4.45	5.53	24.30	22.12			

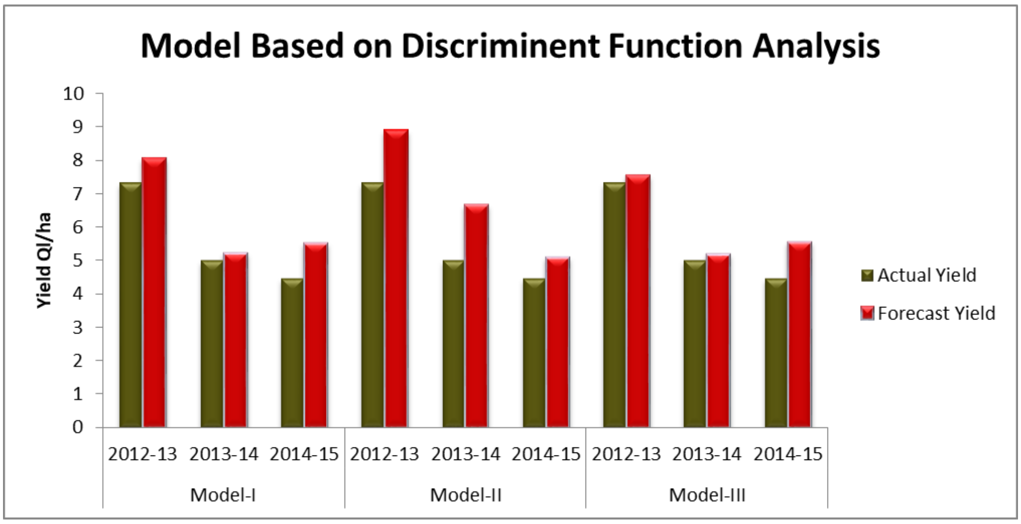

Fig 1: Model Based on Discriminent Function Analysis

Table 2: Best Model from the application of discriminant function analysis of weekly weather data

Model-I	$Y = 7.150+0.227ds_1+0.064ds_2+0.037T$	$R^2 = 79.5$	Adj $R^2 = 76.0$
Model-II	$Y = 7.015+ 0.651\ ds_1- 0.214\ ds_2 + 0.055T$	$R^2 = 84.8$	Adj $R^2 = 82.3$
Model-III	$Y = 7.261– 0.021ds_1+0.547\ ds_2 + 0.028\ T$	$R^2 = 79.4$	Adj $R^2 =75.9$

Results and Discussion

The forecast models developed under the six procedures along with R^2 and AdjR^2 are given in Table 1. In all the models, the time trend variable (T) has been found to be significant at one percent probability level of significance ($P<0.01$). First discriminant score (ds_1) has been found to be significant at $P<0.01$ in all the models except in model 5 where it is significant at $P<0.05$. The second discriminant (ds_2) has been found to be significant at $P<0.01$ in model. Adjusted coefficient of determination (R^2 adj) was found to be maximum of 82.3 percent in model 5 (2nd in the table 1 Model – II) while it was minimum (75.9%) in model 6 (3rd in the table 1 Model – III). Based on these forecast models, the forecast yields for the 2012-13, 2013-14 and 2014-15 were obtained and the results are presented in Table 2. The percent standard error (cv) of forecast yields have also been computed for best three models and are presented in the Table 2. The model 2 (2nd in the Table 1) suggested by Agrawal *et al*. (2012) [4] comes out to be the second best on account of high values of R^2 adj and lower value of percent standard error (cv), and it is close to model 5 (2Nd in the table 1 Model – II). However, on the basis of the overall results of the Table- 1 and 2, it can be concluded that the proposed model 5(2nd in the table 1 Model – II) is the most suitable model among all the models to forecast Rapeseed and Mustard yield in Sultanpur district of Eastern Uttar Pradesh. Hence, a reliable forecast of Rapeseed and Mustard yield about one and half months before the harvest can be obtained from the proposed model 5 (2Nd in the table 1 Model – II).

References

1. Agrawal Ranjana, Jain RC, Singh D. Forecasting of rice yield using climatic variables. Ind. J Agric. Sci. 1980; 50(9):680-84.
2. Agrawal R, Jain RC, Jha MP. Joint effects of weather variables on rice yields. Mausam. 1983; 32(2):177- 81.
3. Agrawal R, Jain RC, Jha MP. Models for studying rice crop weather relationship. Mausam. 1986; 37(1):67-70.
4. Agrawal R, Chandrahas, Kaustav A. Use of discriminant function analysis for forecasting crop yield, *Mausam*, 2012; 63(3):455-458.
5. Anderson TW. An introduction to applied Multivariate Statistical Analysis, John Wiley & Sons Inc. New York, 1984.
6. Draper NR, Smith H. Applied Regression Analysis. 3rd edition, John Wiley & Sons Inc. New York, 1998
7. Fisher RA. The influences of rainfall on the yield of wheat at Rothamsted. Philosophical Transaction of Royal Society of London, Series B, 1924, 213:89-142.
8. Hendricks WA, Scholl GC. Technique in measuring joint relationship: The joint effects of temperature and precipitation on crop yield. N. Carolina Agric. Exp. Stat, 1943.
9. Huda AKS, Ghildyal RA, Tomar VS, Jain RC. Contribution of climatic variable in predicting rice yield. Agric. Met. 1975; 15:71-86.
10. Jain RC, Agrawal Ranjana, Jha MP. Effect of climatic variables on rice yield and its forecast. Mausam. 1980; 31(4):591-96.
11. Sisodia BVS, Yadav RR, Kumar S, Sharma MK. Forecasting of Pre- harvest crop yield using discriminant function analysis of meteorological parameter. Journal of Agrometeorology. 2014; 16(1).121-125.
12. Yadav RR, Sisodia BVS, Sunil Kumar. Application of principal component analysisin developing statistical models to forecast crop yield using weather variables. *Mausam*. 2014; 65(3):357-360.