

Forecasting of pre-harvest crop yield using discriminant function analysis of meteorological parameters

B. V.S. SISODIA, R. R. YADAV, SUNIL KUMAR and M. K. SHARMA

Department of Agricultural Statistics,
Narendra Deva University of Agriculture and Technology, Kumarganj, Faizabad-224 229 (U.P.)
Email: bvssisodia@gmail.com

ABSTRACT

In the present paper, an application of discriminant function analysis of meteorological parameters for developing suitable statistical models to forecast wheat yield in Faizabad district of Eastern Uttar Pradesh has been demonstrated. Time series data on wheat yield for 20 years (1990-91 to 2009-10) 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 2008-09 and 2009-10 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-5 has been found to be the most suitable among all the models developed.

Key word: Meteorological parameters, Crop yield, Discriminant function analysis, Forecast model

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 Chandras (2000) 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. Agrawal *et al.* (2012) have recently developed forecast models for wheat yield in Kanpur district (U.P.) using discriminant functions analysis of weekly 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 yield in Faizabad district of Uttar Pradesh using discriminant functions analysis of weekly data on

weather variables.

MATERIALS AND METHODS

The study was conducted for Faizabad district of Eastern Uttar Pradesh, India which is situated between 26° 47' N latitude and 82° 12' E longitudes. It lies in the Eastern plain zone of Uttar Pradesh. It has an annual rainfall of about 1002 mm and is liberally sourced by the Saryu (Ghaghra) river and its tributaries. Soils are deep alluvial, medium to medium heavy textured but are easily ploughable. The favorable climate, soil and the availability of ample irrigation facilities make growing of rice and wheat a natural choice for the area. The objective is to develop pre-harvest forecast model for wheat yield.

Time series data on yield for wheat crop of Faizabad district of Uttar Pradesh for 20 years (1991-2010) were collected from the bulletins of Directorate of Agricultural Statistics and Crop Insurance, Govt. of Uttar Pradesh. Weekly weather data for the period (1991-2010) on the weather variables of Faizabad district of Uttar Pradesh during the different growth phases of wheat crop were obtained from the Department of Agrometeorology, N.D. University of Agriculture & Technology Kumarganj, Faizabad. The data were collected up to the first 15 weeks

of the crop cultivation which included 44th standard meteorological week (SMW) to 52nd SMW of a year and 1st SMW to 6th SMW of the next year. The data on five weather variables viz. minimum temperature, maximum temperature, relative humidity, wind-velocity and sunshine hours were 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.

Agrawal et al (2012) 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 yield in Faizabad district of Uttar Pradesh.

In order 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 year 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 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 $-s$ would form adverse group, the adjusted crop yields between $-s$ and $+s$ would form normal group and adjusted crop yields above or equal to $+s$ would form congenial group.

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 no. 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).

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

where Z_{ij} is un-weighted (for $j=0$) and weighted (for $j=1$) weather indices for i^{th} weather variable and 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. Here, $p=5$ and $n=15$, i.e. 15 weeks data from 44th week to 52nd week of a year and 1st week to 6th week of the next year 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 weather indices and 10 weighted interaction indices, and 5 un-weighted and 10 un-weighted interaction indices have been constructed. Besides, some more suitable strategies have been suggested. In all, six possible models are attempted. First two are the best models reported by Agrawal *et al.* (2012) and last four are newly proposed. Models are developed using regression analysis. Only the first 18 years data from 1991 to 2008 have been utilized for modeling the yield and remaining two years yield data of 2009 and 2010 have been used for validation of the models.

Model 1

This model is the 2nd model of Agrawal et al (2012). Using five weighted weather indices of five weather variables, discriminant function analysis was 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:

$$y = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + e$$

where y is un-trended crop yield, $\hat{\alpha}_i^s$ ($i = 0, 1, 2, 3$) are model parameter, ds_1 and ds_2 are two sets of discriminant scores, T is the trend variable and e is error term assumed to follow independently $N(0, \sigma^2)$. This model utilizes the complete data over 15 weeks and also considers relative importance of weather variables in different weeks.

Model 2

This model is 4th model of Agrawal et al (2012). Two discriminant functions and there from two sets of discriminant scores were obtained using the first week data (44th SMW) on five weather variables. Two sets of discriminant scores obtained from first week data and data on five weather variables in the second week (45th SMW) were used as discriminating variables, so in all there were 7 discriminating variables, and based on these 7 discriminating variables the discriminant analysis has been done and, therefore, two sets of discriminant scores were obtained. This process was repeated up to the last week till the time of forecast (6th SMW or 15th week) and finally two sets of discriminant scores (ds_1 and ds_2) were obtained. Based on these two sets of scores obtained at the 15th week, the forecasting model taking yield as the regressand and the discriminant scores and the trend as the regressor variables has been fitted similar to that given in Model 1.

Model 3

In this model, five weighted and five un-weighted weather indices of five weather variables were used as discriminating variables in the discriminant function analysis. Two sets of scores from two discriminant functions were obtained. The forecasting model were fitted taking the yield as the regressand and the two sets of scores (ds_1 and ds_2) and the trend as the regressors.

Model 4

In this model, all 30 indices (weighted and un-weighted including interaction indices) were used as discriminating variables in discriminant analysis and two sets of discriminant scores from two discriminant functions were obtained. Forecasting model were fitted taking un-trended yield as the regressand variable and the two sets of discriminant scores (ds_1 and ds_2) and the trend as the regressor variables.

Model 5

In this model, discriminant function analysis were carried out using the data on the first weather variable spread over 15 weeks using (44th SMW to 6th SMW of next year). Using two sets of discriminant scores obtained from two discriminant function of data on the first weather variable and 15 weeks data of the second variable, discriminant function analysis were again performed and two sets of discriminant scores were obtained (here the discriminating variables will now become 17). Using these two sets of discriminant scores and 15 week data of third variable were again used to carry out discriminant analysis and subsequently two sets of discriminant scores were obtained. This process was continued up to fifth weather variable, and ultimately we got two sets of discriminant scores ds_1 and ds_2 and forecast model was developed.

Model 6

In this model, discriminant function analysis were carried out using the unweighted and weighted averages (weather indices) for the first weather variable (here discriminating factors will be only two). Using the two sets of discriminant scores obtained on the basis of first weather variable and unweighted and weighted averages (weather indices) for the second weather variable, discriminant function analysis were further carried out (here, the discriminating factors will be four). This process was continued up to fifth weather variables, and finally we got two sets of discriminant scores ds_1 and ds_2 and forecast model was developed.

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

Table 1: Wheat yield forecast models

Model	Forecast regression equation	R ² (%)	R ² adj(%)
1.	Yield = 21.619 - 0.757**ds ₁ + 0.464 ds ₂ + 0.413**T (0.529) (0.158) (0.272) (0.049)	85.8	82.8
2.	Yield = 22.313 - 0.094**ds ₁ + 0.002 ds ₂ + 0.351**T (0.322) (0.010) (0.028) (0.030)	94.5	93.3
3.	Yield = 21.468 - 0.748**ds ₁ - 0.318ds ₂ + 0.422**T (0.483) (0.133) (0.170) (0.045)	88.4	86.0
4.	Yield = 22.227 - 0.123**ds ₁ + 0.063ds ₂ + 0.360**T (0.351) (0.015) (0.100) (0.032)	93.4	92.0
5.	Yield = 22.247 + 0.004*ds ₁ + 0.016**ds ₂ + 0.359**T (0.321) (0.001) (0.002) (0.030)	95.6	94.3
6.	Yield = 21.558 - 0.742**ds ₁ - 0.329ds ₂ + 0.436**T (0.537) (0.159) (0.195) (0.050)	85.4	82.2

Note: Figures in brackets denote Standard Error of regression coefficients.*P < 0.05,**P < 0.01

Table 2: Actual and forecasts of wheat yield (qha⁻¹ by models)

Year	Actual yield	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
2008-09	30.82 (0.62)	30.63 (0.68)	30.61 (1.68)	31.03 (0.32)	30.72 (0.19)	30.88 (1.56)	31.30
2009-10	28.32 (1.55)	28.77 (2.53)	27.60 (0.33)	28.03 (2.62)	27.68 (2.34)	27.66 (1.77)	28.38
2008-09	PSE (cv)	2.316	1.304	1.950	1.476	1.204	2.359
2009-10	PSE (cv)	2.977	1.567	2.064	1.735	1.511	2.219

Note: Figures in brackets denote % deviation of forecast. CV : coefficient of variation

deviation of forecast from actual, and percent standard error (SE).

RESULTS AND DISCUSSION

The forecast models developed under the six procedures along with R² and R²_{adj} 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 (ds₁) 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 (ds₂) has been found to be significant at P < 0.01 in only model 5. Adjusted coefficient of determination (R²_{adj}) was found to be maximum of 94.3

percent in model 5 while it was minimum (82.2%) in model 6. Based on these forecast models, the forecast yields for the 2008-09 and 2009-10 were obtained and the results are presented in Table 2. It is evident from the results of the Table 2 that percent deviation of forecast yield from actual yield varied from 0.19 (in model 5) to 2.62 (in model 4) over two years. The percent standard error (cv) of forecast yields have also been computed for all the models and are presented in the Table 2. The percent standard error (cv) of forecast has been found to be minimum, i.e. 1.204 for 2008-09 and 1.511 for 2009-10 in model 5 as compared to other models. The model 2 (2nd in the Table 1) suggested by Agrawal et al(2012) comes out to be the second best on account of high values of R² adj

and lower value of percent standard error (cv), and it is close to model 5. However, on the basis of the overall results of the Table 1 and 2, it can be concluded that the proposed model 5 is the most suitable model among all the models to forecast wheat yield in Faizabad district of Eastern Uttar Pradesh. Hence, a reliable forecast of wheat yield about two months before the harvest can be obtained from the proposed model 5.

REFERENCES

- Agrawal, R.; Jain, R.C. and Jha, M.P. (1983). Joint effects of weather variables on rice yields. *Mausam*, **32**(2): 177-81.
- Agrawal, R., Jain, R.C. and Jha, M.P. (1986). Models for studying rice crop weather relationship. *Mausam*, **37**(1): 67-70.
- Agrawal, R., Chandrahas and Aditya K. (2012) Use of discriminant function analysis for forecasting crop yield. *Mausam* **36** (3): 455-458.
- Anderson, T. W. (1984) An introduction to applied Multivariate Statistical Analysis, *John Wiley & Sons Inc.* New York.
- Draper, N.R. and Smith, H. (1998). Applied Regression Analysis. 3rd edition, *John Wiley & Sons Inc.* New York.
- Hair, Joseph F., Anderson, Rolph, E., Tatham, Ronald L., and Black, William C., (1995). "Multivariate Data Analysis with Readings", Prentice Hall, Inc New Jersey
- Johnson ,R.A. and Wichern, D.W. (2001). "Applied Multivariate Statistical Analysis" Third Edition, Prentice Hall of India Private Limited, New Delhi.
- Rai, T. and Chandrahas (2000). "Use of discriminant function of weather parameters for developing forecast model of rice crop". Publication of IASRI, New Delhi.
- Sharma, S.(1996). "Applied Multivariate Techniques", *John Wiley & Sons Inc.* New York.

Received : October 2012 ; Accepted : January 2014