International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2018; 6(4): 2205-2208

Received: 21-05-2018
Accepted: 26-06-2018

Annu
Department of Agricultural Statistics Narendra Deva University of Agriculture & Technology, Kumarganj, Faizabad, Uttar Pradesh, India

BVS Sisodia
Department of Agricultural Statistics Narendra Deva University of Agriculture & Technology, Kumarganj, Faizabad, Uttar Pradesh, India

Sunil Kumar
Department of Agricultural Statistics Narendra Deva University of Agriculture & Technology, Kumarganj, Faizabad, Uttar Pradesh, India

Correspondence
Annu
Department of Agricultural Statistics Narendra Deva University of Agriculture & Technology, Kumarganj, Faizabad, Uttar Pradesh, India

Pre-harvest forecast model for rice yield based on biometrical characters: An application of discriminant function analysis

Annu, BVS Sisodia and Sunil Kumar

Abstract
In the present paper, an application of discriminant function analysis has been demonstrated to develop pre- harvest forecast models based on biometrical characters for rice yield using experimental data. The results have shown that the reliable forecast of rice yield can be obtained using these models with percent standard errors of forecast yields below 5 percent.

Keywords: forecast model, biometrical characters, discriminant function, experimental data, and rice crop

1. Introduction

Rice is one of the major staple food of India being grown in about 44 million hectares land and its production touches about 100 million tonnes (2013-14). Uttar Pradesh ranks first in its area and second in production. However, productivity of rice in Uttar Pradesh (24.60 Q/ha) is almost at par with the national average (2012-13) but it ranks eighth position among the major producing States. Its share in area and production has been about 13.82 and 14.41 percent, respectively, of the total area and production in the country (2012-13). Frequent drought and floods in the various regions of the country make production of rice more vulnerable and this affects its area and production considerably. Therefore, reliable forecast of rice production before the harvest constitutes a problem of topical interest. Such forecast is needed by the Government, and other private and public sector for making policy decisions in regards to procurement, distribution, buffer- stocking, import- export, price fixation etc. Various researchers have made effort in the past to develop statistical models using time series data on the crop- yield and weather variables for pre- harvest forecast of crop yield. Notably among them are Agrawal *et al.* (1980, 1982, 1983, 1986, 2001) [2-7], Agrawal and Mehta (2007) [5], Jain *et al.* (1980), Kumar and Bhar (2005) [13], Khan *et al.* (2006), Patel *et al.* (2007), Singh *et al.*(2007), Mohd. Azfar *et al.* (2015), Yadav *et al.* (2014) [20] etc. Data on biometrical characters of crops from experiments or surveys have also been used to develop pre-harvest model for crop yield applying statistical tools like regression model and principal component regression model by Jain *et al.* (1985), Singh *et al.*(1986), Singh and Bapat (1988), Aneja *et al.*(2008) etc. Application of discriminant function analysis of weather indices and weekly data of weather variables for development of statistical models to forecast crop yield has been attempted by Yadav *et al.* (2008) [21], Agrawal *et al.* (2012) [8], Sisodia *et al.* (2014) [22] and Mohd. Azfar *et al.* (2014). An attempt has been made in the present paper to develop pre-harvest forecast model for rice yield using measurements on biometrical characters from experiment by applying discriminant function analysis.

2. Materials and statistical methods

2.1 Sources and description of data

The data on yield of rice and related biometrical characters were obtained from two experiments conducted at Main Experimental Station of Narendra Deva University of Agriculture & Technology Kumarganj, Faizabad U. P. India. The details of the experiment are described below.

Table 1: Detail of experiments

S. No	Experiment	Design	Treatment	Replication	Plot size	Date of sowing
1.	I	Randomized Block Design	21 Varieties	03	2.0m.X 3.0m.	10th August, 2011
2.	II	Randomized Block Design	28 Varieties	03	4.0m.X 2.0m.	16th August, 2011

The names of 21 varieties of rice in the experiment-I are as follows

1- NDRK-11-1, 2-NDRK-11-8, 3 - NDRK-11-9, 4 - NDRK-11-10, 5- NDRK-11-11, 6 -NDRK-11-5, 7- NDRK-11-6, 8-NDRK-11-12, 9- NDRK- 11-13, 10-NDRK- 11-14, 11 - NDRK - 11-15, 12- NDRK - 11-16, 13- NDRK - 11-17, 14-NDRK-11 –18, 15- NDRK- 11-18, 16- NDRK- 11-19, 17-NDRK- 11-4, 18- NDRK -11-20, 19- CSR-36, 20- SARJOO-52, 21- IR-28

The names of 28 varieties of rice in the experiment-II are as follows

1- RAU-1428, 2- RAU-1-16, 3- CR-2218-64-1, 4 -CR-2218-207, 5- CR-2461-1, 6 - CR-2462-1, 7-CR-2219, 8- CARI Dhan-2, 9-CARI Dhan -5, 10- NDRK-11-1, 11- NDRK-11-2, 12- NDRK-11-3, 13-NDRK-11-4, 14 - NDRK-11-5, 15-NDRK-11-6, 16- RP-4353, 17-RP-4631, 18-PNL-9, 19- CSR-2K-219, 20-CSR-2K-242, 21-CSR-2K-255, 22-CSR-2K-262, 23- Check (CST-7-1), 24-Check (CST-27), 25-Check (CST-36) 26- Pusa Sugandha-1121, 27- Pusa Sugandha, 28-Narendra Usar Dhan-3.

The following biometrical characters for rice were measured from each plot of the experiment using standard methods of measurement. Their average values corresponding to the each variety were used for the study.

1. X_1: Plant population /plot, 2. X_2: Plant height, 3. X_3: No. of tillers/plot, 4. X_4: Length of ear head/plant, 5. X_5: Green leaves/plant, 6. X_6: Basal girth, 7. X_7: No. of grains/ear head.

2.2 Development of pre- harvest forecast model using discriminant function analysis of biometrical characters

Discriminant function analysis is a multivariate statistical technique to describe the differential feature of objects form several known population such that the populations are separated as much as possible. The theory of the technique is available in many standard books on multivariate analysis like Johnson and Wichern (2014). However, how it has been used in the present study is describe below.

Let n be the number of varieties involved in an experiment. Let Y_i be the yield (q/ha) of the i^{th} variety (i= 1, 2, ...,n) and X_{ij} be the measurement on j^{th} biometrical character (j=1, 2,..., p) corresponding to i^{th} variety. The yields of varieties Y_i's are classified into three groups, viz., below normal, normal and above normal as given below:

(i) Below normal group: This group consists those varieties having yield below or equal to $\bar{y}$ - S.D. (ii) Normal group: This group consists of those varieties having yield between $\bar{y}$ - S.D. and $\bar{y}$ + S.D.

(iii) Above normal group: This group consists those varieties having yield greater or equal to $\bar{y}$ + S.D. , where $\bar{y} = \sum_{i=1}^{n} y_i / n$, average yield of varieties and S.D. is standard deviation of yields of varieties (y_i). Therefore it may be noted that the varietal effects are taken can of by grouping these varieties as. These groups are coded as 1, 2 and 3, respectively. These three groups were considered as 3 populations for carrying out discriminant function analysis. Using these codes for yields of varieties as per the above groups and the corresponding measures on biometrical characters, the discriminant function analysis will be carried out, and we get two estimated discriminant functions from three populations. These discriminant function are expressed as Anderson's classification function (Statistic) as

$$D_1 = \hat{a}_1 + \hat{l}_{11} x_1 + \hat{l}_{12} x_2 + \dots\dots\dots .. + \hat{l}_{1p} x_p \quad (1)$$

and

$$D_2 = \hat{a}_2 + \hat{l}_{21} x_1 + \hat{l}_{22} x_2 + \dots\dots\dots .. + \hat{l}_{2p} x_p \quad (2)$$

where x_{ij}'s are biometrical characters. a_i's and l_{ij}'s are estimated constants.

From these two estimated discriminant functions, we compute two sets of discriminant scores, say ds_1 and ds_2, each consisting of n scores corresponding to n verities. Using ds_1 and ds_2 as regressor variables, and variety yield (y) as regressand, the following multiple linear regression model is considered for pre- harvest forecast of yield.

$$y_i = \beta_0 + \beta_1 ds_{1i} + \beta_2 ds_{2i} + e_i, \; i=1, 2,....n \quad (3)$$

where y_i is yield of the crop for i^{th} variety; β_0 , β_1 and β_2 are parameters of the model and e_i is the error term assumed to follow independently normal distribution with mean 0 and variance σ^2. The model (3) will be fitted with the data by ordinary least square technique. The data yield of 18 varieties from experiment- I and 25 varieties of experiment –II will be use for the fitting of the model (3) and the data on yield of the remaining of last three varieties of both the experiments will be use for the validation of the model.

2.3 Measures for Validation of the forecast model

Different statistical measures have been used for the validation of the model, which are given bellow.

(i) Coefficient of Determination (R^2)

It is in general used for evaluating the adequacy of the model. R^2 is given as

$$R^2 = 1 - \frac{SS_{res}}{SS_t}$$

where SS_{res} and SS_t are the residual sum of square and the total sum of squares, respectively, in analysis of variance of regression model.

(ii) Percent Deviation of forecast yield from actual yield

The percent deviation of forecast yield is computed as follows.

$$\text{Percent deviation of forecast yield} = \frac{(\text{actual yield} - \text{forecasted yield})}{(\text{actual yield})} \text{X } 100$$

(iii) Root Mean Square Error (RMSE)

It is also a measure validation and comparing two models. The formula of RMSE is given bellow

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

where o_i and the e_i are the observed and forecasted value of the crop yield, respectively, and n is the number of years for which forecasting has been done.

(iv) Percent Standard Error (PSE) of the forecast

Let $\hat{y}_f$ be forecast value of crop yield and X_0 be the column vector of values of P independent variables at which y is forecasted. Variance of $\hat{y}_f$ is given by (Draper and Smith, 1998).

$$V(\hat{y}_f) = \hat{\sigma}^2 X_0^{'} (X'X)^{-1} X_0$$

where $(X'X)$ is the matrix of the sum of square and cross products of regressors matrix X (independent variables) and $\hat{\sigma}^2$ is the estimated residual variance of the model. Therefore, the percent standard error (cv) of forecast is given by

$$PSE = \frac{\sqrt{V(\hat{y}_f}}{\text{forecast value}} \times 100$$

3. Results and Discussion

The average yield of the varieties $(\bar{y})$ along with their standard deviation (S.D.) were obtained to be 22.50 q/ha (S.D.=2.79) and 27.92 q/ha (S.D.= 5.42) with respect to experiment-I & II, respectively. The varieties of both experiments were classified into three population as per procedure described in Sub Section 2.2.The details of the populations are as follows.

Experiment	Ist Population below normal	IInd Population Normal	IIIrd Population above normal
I	Varieties having yield below 19.71 q/ha	Varieties having yield between 22.36 and 25.29 q/ha	Varieties having yield more than 25.29q/ha
II	Varieties having yield below 22.50 q/ha	Varieties having yield between 22.50 and 33.34 q/ha	Varieties having yield more than 33.34q/ha

The above mentioned three populations have been coded as 1, 2 and 3.Using these codes for three populations and measures on biometrical characters X_1, X_2, X_3, X_4, X_5, X_6 and X_7, discriminant function analysis has been carried out for both experiments. We got finally two discriminant functions for each experiment. From these two discriminant functions, two sets of discriminant scores, ds_1 & ds_2 for corresponding varieties have been computed. The estimated discriminant functions and set of score are presented in Appendix- I.

Using these two sets of discriminant scores as regressor variables and rice yields of varieties as regressand, the model-3 for both the experiments has been fitted with the data by applying ordinary least square technique. The fitted models for both the experiments along with the values of R^2 are given in the Table 1.

Table 1: Forecast model for rice experiment I & II

Experiment	Forecast model	R^2 (%)
I	Yield= 25.25+.977**ds_1+.696ds_2 (.490) (.265) (.407)	80.00**
II	Yield= 27.39+2.84**ds_1-0.10ds_2 (.73) (.43) (.53)	66.20**

Note: Figures in bracket denote standard error of regression coefficient. ** Significant at $p \leq 0.01$

The forecast yield of rice for remaining three varieties of the experiment- I & II were compute using these forecast models and are presented in the Table -2. The values of percent deviation of forecast, RMSE and percent standard error (PSE) of forecast were also computed and are also given in the Table-2 along with actual yield of varieties of rice for which forecast were made.

Table 2: Actual and forecast yield of rice based on rice experiment- I &II

Experiment	Actual Yield (q/ha)	Forecast Yield (q/ha)	PSE(CV)	RMSE
I	(i) 28.00	27.77(0.82)	3.97	1.05
	(ii) 23.56	25.07(6.40)	4.41	
	(iii) 21.85	22.85(4.57)	4.43	
II	28.82	28.08(2.57)	3.64	1.93
	29.31	27.43(6.42)	2.72	
	26.59	29.26(10.03)	2.96	

Note: Figure in brackets denote % deviation of forecast, CV: Coefficient of variation

It can be observed from the Table- 1 that the first discriminant score (ds_1) showed significant effect on the yield in both the experiments $p \leq 0.01$. However, the second discriminant score showed positive but not significant in the experiment - I but negative in experiment - II. The values of coefficient of determination (R^2) for the model in the experiment - I & II have been found to be reasonably appropriate, i.e. about 80 and 66 percent, respectively.

The perusal of the Table 2 reveals that the forecast models for both the rice experiments have performed well as the values of PSE (CV) of forecast were found reasonably below 5 percent in both the experiments and it ranged between 2.77 and 4.43 percent. The values of RMSE were found to be 1.05 and 1.93 for the models in the experiment-I & II, respectively. The values of percent deviation of forecast were found to be little smaller in case of experiment - I as compared to experiment –II. It can also be observed from the Table -2 that the forecast yields of rice were in general quite close to actual yield. Agrawal *et al.* (2012) and Sisodia *et al.* have developed various forecast models for wheat yield using time series data

on yield and week weather variable s by applying discriminant function analysis. They found that the percent standard error (PSE) of forecast yield based on the models were below 5 percent with reasonably high values of R^2, more than 80 percent. Moh. Azfar *et al.* (2014) have also applied the technique of discriminant function analysis to develop forecast models for yield of rapeseed & mustard using time series data on yield and weather variables. They found that the values of R^2 were reasonably high about more than 80 percent but $PSE_{(s)}$ of forecast yields ranged between 3.98 and 10.67. Yadav *et al.* (2014) and Moh. Azfar *et al.* (2015) have also developed forecast models for wheat and rapeseed & mustard yield, respectively, using time series data on yield and weather variables by applying technique of principal component analysis of weather data. The $PSE_{(s)}$ for forecast wheat yield were found to be below 5 percent but for rapeseed & mustard yield, it varied between 3.96 and 15.59 percent.
However, the present study has dealt with experimental data in order to develop forecast models by applying first time the technique of discriminant function analysis of biometrical characters. The results are almost similar as above of forecast yield with $PSE_{(s)}$ below 5 percent.

4. Conclusion

In views of the above discussion of the results and various measures of validation of the model, i.e. R^2, RMSE, PSE (CV) and percent deviation of forecast yield from actual yield as presented in the Table -2, it can also be concluded that the reliable forecast of rice yield can be obtained from the model obtained by applying discriminant analysis of data on biometrical characters of the experiments.

5. Acknowledgement

The authors are thankful to Dr. O.P. Verma, Assistant Professor & Rice Breeder, of N.D.U.A. & T. Kumarganj Faizabad for providing the data of rice experiment for the present study.

6. References

1. Agrawal R, Jain RC, Singh D. Forecasting of rice yield using climate variables. Ind. J Agri. Sci. 1980; 50(9):680-684.
2. Agrawal R, Jain RC. Composite model for forecasting rice yield. Ind. J Agri. Sci. 1982; 52(3):177-181.
3. Agrawal R, Jain RC, Jha MP. Joint effects of weather variables on rice yields. Mausam. 1983; 34(2):177-181.
4. Agrawal R, Jain RC, Jha MP. Models for studying rice crop – weather relationship. Mausam. 1986; 37(1):67-70.
5. Agrawal R, Jain RC, Mehta SC. Yield forecast based on weather variable and agriculture inputs on agro-climate zone basis. Ind. J Agri. Sci., 2001; 71(7):487 -490.
6. Agrawal R, Mehta SC. Weather based forecasting of crop yields pest and diseases - IASRI model. J Ind. Soci. of Agril. Stat. 2007; 61(2):255-263.
7. Agrawal R, Chandrahas, Kaustav A. Use of discriminant function analysis for forecasting crop yield, Mausam. 2012; 63(3):455-458.
8. Aneja DR, Rai Lajpat, Grover Deepak, Batra SD. Pre-harvest forecast models for cotton yield. Journal of Cotton Research and development. 2008; 22(1):129-134.
9. Jain RC, Agrawal R, Jha MP. Effect of climate variables on rice yield and its forecast. Mausam. 1980; 31(4):591-596.
10. Jain RC, Sridharan H, Agrawal R. Principal component technique for forecasting of sorghum yield. Indian Journal of Agril. Sci. 1985; LI(1):61-72.
11. Johnson RA, Wichren DW. Applied multivariate statistical analysis 6th edition. PHI Learning Private Limited, New Delhi, 2014.
12. Khan SA, Bhowmick M, Das L. Predicting seed yield of rape and mustard by agro meteorological parameters. Journal of Interacademicia. 2006 10(1), p54-59.
13. Kumar A, Bhar L. Forecasting model for yield of Indian mustard (*Brassica juncea*) using weather parameter. Ind. J Agri. Sci. 2005; 75(10):688-690.
14. Mohd. Azfar, Sisodia BVS, Rai VN, Devi Monika. Pre-harvest forecast of rapeseed & mustard yield based on weather variables: An application of discriminant function analysis. Int. J Agri. and Statistical Science. 2014; 10(2):497-502.
15. Mohd. Azfar, Sisodia BVS, Rai VN, Devi Monika. Pre-harvest forecast models for rapeseed & mustard yield using principal component analysis of weather variables. Mausam. 2015; 66(4):761-766.
16. Patel GB, Vaishnav PR, Patel JS, Dixit SK. Pre-harvest forecasting of rice (*Oryza sativa* L.) yield based on weather variables and technological trend. Journal of Agrometeorology. 2007; 9(2):167-173.
17. Singh BH, Bapat SR. Pre-harvest forecast models for prediction of sugarcane yield. Indian Journal of Agricultural Sciences. 1988; 58(6):465-469.
18. Singh D, Singh HP, Singh P. Pre- harvest forecasting of wheat yield. Ind. J of Agric. Sci. 1986; 46(10):445-450.
19. Singh H, Singh KN, Hassan B, Khan AA. Agro- climatic models for prediction of growth and yield of rice (*Oryza sativa*) under temperature Kashmir condition. Ind. J Agri. Sci., 2007; 80(3):254-257.
20. 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.
21. Yadav HK, Shukla S, Singh SP. Discriminant function analysis for opium and seed yield in opium poppy (*Papaver somniferum* L.), Genetika. 2008; 40:109-120.
22. Yadav RR, Sisodia BVS, Kumar Sunil. Application of principal component analysis in developing statistical models to forecast crop yield using weather variables. Mausam. 2014; 65(3):357-360.