An application of principal component analysis for pre- harvest forecast model for rice crop based on biometrical characters

Annu[1*], B.V.S. Sisodia[1] and V. N. Rai[1]

[1]Department of Agricultural Statistics, Narendra Deva University of Agriculture & Technology, Kumarganj, Faizabad-224229 (U.P.), INDIA
*Corresponding author. E-mail: annustat@gmail.com

Abstract An application of principal component analysis for the development of suitable statistical models for pre-harvest forecast of rice yield based on biometrical characters has been dealt with in the present paper. The data obtained from the two experiments on rice have been utilised to develop the model. The forecast yields of based on these models have been found to be 24.25, 22.60 and 21.10 q/ha against the actual yield of 28.00, 23.56 and 21.85 q/ha, respectively, in experiment –I. For experiment –II the forecast yields were found to be 24.62, 28.06 and 29.43 q/ha against the actual yield of 28.82, 29.31 and 26.59 q/ha, respectively. These forecast yields are subject to maximum of almost 10 percent standard error. In most of the cases, the forecast yields were found to be close to the actual yield except in some cases. The values of R^2, i.e. 79.80 and 72.60 for experiment –I and II, respectively, indicate the validity of the models. Statistical tool like viz. principal component analysis (PCA) has been first time applied to develop pre-harvest forecast model based on experimental data.

Keywords : Biometrical characters, Pre-harvest forecast model, Principal component Analysis, Rice experiment

INTRODUCTION

A reliable forecast of crop yield before the harvest are needed by the Government for making policy decision in regards to procurement, distribution, buffer- stocking, import- export, and marketing of agricultural commodities, while agro –based industries, traders and agriculturists need them for proper planning of their operations. Various research workers in the recent past have developed pre-harvest forecast models for several crops based on time series data on crop yield and weekly data on weather variables. Notably among them are Agrawal et al. (1980, 1983, and 1986), Singh and Bapat (1988), Singh et al. (1986), Yadav et al. (2014), Mohd. Azfar et al. (2014), Pandey et al. (2014), Azfar et al. (2015), Yadav et al. (2015) and Annu et al. (2015). Jain et al. (1984, 1985, 1992b) have developed statistical models for forecasting crop yield based on biometrical characters using experimental and survey data in different regions of the country. Rice and wheat is major cereal crop of the Eastern Uttar Pradesh. An attempt has been made in the paper to develop pre-harvest forecast model for rice yield using experimental data by applying the technique of principal component analysis in Faizabad district of Eastern Uttar Pradesh. The explanatory variables used in the general regression model are generally correlated and it may create problem in estimating model parameters. PCA of explanatory variables provides principal component (PC) which are independent.

These PC(s) are used in the model as explanatory variables and model parameters and easily estimated with accuracy for pre-harvest forecasting of crop yield.

MATERIALS AND METHODS

Study area: The present study is related to Faizabad district (Eastern Uttar Pradesh, India) which is situated between 26^0 47' N latitude and 82^0 12' E longitudes. It lies in the Eastern plain zone of Uttar Pradesh. It has an annual rainfall of about 1002 mm. Nearly 85% of total precipitation is received from south- west monsoon during the month of July to September. However, occasional mild shower occur during winter season. The average minimum temperatures are 18.6 ^{0}C and 31.3^0C, respectively. It is liberally sourced by the Saryu (Ghaghara) river and its tributaries. Soils are deep alluvial, medium to medium heavy textured but are easily ploughable. The favourable climate, soil and the availability of ample irrigation facility make growing of wheat and rice a natural choice for the area. Wheat crop is generally cultivated during the Rabi season and rice is cultivated during Kharif session.

 Sources and description of data : The data on yield of rice and related biometrical characters were obtained from different experiments conducted at Main Experimental Station of Narendra Deva University of Agriculture and Technology Kumarganj, Faizabad U. P. India. The details of the experiment are described below. The name of 21 varieties of rice in experiment –I are

as under.

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 name of 28 varieties of rice in experiment –II are as under.

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 was measured from each plot of the experiment following standard methods of measurement

1. Plant population /plot 2. Plant Height 3. No. of tillers/plot 4. Length of ear head/plant 5. Basal Girth 6. Green leaves/plant 7. No. of grain/ear head

Development of preharvest forecast model using orincipal component analysis: PCA is a multivariate technique for data reduction. It is a mathematical function, which does not require user to specify the statistical model or assumption about distribution of original variables. It may also be mentioned that principal components are artificial variables and often it is not possible to assign physical meaning to them. Further, since principal component analysis transforms original set of correlated variables to new set of uncorrelated variables, it is worth stressing that if original variables are uncorrelated, than there is no point in carrying out principal component analysis. The theory of principal component analysis is available in many standard books on multivariate analysis (Anderson, et al., 1974) so its theoretical aspects are not presented here. Let x_{ij} be the value of jth biometrical character (j= 1, 2,...p) corresponding to i^{th} varieties of experiment (i= 1, 2,...n). The PCA for x_{ij}'s will be carried out. Let PC_1, PC_2, ...PC_K be first K (K< P) principal components explaining variability about 90 percent of the total variation in x_{ij}'s. Using these K principal components as regressor variables and varieties yield (y_i) as regressand, the following linear multiple regression model for pre-harvest forecast of crop yield has been proposed.

$$y_i = \beta_0 + \beta_1 PC_{1i} + \beta_2 PC_{2i} + ... \beta_K PC_{Ki} + e_i \quad , i= 1, 2, ...n.$$

where y_i is the crop yield of the ith plot;

$\beta_0, \beta_1, \beta_2 ... \beta_K$ are model parameter and e_i is error term assumed to follow independently normal distribution with mean 0 and variance σ^2. The aforesaid model is fitted with the data by least square technique.

Measures for validation and comparison of the models Different procedures have been used for the comparison and the validation of the models developed. These procedures are given bellow.

Coefficient of determination (R^2 and R^2_{adj}) The models were validated on the basis of R^2 and R^2_{adj} which can be computed from the formula given in Drapper and Smith (1988)

For computation of Percent Deviation of forecast yield from actual yield, Root Mean Square Error (RMSE) and Percent Standard Error of the forecast yield, the formulae are given in Mohd. Azfar et al. (2015) and Yadav and Sisodia (2015).

RESULTS AND DISCUSSION

Using the data on plant biometrical characters X_1, X_2, X_3, X_4, X_5, X_6 and X_7 the principal component analysis has been carried out for the data of both the experiments. Since first five principal components have explained about 87.86 percent of total variability in the both the experiments of rice, these first five principal components have been used as regressor variables and varieties yield as regressand in the development of the model. The yields of first 18 and 25 varieties of rice in experiment-I and II, respectively were used for model fitting and the yields of last three varieties of the experiments were used for validation of the models. The model has been fitted with the data of the rice experiments–I and II by applying least square technique. The fitted models along with values of R^2 and R^2_{adj} are presented in Table 2.The forecast yield for remaining three varieties of the rice experiment – I and II were computed by applying the forecast models given in the Table 2. The values of percent deviation of forecast, RMSE and PSE (CV) for each forecast from the models have been computed and are presented in the Table 3 along with actual and forecast yield.

For experiment- I, it can be observed from the Table-2 that only first principal component (PC_1) has shown positive significant effect at 1 percent significant level on the yield. The values of R^2 and R^2_{adj} have been obtained reasonably high, i.e. 79.80 and 71.40 percent, respectively. For experiment-II, the first principal component (PC_1) has shown positive significant effect at 1 percent significant level while fifth principal component (PC_5) has shown negative significant effect at 5percent level of significant. The values of R^2 and R^2_{adj} have been obtained to be 72.60 and 61.20 percent, respectively. However, the relationship between principal components and yields do not exhibit actual relationship between yield (Y) and biometrical characters ($X_i's$). The perusal of the Table 3

Table 1. Detail of experiments.

Sr. 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

Table 2: Forecast model for rice experiment I & II.

Experiment	Forecast model	R^2 (%)	R^2_{adj} (%)
I	Yield= 24.89+ 2.213** pc_1-0.210pc_2 (0.590) (0.359) (0.416) +1.171pc_3+0.897pc_4+0.788pc_5 (0.650) (0.510) (0.638)	79.80**	71.40
II	Yield= 27.55+ 1.57** pc_1-0.6920pc_2 (0.809) (0.529) (0.763) +1.453pc_3+2.143pc_4-2.312*pc_5 (0.769) (0.813) (0.946)	72.60**	61.20

Note: Figures in bracket denote Standard Error of regression coefficient, *Significant at $p \leq 0.05$ ** Significant at $p \leq 0.01$

Table 3. Actual and forecast yield of rice on based rice experiment I &II.

Experiment	Actual Yield (q/ha)	Forecast Yield (q/ha)
I	(i) 28.00	24.25(13.38)
	(ii) 23.56	22.60 (4.03)
	(iii) 21.85	21.10 (3.38)
	RMSE	2.27
	PSE(CV)	6.75
	PSE(CV)	9.27
	PSE(CV)	10.15
II	(i) 28.82	24.62(14.53)
	(ii) 29.31	28.06(4.26)
	(iii) 26.59	29.43(10.68)
	RMSE	3.01
	PSE(CV)	9.57
	PSE(CV)	9.91
	PSE(CV)	9.23

Note: Figure in Brackets denoted % deviation of forecast, CV: Coefficient of variation

reveals that the proposed models based on R^2 biometrical characters by applying principal component analysis has provided forecast yield very close to the actual yield of rice in both the situations of normal and late sowing of rice. The percent standard errors of the forecast in both situations have been found to be below 10 percent. Mohd. Azfar et al (2015) has shown that standard errors of forecast for rapeseed and mustard yield using PCA of weather variable have been found to be below 15 percent. In another study by Mohd. Azfar et al (2014), it has been shown that standard errors of forecast for rapeseed and mustard yield were below 10 percent when application of discriminant function analysis of weather variables was used, Yadav et al (2014) have found standard errors of forecast of wheat below 5 percent while applying technique of PCA of weather variables. The percent RMSE has been also found to be below 3 percent in the pre-

sent study. Thus, on the basis of the overall results of the Table 2 and 3, it can be concluded that the application of technique of PCA has provided a suitable forecast model using biometrical characters. Therefore, the proposed model can be used to obtain reliable preharvest forecast of rice yield in both the situations if the proper measurement on biometrical characters under consideration are available.

Conclusion

The results and discussion as above reveals that the application of technique of principal component analysis of biometrical characters has provided suitable forecast model with standard error of forecast yield below 10 percent and RMSE below3.00 percent. Therefore, it may be concluded that the proposed model can provide reliable forecast of the rice yield in both early and late sowing of rice with proper measurement of biometrical characters.

ACKNOWLEDGEMENTS

REFERENCES

Agrawal, R., Jain, R.C. and Jha, M.P. (1986). Models for studying rice crop weather relationship. *Mausam,* 37 (1), 67-70.

Agrawal, R., Jain, R.C., Jha, M.P. and Singh, D. (1980). Forecasting of rice yield using climatic variables. *Ind. J. Agric. Sci.,* 50 (9), 680-684.

Agrawal, R.; Jain, R.C. and Jha, M.P. (1983). Joint effects of weather variables on rice yields. *Mausam,* 34 (2),189-194.

Anderson, T. W. (1974) An introduction to multivariate statistical analysis 2nd edition .Wiley eastern private limited New Delhi.

Annu, Sisodia, B.V.S. and Kumar, Sunil (2015). Pre-harvest forecast models for wheat yield based on biometrical characters. *Economic Affairs,* 60(1):89-93.

Drapper, N. R. and Smith, H. (1988). Applied regression analysis, second edition John Willy and sons, New York.

Jain, R. C., Agrawal, Ranjana and Singh, K.N.(1992b). A within year growth model for crop yield forecast. *Biometrical Journal*. 34(7), 789-799.

Jain, R. C., Sridharan, H. and Agrawal Ranjana (1984). Principal component technique for forecasting of sorghum yield. *Indian Journal of Agril. Sci*. 54 (6), 467-470.

Jain, R.C., Sridharan, H. and Agrawal, Ranjana, (1985),. Principal component technique for forecasting of sorghum yield. *Ind. J. Agric. Sci.,* 54 (6), 467-470.

Mohd. Azfar, Sisodia, B. V. S., Rai, V. N. and Devi, Monika (2014). Pre-harvest of rapeseed and mustard yield based on weather variables- An application of discriminant function analysis. *Int. J. Agri. and Statistical Science*, 10(2): 497-502.

Mohd. Azfar, Sisodia, B. V. S. Rai, V. N. and Devi, Monika (2015). Pre-harvest of rapeseed & mustard yield based on weather variables- An application of principal component analysis of weather variables. Mausam Vol. 66 (4):761-766.

Pandey, K. K., Rai, V. N. and Sisodia, B. V. S. (2014). Weather variable based rice yield forecasting models for Faizabad district of eastern Uttar Pradesh. Int. *J. Agri. and Statistical Science*, 10(2): 381-385.

Singh, B. H. and Bapat, S. R. (1988)..Pre-harvest forecast models for prediction of sugarcane yield. *Indian Journal of Agricultural Sciences*, 58(6): 465-469.

Singh, D., Singh, H.P. and Singh, P (1986). Pre- harvest forecasting of rice yield. *Ind. J. of Agric. Sci.* 46, (10): 445-450.

Yadav, R. R., and Sisodia, B. V. S.(2015). Predictive models for Pigeon-pea yield using weather variables. *Int. J. Agri. and Statistical Science,* 11(2): 462-472

Yadav, R. R., Sisodia, B. V. S. and Kumar Sunil (2014). Application of principal component analysis in developing statistical models to forecast crop yield using weather variables. *Mausam,* 65 (3): 357-360.