Visit us - www.researchjournal.co.in ■ DOI : 10.15740/HAS/IRJAES/8.1/83-87
International Research Journal of Agricultural Economics and Statistics
Volume 8 | Issue 1 | March, 2017 | 83-87 ■ e ISSN-2231-6434

Research Paper

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

■ ANNU, B.V.S. SISODIA AND V. N. RAI

See end of the paper for authors' affiliations

Correspondence to :
ANNU
Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, FAIZABAD (U.P.) INDIA
Email : annustat@gmail.com

Paper History :
Received : *01.01.2017;*
Revised : *20.01.2017;*
Accepted : *30.01.2017*

ABSTRACT : An application of principal component analysis for the development of suitable statistical models for pre-harvest forecast of wheat yield based on biometrical characters has been illustrated in the present paper. The data obtained from the experiments on wheat under normal and late sowing situations have been utilised to develop the model. The result have revealed that the proposed model can provide reliable pre- harvest forecast of wheat yield in both the situations within the reasonable range of per cent standard error of 2.16 to 4.96 per cent.

KEY WORDS : Principal component analysis, Biometrical characters, Pre-harvest forecast model, Wheat experiment

HOW TO CITE THIS PAPER : Annu, Sisodia, B.V.S. and Rai, V.N. (2017). An application of principal component analysis for pre- harvest forecast model for wheat crop based on biometrical characters. *Internat. Res. J. Agric. Eco. & Stat.*, **8** (1) : 83-87, **DOI : 10.15740/HAS/IRJAES/8.1/83-87.**

INTRODUCTION :

A reliable forecast of crop yield before the harvest is required 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 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); Mohd. Azfar *et al.* (2015); Yadav *et al.* (2015) and Annu *et al.* (2015). Jain *et al.* (1984, 1985 and 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. The explanatory variables used in the general regression model are generally correlated and it may create problem in estimating model parameters. Principal component analysis (PCA) of explanatory variables provides principal components (PC) which are independent. First few $PC_{(s)}$ which explain maximum variability in explanatory variables are used in the model as explanatory variables and model parameters are easily estimated with reasonable.

Therefore an attempt has been made in the paper to develop pre-harvest forecast model for wheat yield using experimental data by applying the technique of principal component analysis in Faizabad district of Eastern Uttar Pradesh.

MATERIALS AND METHODS :

The materials used and the methodologies employed for develop of forecast model based on biometrical characters are described below.

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 per cent 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 a natural choice for the area. Wheat crop is generally cultivated during the *Rabi* season.

Source and description of data :

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

The 25 varieties of wheat were same in the both the experiments. The name of varieties are

1- AKDW4021, 2- NIAW1415, 3- DBW46, 4- USA316, 5- DBW51, 6- DBW52, 7- HD2864, 8- HD2932, 9-HD2997, 10-HD2985, 11- HI1563, 12- HI869, 13-HI977, 14-HUW234, 15- HW5207, 16- MP4010, 17- MP4106, 18- MACS3742, 19-NW4035, 20-PDW317, 21-PBW590, 22-PBW621 23- PBW315, 24-RSP561 25-WHD943

The following biometrical characters were measured by standard techniques used by Plant breeders and agronomists.

1. Plant population /plot (X_1) 2. Plant height (X_2), 3. No. of tillers/plot (X3), 4. Length of ear head/plant (X_4), 5. Basal girth(X_4), 6. Green leaves/plant(X_6) 7. No. of grain/ear head (X_7).

Development of pre- harvest forecast model using principal component analysis :

Principal 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, and then 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, 1974 and Johnson and Wichern, 2001). So its theoretical aspects are not presented here.

Let x_{ij} be the value of j^{th} biometrical character (j= 1, 2,...P) corresponding to i^{th} variety of experiment (i= 1, 2,....n). The principal component analysis 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 per cent of the total variation in x_{ij}’s. Using these K principal components as regressor variables and variety yield (y_i) as regressand, the following linear multiple regression model for pre-harvest forecast of crop yield has been proposed.

$$y_i = {}_0 + {}_1PC_{1i} + {}_2PC_{2i} + \ldots\ldots\ldots\ldots {}_kPC_{Ki} + e_i, \; i= 1, 2, \ldots\ldots n$$

where y_i is the crop yield of i^{th} variety; β_0, β_1, β_2, ... β_k are model parameter and e_i is error term assumed to follow independently normal distribution with mean 0 and variance σ^2.

Table A : Details of the experiments

Sr. No.	Experiment	Design	Treatment	Replication	Plot size	Date of sowing
1.	Experiment-I	Simple lattice design	25 varieties	2	4.5 x3.0m	25^{th}Nov. 2010 (normal sowing period)
2.	Experiment-II	Simple lattice design	25 varieties	2	3.5 x2.5m	25^{th}Dec. 2010 (late sowing period)

The aforesaid model is fitted with the data by ordinary least square technique.

Measures for validation the model:

Different measures for the validation of the model are given bellow.

Co-efficient of determination (r^2) :

The co-efficient of determinant, *i.e.* R^2 is compute by

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

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

Per cent deviation of forecast yield from actual yield:

The per cent deviation of forecast yield from actual yield has been computed as

$$\text{Per cent deviation of forecast yield} = \frac{\text{Actual yield - Forecast yield}}{\text{Actual yield}} \times 100$$

Root mean square error (RMSE):

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

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

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.

Per cent standard error 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 then 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 per cent standard error (CV) of forecast is given by

$$\text{Per cent S.E.} = \frac{\sqrt{V(\hat{y}_f)}}{\text{Forecast value}} \times 100$$

RESULTS AND DATA ANALYSIS :

Using the data on 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 experiments. The results of the principal component analysis for experiment –I and II are given in the Table 1. Since first five principal components have explained about 93 per cent of the total variability (Table 1), these first five principal components have been used as regressor variables and variety yield as regressand in the development of the model. The model has been fitted with the data of the experiment – I and II using yield of first 22 varieties by applying ordinary least square technique. The yields of last three varieties were left for validation of the model. The fitted models along with value of R^2 are presented in the Table 2.

The forecast of wheat yield for the remaining three varieties of the wheat experiment –I and II were computed by applying the forecast models given in the Table 2. The per cent deviation of forecast, RMSE and per cent standard error (CV) of each forecast yield for both the experiments were computed and are presented in the Table 3 along with actual and forecast yield.

For experiment –I it can be observed form the Table 2 that the first principal component (PC_1) and fourth

Table 1 : Principal component analysis

	F_1	F_2	F_3	F_4	F_5	F_6	F_7
Experiment-I							
Eigenvalue	2.107	1.760	1.182	0.886	0.584	0.274	0.208
Variability (%)	30.094	25.139	16.893	12.654	8.345	3.907	2.968
Cumulative %	30.094	55.233	72.126	84.780	93.125	97.032	100.000
Experiment-II							
Eigenvalue	2.484	1.413	1.137	0.836	0.601	0.369	0.160
Variability (%)	35.487	20.189	16.240	11.940	8.586	5.275	2.283
Cumulative %	35.487	55.677	71.917	83.857	92.443	97.717	100.000

NB: F_i's are factors (explanatory variable)

Table 2: Forecast model for wheat based on experiment I and II

Experiment	Forecast model	R^2 (%)
I	Yield = 35.747 + 0.464** pc_1 - 0.086pc_2 - 0.884pc_3 + 1.812**pc_4	63.80*
	(0.462) (0.309) (0.334) (0.461) (0.505)	
	-1.173*pc_5	
	(0.570)	
II	Yield = 32.142 + 0.9.62*pc_1 - 0.645pc_2 + 0.136pc_3 + 1.395*pc_4	62.10*
	(0.444) (0.271) (0.375) (0.391) (0.463)	
	+0.332pc_5	
	(0.569)	

Note: Figures in bracket denote standard error of regression co-efficient. * and ** indicate significance of values at P≤0.05 and ≤0.01,respectively

Table 3 : Actual and forecast yield of wheat based on wheat experiment I and II

Experiment	Actual yield (q/ha)	Forecast yield (q/ha)	RMSE	PSE(CV)
I	37.50	36.10 (3.37)	1.11	4.96
	35.00	35.50 (1.43)		3.10
	35.70	34.46 (3.47)		3.80
II	30.00	31.43 (4.70)	0.96	3.46
	30.50	31.05 (1.80)		2.81
	32.50	33.16 (2.03)		2.16

Note: Figure-s in brackets denote % deviation of forecast, CV: Co-efficient of variation

principal component (PC_4) showed positive significant effect on the yield. However, the fifth principal component (PC_5) showed negative and significant effect on wheat yield. However, they do not carry any physical meaning about the relationship between y and x_{ij}'s (biometrical characters). The value of R^2 has been found to be 63.80, have which in reasonably appropriate. For experiment-II, the first principal component (PC_1) and fourth principal component (PC_4) showed positive significant effect on the yield of wheat. The values R^2 was found to be 62.10 per cent (Table 2).

The perusal of the Table 3 reveals that the proposed models based on biometrical characters by applying principal component analysis has provided forecast yield very close to the actual yield of wheat in both the situations of normal and late sowing of wheat. The per cent standard errors of the forecast yield in both the situations have been found to be within the reasonable range of 2.16 to 4.96 per cent. The per cent RMSE has been also found to be below 1.11 per cent. Thus, on the basis of the overall results of the Table 2 and 3, it can be concluded that the application of technique of principal component analysis has provided a suitable forecast model using biometrical characters. Therefore, the proposed model can be used to obtain reliable pre-harvest forecast of wheat yield in both the situations if the proper measurements on biometrical characters under consideration are available.

Acknowledgement :

The authors are thankful to Dr. B.N. Singh, Wheat Breeder, of N. D. U. A. and T. Kumarganj, Faizabad for providing the data of wheat experiment for the present study.

Authors' affiliations:

B.V.S. SISODIA AND V.N. RAI, Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, FAIZABAD (U.P.) INDIA

LITERATURE CITED :

Agrawal, R., Jain, R.C., Jha, M.P. and Singh, D. (1980). Forecasting of rice yield using climatic variables. *Indian 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.

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

Anderson, T.W. (1974). An introduction to multivariate statistical analysis 2nd Ed. Wiley Eastern Private Limited, NEW DELHI, INDIA.

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

Draper, N. R. and Smith, H. (1981) *Applied regression analysis*, II[nd] Ed. John Willy and Sons, NEW YORK, U.S.A.

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

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

Johnson, R.A. and Wichern, D.W. (2001). *Applied multivariate statistical analysis*. 3[rd] Ed., Prentice-Hall of India.

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. *Internat. J. Agric. & Stat. Sci.*, **10**(2): 497-502.

Mohd. Azfar, Sisodia, B.V.S., Rai, V.N. and Devi, Monika (2015). Pre-harvest of rapeseed and mustard yield based on weather variables- An application of principal component analysis of weather variables. *Mausam*, **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. *Internat. J. Agric. & Statistical Sci.*, **10**(2): 381-385.

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

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

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.

Yadav, R.R. and Sisodia, B.V.S. (2015). Predictive models for Pigeon-pea yield using weather variables. *Internat. J. Agric. & Stat. Sci.*, **11**(2): 462-472.