MAUSAM, **67**, 3 (July 2016), 577-582

551.509.32 : 635.654

Pre-harvest forecast of pigeon-pea yield using discriminant function analysis of weather variables

R. R. YADAV, B. V. S. SISODIA and SUNIL KUMAR

Narendra Deva University of Agriculture and Technology, Narendra Nagar (Kumarganj), Faizabad – 224 229 (U. P.)

(*Received 10 March 2014, Accepted 18 November 2015*)

e mail : rajaramy22@gmail.com

सार – वर्तमान शोध पत्र में मौसम चर न्यूनतम और अधिकतम तापमान वर्षा, वर्षा के दिनों, 7 घंटे व 14 घंटों की सापेक्षिक आर्द्रता, तेज धूप के घंटे और पवन वेग की विभेदक फलन विश्लेषण का उपयोग पूर्वी उत्तर प्रदेश के फैजाबाद जिले में अरहर उपज के पूर्वानुमान के लिए उपयुक्त सांख्यिकीय मॉडल के विकास के लिए के बारे में बताया गया है। इस अध्ययन में कुल आठ मौसम चरों के 22 वर्षों के श्रृंखला वार साप्ताहिक आँकड़ों का उपयोग किया गया है। 22 वर्षों (1990-91 से 2011-12) के अरहर उपज के श्रृखलाबद्ध आँकड़ों को तीन वर्गों में अर्थात अनुकूल, सामान्य एवं प्रतिकूल में विभाजित किया गया। मौसम चर के साप्ताहिक आँकड़ों के उपयोग के लिए विभिन्न नीतियॉ प्रस्तावित कर विभेदक फलन विश्लेषण किया गया तथा उससे प्राप्त विभेदक स्कोरों को समाश्रयण के रूप में लेते हुए बहुसमाश्रयण पद्धति का उपयोग करके उपज के पूर्वानुमान मॉडलों का विकास किया गया है। कुल 9 मॉडल विकसित किए गए है। वर्ष 2009-10, 2010-11 और 2011-12 के लिए इन मॉड़लों से अरहर की पैदावार का पूर्वानुमान प्राप्त किया गया, जिन्हें इन मॉड़लों के विकास में शामिल नहीं किया गया था। R^2adj., प्रतिशत वर्ग त्रुटि (% आर एम एस ई) एवं प्रतिशत मानक त्रुटि (पी एस ई) के मानों के आधार पर मॉडल 4 व 9 को सबसे उपयुक्त पाया गया है। इन मॉडलों का उपयोग कर अरहर उपज का विश्वसनीय पूर्वानुमान फसल की कटाई के ढाई महीने पूर्व किया जा सकता है।

ABSTRACT. In the present paper, an application of discriminant function analysis of weather variables (minimum & maximum temperature, Rainfall, Rainy days, Relative humidity 7 hr & 14 hr, Sunshine hour and Wind velocity)for developing suitable statistical models to forecast pigeon-pea yield in Faizabad district of Eastern Uttar Pradesh has been demonstrated. Time series data on pigeon-pea yield for 22 years (1990-91 to 2011-12) 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 on eight weather variables in different forms has been carried out. The sets of discriminant scores obtained from such analysis have been used as regressor variables along with time trend variable and pigeon-pea yield as regressand in development of statistical models. In all nine models have been developed. The forecast yield of pigeon-pea have been obtained from these models for the year 2009-10, 2010-11 and 2011-12, which were not included in the development of the models. The model 4 and 9 have been found to be most appropriate on the basis of R^2adj, percent deviation of forecast, percent root mean square error (%RMSE) and percent standard error (PSE) for the reliable forecast of pigeon-pea yield about two and half months before the crop harvest.

Key words – Weather variables, Pigeon-pea yield, Discriminant function analysis, Forecast models.

1. Introduction

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 Chandrahas (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 applied the technique of discriminant function analysis to develop forecast models for wheat yield in Kanpur district (U. P.). Sisodia *et al*. (2014) have also made use of discriminant functions analysis of weakly data of weather variables for developing pre-harvest forecast models for wheat yield in Faizabad district (U. P). The pigeon-pea is most vulnerable crop to environment fluctuations. It is a long duration crop comprising about 44 weeks. It faces a lot of day to day weather variation during the entire period of crop-season. Therefore, an attempt has been made in the present paper to develop suitable statistical models for forecasting of pre-harvest pigeon-pea yield in Faizabad district of Uttar Pradesh (India) using discriminant functions analysis of weekly data of weather variables.

2. Materials and statistical methodology

2.1. *Area and crop covered*

The study has been 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 with an annual rainfall of 1002 mm, 90% of which is received during mid-June to mid-October. Pigeon-pea is the principal pulse crop of the kharif season in Faizabad district.

2.2. *Data*

Time series data of pigeon-pea yield of Faizabad district of Uttar Pradesh for 22 years (1990-91 to 2011-12) have been used for development of the models. These data have been collected from the Bulletins of Directorate of Agricultural Statistics and Crop Insurance, Govt. of Uttar Pradesh. Weekly weather data for the same period on eight weather variables, *viz*., Minimum Temperature, Maximum Temperature, Rainfall, Number of rainy days, Relative Humidity at 7 and 14 hrs, Sun-shine hours and Wind-Velocity have been used in the study. The weekly data on these weather variables have been obtained from the Department of Agro-meteorology, N. D. University of Agriculture & Technology Kumarganj, Faizabad, U. P., India.

2.3. *Crop season*

Preparation for sowing of pigeon-pea starts roughly from the first week of June in Faizabad districts and its harvesting starts from the first week of April of the next year. The entire crop season has been divided broadly into four phases. Phase I : pre-sowing, sowing, emergence and initial growth phase that includes the period from 28th may to about 22 July. Phase II : vegetative growth phase that includes the period from about 23rd July to 18th November. Phase III: flowering, reproductive and pod formation phase that includes the period from about 19th November to 25th February. Phase IV : ripening, maturity and harvesting period that start roughly from 26th February to 15th April. Therefore, the weekly data on weather variables have been collected for 46 weeks of the crop production which included 22th Standard Meteorological Week (SMW) that starts from 28th May to 52nd SMW of a year and 1st SMW to 15th SMW of the next year which ends by the second week of April.

2.4. *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 function observations into one of the previously defined groups. This technique is a multivariate technique discussed in many standard books, to mention a few, Anderson (1984); Johnson & Wichran (2001), etc. Therefore, theoretical developments of this technique need not to be presented here. However, few conceptual aspects of technique is given below.

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 classify the observations into different groups.

Agrawal *et al.* (2012) and Sisodia *et al*. (2014) have applied the technique of discriminant function analysis to develop pre-harvest forecast models for wheat yield in Kanpur and Faizabad district of U. P., respectively. Some of the models have provided reliable yield forecast about two months before harvest. This paper applies the same technique used by them along with a few modifications for the development of suitable models for pre-harvest forecast of pigeon-pea yield in Faizabad district of Uttar Pradesh.

In order to apply discriminant function analysis for modeling yield using weather variables, crop years under consideration 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 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 eight, 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 from normal group and adjusted crop yields above or equal to $\bar{y}$ + s would from 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 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).

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

$$Z_{ii',j} = \frac{\sum_{w=1}^{n} r_{ii'}^{j} X_{iw} X_{i'w}}{\sum_{w=1}^{n} r_{ii'w}^{j}}, \; j = 0,1 \text{ and } i = 1,2,...,p \qquad (1)$$

where, Z_{ij} is unweighted (for j = 0) and weighted (for j = 1) weather indices for i^{th} 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. 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 = 8 and n = 35, *i.e*., 35 weeks data from 22nd week to 52nd week of a year and 1st week to 4th 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 72 indices (36 weighted and 36 un-weighted) consisting of 8 weighted weather indices and 28 weighted interaction indices and 8 un-weighted weather variables and 28 un-weighted interaction indices have been constructed. Besides, some more suitable strategies have been suggested. In all, nine possible models are attempted. Models are developed using regression analysis. Only the first 19 years data from 1990-91 to 2008-09 have been utilized for modeling the yield and remaining three years yield data of 2009-10,2010-11 and 2011-12 have been used for validation of the models.

Model 1

In this model, using eight weighted weather indices of eight 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:

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + e \qquad (2)$$

where, y is untrended crop yield, $\beta_i^{'s}$ (i = 0, 1, 2, 3) are model parameters, 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 $N(0, \sigma^2)$.

Model 2

In this model, two discriminant functions and therefrom two sets of discriminant scores have been obtained using the first week data (22nd SMW) on eight weather variables. Again two sets of discriminant scores obtained from first week data and data on eight weather variables in the second week (23rd SMW) have been used as discriminating variables (in all 10 discriminating variables) and using these 10 discriminating variables the discriminant analysis has been done and there from two sets of discriminant scores have been obtained. This process was repeated up to the last week till the time of forecast (4th SMW or 35th week) and finally two sets of discriminant scores have been obtained. Based on these two sets of scores obtained at the 35th week, the forecasting model taking yield as the regressand and the discriminant scores and the trend variable (T) as the regressor variables has been fitted. The form of model is similar to the model given in equation (2).

Model 3

In this model, eight weighted and eight un-weighted weather indices of eight weather variables have been used as discriminating variables in the discriminant function analysis. Two sets of discriminant scores from two discriminant functions have been obtained. The forecasting model has been fitted taking the yield as the regressand and the two sets of scores and the trend variable (T) as the regessors. The form of model fitted is similar to given in equation (2).

Model 4

In this model, all 72 indices (weighted and un-weighted including interaction indices) have been used as discriminating variables in discriminant 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 similar to given in equation (2).

Model 5

In this model, discriminant function analysis has been carried out using the data on the first weather variable spread over 35 weeks (22nd SMW to 4th SMW of next year). Using two sets of discriminant scores obtained from two discriminant function of data on the first weather variable and 35 weeks data of the second variable, discriminant function analysis has been again performed and two sets of discriminant scores are obtained (here the discriminating variables will now become 37). Using these two sets of discriminant scores and 35 week data of third variable have been again used to carry out discriminant analysis and subsequently two sets of discriminant scores have been obtained. This process is continued up to eighth weather variable, 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 similar model as in equation (2).

Model 6

In this model, discriminant function analysis has been carried out using the un-weighted and weighted 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 un-weighted and weighted weather indices for the second weather variable, discriminant function analysis has been further carried out (here, the discriminating factors will be four). This process is continued up to eighth 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 and ds_2 and the time trend variable (T) as regressor variables, the similar model as in equation (2) has been fitted.

Model 7

This model is based on the method given by Rai and Chandrahas (2000). In fact, the crop season is divided into four phases where each phase consists of different number of weeks (see sub-section 2.3). However, we utilize the weekly data of weather variables of first three phases for the development of forecast model because it is intended to forecast the crop yield at the beginning of fourth phase. Weather indices have been constructed separately in the different phases of the crop growth by taking the simple average of the weekly weather variables. Then, at each phase discriminant function analysis has been carried out using these simple averages of weekly data of 8 weather variables as the discriminating variables, and two discriminant scores were obtained. In all 6 sets of discriminant scores have been obtained from 3 phases. Using these 6 sets of discriminant scores and trend variable (T) as regressor variables and yield as regressand, the following regression model has been fitted to develop the forecast model :

$$Y = \beta_0 + \sum_{l=1}^{2}\sum_{m=1}^{3}\beta_{lm} ds_{lm} + \beta_7 T + \varepsilon$$

where, β_0 = intercept of the model, β_{lm}'$s (l = 1, 2; m = 1, 2, 3)$ and β_7 are the regression coefficients, ds_{lm} is the l^{th} discriminant score obtained at m^{th} phase and T is the trend variable, and ε is error $\sim N\left(0, \sigma^2\right)$.

Model 8

The model 8 is newly proposed one but similar to model 7. The difference is only that two sets of discriminant scores have been obtained by carrying out discriminant function analysis using weighted weather indices of weekly data of weather variables computed by the formula given in equation (1) in each phase.

Model 9

The model 9 is newly proposed one but it is also similar to the model 7. Here two sets of discriminant scores have been obtained in each phase by carrying out discriminant function analysis using weighted and un-weighted weather indices of weekly data of weather variables computed by the formula given in equation (1).

Remarks - The models 1 and 2 were considered by Agrawal *et al*. (2012) and the model 3 to 6 were considered by Sisodia *et al*. (2014) for forecasting wheat yield in Kanpur and Faizabad district of U. P., respectively. The model 7 was proposed by Rai and Chandrahas (2002) for rice yield in Raipur district of Chhattisgarh.

2.5. *Comparison and validation of forecast models*

Different statistical measures used for the comparison and the validation of the models are described below:

(*i*) R_{adj}^2 : The nine models were compared on the basis of adjusted coefficient of determination (R_{adj}^2) which is as follows : $R_{adj}^2 = 1 - \dfrac{ss_{res}/(n-p)}{ss_t/(n-1)}$

TABLE 1

Pigeon-pea yield forecast models

Model	Forecast regression models
IV	$Y = 10.0 + 0.03ds_1 + 0.635^{**}ds_2 + 0.074^{*}T$
IX	$Y = 9.89 - 0.166ds_1 - 0.099ds_2 + 0.006ds_3 - 0.456^{*}ds_4 + 0.153ds_5 - 0.075ds_6 + 0.082T$

Significant at *P < 0.05, **P < 0.01

TABLE 2

Actual & Forecasts of Pigeon-pea Yield (Q/ha)

Model	Year	Actual yield (Q/ha)	Forecast yield (Q/ha)	Percent Deviation	RMSE	PSE (CV)	R^2	R^2adj
	2009-10	10.53	10.94	3.85		3.087		
IV	2010-11	16.26	15.31	5.85	0.623	2.690	87.3	84.8
	2011-12	11.53	11.84	2.69		3.501		
	2009-10	10.53	11.16	6.02		4.793		
IX	2010-11	16.26	15.12	7.01	0.756	5.779	87.2	82.0
	2011-12	11.53	11.42	0.92		4.843		

where, $ss_{res}/(n\text{-}p)$ is the residual mean square and $ss_t/(n\text{-}1)$ is the total mean square

(*ii*) *The percent deviation of forecast from actual yield* : It has have been computed as

$$\text{Percentage deviation} = \frac{\text{Actual yield - Forecast yield}}{\text{Actual yield}} \times 100$$

(*iii*) *Root Mean Square Error (RMSE)* : It is also a measure for 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 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 (independent variables) and $\hat{\sigma}^2$ is the estimated residual variance of the model. Therefore, the percent standard error (c.v.) of forecast is given by

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

3. Results and discussion

The forecast models have been developed under nine procedures described in sub-section 2.4. In Model IV only, time trend variable (T) has shown significant effect at $P \leq 0.05$. Based on the performance of these models as per criteria of statistical measures described in sub-section 2.5, it has been found that the model IV and IX are the best among all models. These two models are presented in the Table 1 Using the models IV and IX, the forecast yields for the years 2009-10, 2010-11 and 2011-12 have been computed and the results along with the actual yields and different statistical measures are presented in the Table 2. It is obvious form the results of the Table 2 that the forecast yields based on these models for the years under consideration are very close to the actual yields. The values of R^2_{adj} have been found to be considerably high, *i.e.*, 84.8 and 82.0 per cent for the

models IV and IX, respectively. The values of percent standard error (PSE) of forecast yield have been obtained to be below 5 percent in most the cases except one where it is 5.779 for the year 2010-11 in case of the model IX.

As per criteria of statistical measures, performance of other models have been found to be interior as compared to the models IV & IX. For instance, the values of R^2adj varied between 60.9 to 78.6 per cent for other models. The values of PSE of forecast yields were found to be above 5.00 per cent is most of cases for these models. The per cent deviation of forecast yields form actual yield was also obtained to be quite high ranging between 3 to 27.30 per cent for these models. Therefore, these models have not been presented.

On the basis of the overall results and discussion as above it can be concluded that the model IV and IX are the most suitable models among all the models to forecast pigeon-pea yield in Faizabad district of Eastern Uttar Pradesh. Hence, a reliable forecast of pigeon-pea yield about two and half months before the harvest can be obtained from the model IV and IX.

Acknowledgement

Authors are thankful to the referee for his valuable suggestions which has brought considerable improvement in the earlier manuscript of the paper.

References

Agrawal, R., Jain, R. C. and Jha, M. P., 1983, "Joint effects of weather variables on rice yields", *Mausam,* **32**, 2, 177-181.

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 multivariate statistical analysis", Wiley Eastern Private Limited, New Delhi.

Draper, N. R. and Smith, H., 1998, "Applied Regression Analysis", 3rd edition, John Wiley & Sons Inc.

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.

Sisodia, B. V. S., Yadav, R. R., Kumar, Sunil and Sharma, M. K., 2014, "Forecasting of Pre-harvest crop yield using discriminant function analysis of meteorological parameters", *Journal of Agrometeorology* **16**, 1, 121-125.