Plant Archives Vol. 18, Special Issue (ICAAS-2018), 2018 pp. 117-120 *ISSN 0972-5210*

WEATHER FORECAST MODELS OF POTATO YIELD USING PRINCIPAL COMPONENT ANALYSIS FOR BARABANKI DISTRICT OF EASTERN UTTAR PRADESH, INDIA

Snehdeep, B. V. S. Sisodia, V. N. Rai, Sunil Kumar and Ajit Singh

Department of Agricultural Statistics, Narendra Dev University of Agriculture and Technology, Kumarganj, Faizabad - 224 229 (U.P.), India.

E-mail: snehdeepsnehdeep@gmail.com, bvssisodia@gmail.com.

Abstract

The present investigation entitled "Forecast Models of Potato Yield Using Principal Componant Analysis for Barabanki District of Eastern Uttar Pradesh." Time series data on yield of potato and weekly data from 40th SMW of the previous year to 6th SMW of the following year on five weather variables *viz.,* Minimum temperature, Maximum temperature, Relative humidity 08.30hrs, Relative humidity 17.30 hrs and Wind-velocity covering the period from 1990-91 to 2011-12 have been utilized for development of pre-harvest forecast model. Statistical methodologies using multiple regression, principal component analysis for developing pre-harvest forecast model have been described. In aboth models (one based on regression and one from principal componant) have been developed. The Model-Ist is based on step wise regression, and IInd based on principal component analysis. Models have been developed on the basis of adjR^2, RMSE and %SE, the best model obtained by the application of step-wise regression analysis of weekly weather data are Model-IInd have further reduced the percentage standard error of the forecast yield to some extent. These models can be used to get the reliable forecast of potato yield two and half months before the harvest.

Key words : Pre-harvest forecast, statistical model, weather variables, principal componant.

Introduction

Potato (*Solanum tuberosum* L.) is the most important vegetable crop of the India and known as "The King of Vegetable". It is most important cash crop of Uttar Pradesh. Potato is nutritionally superior vegetable. Being a short duration crop, it produces more quantity of dry matter, edible energy and edible protein in lesser duration of time compared to cereals like rice and wheat. It is a native of tropical South America. India produced about 453.44 lakh tonnes of potato from 19.92 lakh hectares under the crop in the year 2012-13. The bulk of the produce come from state of Uttar Pradesh, West Bengal, Bihar and Punjab contributing 32, 26, 15 and 5% respectively in the year 2012-13. The area, production and potato yield at the national level increased during the period 1979-80 to 2010-11 by 172, 408 and 87%, respectively. The heat sensitive potato crop is mostly confined to Indo-Gangetic plains under irrigated conditions due to climate constraints. Small scattered area as rainfed crop are grown in hill during summers and in *kharif* season in plateau region, whereas winter season crop in the plateau region is irrigated. Usually the pre-harvest estimate of crop yield is obtained on the basis of visual observation which is not objective. There are two major objective approaches for forecasting crop yields one by using weather variables and the other by using weather variables and agriculture inputs jointly. These approaches can be used individually or in combination to give a composite model. Weather is one of the most important factors influencing crop growth. It may influence production directly through affecting the growth structural characteristics of crop such as plant population, numbers of tillers leaf area etc., and indirectly through its effect on incidence of pest and diseases. The effect of weather parameter at different stages of growth of crop may help in understanding their response in term of final yield and also provide a forecast of crop yields in advance before the harvest. The extent of weather influence on crop yields depends only on magnitude of weather parameters but also on their frequency distribution. Therefore, the knowledge of the frequency distribution of weather parameter is also essential while developing the pre-harvest model. Several studies have been carried out in past both in India and abroad on the crop weather relationship and forecasting crop yield, Fisher (1924) made first attempt to develop crop-weather relationship Hendrics and Scholl (1943) modified the Fisher's technique. Agarwal *et al.* (1980) further modified the technique of Hendrics and Scholl (1943) by developing forecast model using weather indices for rice crop in Raipur district & Chhatisgarh such technique of Agarwal *et al.* (1980) has been used by various author in the past for developing forecast yield of various crops in different region of the country. Notable among them are Sisodia *et al.* (2014), Azfar *et al.* (2014), Azfar *et al.* (2015), Yadav *et al.* (2016), Yadav *et al.* (2014), etc.

Materials and Methods

This chapter consists of the material used and the methodology employed for developing models to study the relationship between crop yield and weather variables, and to develop models for making pre-harvest forecast of yield. In order to facilitate systematic presentation, the chapter is divided into following sub-sections :

3.1 General information of the study area

3.2 Sources and description of data

3.3 Statistical methodology used for the development of models.

2.1 Description of the Study Area

Barabanki district is one of four districts of Faizabad division, lies at the very heart of Awadh region of Uttar Pradesh state of India and forms as it were a centre from which no less than seven other districts radiate. It is situated between 27°19' and 26°30' north latitude and 80°05' and 81°51' east longitude; it runs in a south-easterly direction, confined by the nearly parallel streams of the Ghaghara and Gomti. The extreme length of the district from east to west may be taken at 57 miles (92 km), and the extreme breadth at 58 miles (93 km); the total area is about 1,504 sq miles (3,900 km^2).

2.2 Yield data

Time series data on yield of potato for Barabanki district of Uttar Pradesh for 22 years (1990-91 to 2011-12) have been collected from the Bulletins of Directorate of Agricultural Statistics and Crop Insurance, Govt. of Uttar Pradesh.

2.3 Weather data

Weekly weather data for the same period on five weather variables *viz*., Minimum temperature, Maximum temperature, Relative humidity at 8.30 and 17.30 hrs and Wind-velocity have been used in the study. The weekly data on these weather variables have been obtained from the Department of meteorological centre Amausi Airport Lucknow, U.P., India.

2.4 Statistical Tools Used In the Analysis

Keeping in view the objectives set out for the study, following statistical tools and methods have been used. The data are analyzed by using software like SPSS and MS-EXCEL.

2.4.1. Development of the forecast model

This is based on the method given by Agrawal *et al.* (1986) for developing forecast using weather indices. In this procedure, the entire 19 weeks data from 40th 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; 5 un-weighted indices and 10 un-weighted interaction indices have been obtained. Considering these 30 indices and trend variable (T) as regressors and yield as dependent variable, forecast has been developed. The fitted formula is :

$$y = a_0 + \sum_{i=1}^{p}\sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i\neq i'-1}^{p}\sum_{j=0}^{I} a'_{iij} z_{IIJ} + {}_{c}T + \varepsilon$$

Where, $Z_{i\,j} = \sum_{w=1}^{n} r_{iw}^{j} X_{iw} \Big/ \sum_{w=1}^{n} r_{iw}^{j} \quad j = 0,1$

$$Z_{ii'j} = \sum_{w=n_1}^{n_2} r_{ii'w}^{j} X_{iw} X_{i'w} \Big/ \sum_{w=n_1}^{n_2} r_{ii'w}^{j}$$

y is the original crop yield, X_{iw} is the value of the ith weather variable in wth week, $r_{iw}/r_{ii'w}$ is correlation coefficient of yield adjusted for trend effect with ith weather variable/product of ith and i'th weather variable in wth week, n is the number of weeks considered in developing the weather indices, and p is number of weather variables used. a_0, a_{ij}, $a_{ii,j}$ and c are the parameters. ε is error term assumed to follow N (0, σ^2). The step-wise regression analysis was employed to develop the forecast.

2.4.2 Development of Statistical forecast models

To start with the crop years have been divided into three groups namely congenial, normal and adverse on the basis of crop yield. Here, only the first 19 years data from 1990-91 to 2008-09 have been utilized for model fitting and remaining three years were left for the validation of the model. The growth process of the crop has various phases and weeks within phases. At each week the weather variables corresponding to the three per defined groups have been used for the development of weather scores for each year through discriminant analysis technique. In the present study, the number of groups is three and number of weather variables is 5. Therefore, only two discriminant scores will be obtained. Discriminant analysis approach predicts the future observations qualitatively in different groups. For quantitative forecasting, regression models are fitted by taking the discriminant scores and the trend variable as the regressors and crop yield as the regressand. The models are developed using stepwise regression procedure. Models are developed by adopting five existing procedures and one proposed procedure.

2.4.5 Measures for validation and comparison of models

(i) Percentage Standard Error (%SE)

The % standard error of the (PSE) of the composite forecast of yield is computed as follows :

$$\text{PSE}=\frac{\sqrt{\text{V(ycf)}}}{ycf}\times 100$$

(ii) Percentage Deviation

This measures the deviation (in percentage) of forecast from the actual yield. The formula for calculating the percentage deviation of forecast is given below

$$\text{Percentage deviation} = \frac{\text{(Actual yield - Forecasted yield)}}{\text{(Actual yield)}} \times 100$$

(iii) Percentage Standard Error of the forecast

Let $\hat{y}_f$ be forecast value of crop yield and x_0 be the selected value of X at which the forecast has been done. The variance of $\hat{y}_f$ as given in Draper and Smith (1998) is given by

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

Where, X'X is the dispersion matrix of the sum of square and cross products of regressors and $\hat{\sigma}^2$ is the estimated residual variance of the fitted . The standard error of $\hat{y}_f$ is given by

$$\text{SE}(\hat{y}_f) = \sqrt{\text{V}(\hat{y}_f)}$$

and the % standard error (%SE) of $\hat{y}_f$ is given by

$$\%\text{SE} = \frac{\text{SE}(\hat{y}_f)}{\hat{y}_f} \times 100$$

(iv) Root Mean Square Error (RMSE)

It is also a measure for comparing two s. The formula of RMSE is given below

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

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.

Results and Discussion

This chapter deals with results, salient finding and discussion of the study undertaken. Pre-harvest forecast models as described in the preceding chapter have been developed. The results and findings and relevant discussion are presented as follows. Pre-harvest forecast models using principal component analysis of weekly data of weather variables.

Results for Barabanki district

Statistical models for pre harvest forecast of the potato yield in Barabanki district of Eastern Uttar Pradesh have been developed on the basis of weekly data on weather variables *viz.,* Minimum temperature, Maximum temperature, Relative humidity 08.30 hrs, Relative humidity 17.30 hrs and Wind-velocity using principle componant analysis. Following the two procedures described in Chapter-III, seven models have been developed. Sowing of potato starts generally from the first week of October in Barabanki district. Therefore, weekly data on the weather variables have been considered from pre-sowing the 40^{th} SMW of crop which fall during the first week of October. It has been proposed to make pre-harvest forecast of the potato yield at the stage of milking/dough, about two months before the harvest. Milking and dough stages generally start after about 130 days of sowing. Therefore, 6^{th} SMW of the next year (Feb. 5- Feb. 11) has been considered the week of pre-harvest forecast. Thus, in all 19 weeks data on the weather variables (40^{th} SMW of the previous year to 6^{th} SMW of the next year) have been utilized to develop the statistical models. In order to carry out discriminant analysis, the potato yields are adjusted for trend effect. The crop years have been divided into three group's namely congenial, normal and adverse groups. The actual potato yields, adjusted potato yield and the groups indicated

Table 1 : Comparison between actual and forecasted yield of different years of Barabanki district.

Model	Year	Actual yield	Forecast yield	Percent Deviation	RMSE	%SE	R^2	R^2adj.
I	2009-10	19.58	16.86	13.9	2.00	8.91	47.93	41.43
	2010-11	20.34	21.72	6.75		9.44		
	2011-12	26.07	24.44	6.26		10.65		
II	2009-10	19.58	14.69	24.95	2.87	10.91	84.92	81.9
	2010-11	20.34	19.46	4.32		12.64		
	2011-12	26.07	25.85	0.82		13.65		

Table 2 :

Model-I	$Y = 11.204 + 0.0792Z21 - 0.116T$	R^2=83.33	R^2adj=80.00
Model-II	$Y=4.981-0.0004P_1+0.0001P_2+0.004P_3-0.010P_4-0.003P_5-0.084T$	R^2=89.17	R^2adj=87.00

by 1, 2, 3 as adverse, normal and congenial, respectively, are given in the Table 1.

Comparison of the model

Based on these eight forecast models, the forecast yields for the 2009-10, 2010-11 and 2011-12 have been computed and result are presented in Table-3.2.2. The values of R^2adj, percent deviation of forecast from actual yield, RMSE and %SE (CV) have also been computed for each model and are also presented in the Table 1.

It is evident from the results of the table 1 that coefficient of determination (R^2) has been found to be 84.92% with RMSE 2.87, less percent S.E. and percent deviation for the Model-II followed by model-I have R^2 47.93 with minimum RMSE and lesser percent deviation with minimum percent S.E. followed. On the basis of the overall results of Table 2, it can be concluded that the Model-II is the most suitable model between both models to forecast potato yield in Barabanki district of Eastern Uttar Pradesh. Hence, a reliable forecast of potato yield about two and half months before the harvest can be obtained.

Summary and Conclusion

The forecast yields for the years 2009-10, 2010-11 and 2011-12 obtained from the aforesaid models including from the best composite forecast model, actual yield and various statistical measures for validation and comparison of the models are presented in the Tables 1 and 2. On the basis of the overall comparison of the results from these tables, it can be concluded that the Model-II for Barabanki based on application of Principal componant analysis of weekly weather data were found best among both models.

References

Fisher, R. A. (1924). *The influences of rainfall on the yield of wheat at Rothamsted*. Philosophical Transaction of Royal Society of London, Series B, Vol. 213,pp. 89-142.

Hendricks, W. A. and G. C. Scholl (1943). Technique in measuring joint relationship: The joint effects of temperature and precipitation on crop yield. *N. Carolina Agric. Exp. Stat. Tech. Bull.*, pp. 74.

Jain, R. C. (1998). Forcasting of Crop Yields using Second Order Markov Chains. *Jour. of the Ind. Sco. of Agril. Stats*., **LI (1)** : 61-72.

Mohd. Azfar, S. B. V. S. Sisodia, V. N. Rai and Monika Devi (2015). Pre harvest forecast models for rapeseed & mustard yield using principal component analysis of weather variables. *Mausam*, **66(4)** : 761-766.

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

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

Tripathi, Ramesh (2013). A Study on Statistical Models for Pre-harvest forecasting of potato yield based on weather variables. *M. Sc. (Agril.) thesis*, N. D. U. A. & T. Kumarganj, Faizabad, Uttar Pradesh.

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