Plant Archives Vol. 20, Special Issue (AIAAS-2020), 2020 pp. 523-525 ***ISSN 0972- 5210***

STATISTICAL MODELS FOR PRE-HARVEST FORECASTING OF POTATO YIELD BASED ON WEATHER VARIABLES

Ramesh Kumar Tripathi[1], Nitin Tanwar[2,*], VN Rai[1] and Lalit Kumar[2]

[1]Department of Agricultural Statistics, ANDUA&T, Kumarganj, Ayodhya (UP)

[2]Department of Statistics, Lady Shri Ram College for Women, University of Delhi, New Delhi, India

Abstract

An application of discriminant function analysis of time series weekly data on weather variables and weather indices together with time series data on potato yield has been demonstrated for development of pre harvest forecast model for potato yield. The Model-7 has been judged best one with considerably high value of R^2 (82.3%) and lower values of percent deviation, RMSE and PSE as compared to other models. The Model-4 and 3, however, also complete with Model-7.

***Keywords*:** Discriminant function analysis, forecast model, potato yield and weather variables.

Introduction

Potato (*Solanum tuberosum L.*) popularly known as "The king of vegetable", has emerged as fourth most important food crop in Uttar Pradesh after rice, wheat and maize. 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. Forecast of the crop production at suitable stages of crop period before the harvest are vital for rural economy. On the other hand, forecasts of crop yields are important for advance planning formulation and its implementation. These are useful to farmers to decide in advance their future prospects and course of action. Thus, reliable and timely pre-harvest forecasting of crop yield is very important. Statistical techniques have been used to develop various statistical models for pre-harvest forecast of crop yield using time series data on crop yield and weekly weather variable in the past, particularly for cereal crops (Agrawal *et al*., 1981, 1983, 1986, 2012; Jain *et al*., 1980; Sisodia *et al*., 2014). In the present paper, an attempt has been made to develop suitable statistical models for pre-harvest forecast of potato yield using time series data on potato yield and weather variables by applying a multivariate statistical tool, *i.e.* discriminant function analysis.

Materials and Methods

The study has been conducted for Faizabad district of Eastern Uttar Pradesh, India which is situated between 260 47' N latitude and 820 12' E longitudes and 113m above mean sea level. It has an annual rainfall of about 1002 mm. The favorable climate, soil and the availability of ample irrigation facilities make growing of potato a natural choice for the area.

Time series data on yield for potato crop of Faizabad district of Uttar Pradesh for 20 years (1991-2010) have been collected from Directorate of Agricultural Statistics and Crop Insurance, Government of Uttar Pradesh. Weekly weather data (1991-2010) on the weather variables of Faizabad district during the different growth phases of potato crop have been obtained from the Department of Agro meteorology, NDUAT, Kumarganj, Faizabad. The data on six weather variables viz. Minimum Temperature, Maximum Temperature, Minimum Relative Humidity, Maximum relative Humidity, Wind-velocity and Sun-shine hours have been used in the study.

Potato is generally sown from the third week of September in Faizabad District when average daily temperature falls around 250c. Growth of a potato plant occurs in several phases namely: sprout development, plant establishment, tuber initiation, and its bulking and maturation.

Statistical Methodology

The crop years have been divided into three groups namely congenial, normal and adverse on the basis of crop yield. Here only the first 17 years data from 1991 to 2007 have been utilized for development of model 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 pre-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 6. Therefore, only two sets of 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 various models are proposed as described below.

Model-1: This model is based on the method given by Agrawal *et al*. (1986) for developing forecast model using weather indices. The model fitted is

$$Y = A_0 + \sum_{i=0}^{p}\sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i \neq i'=1}^{p}\sum_{j=0}^{1} a_{ii'j} Z_{ii'j} + cT + e$$

*Corresponding author Email: nitintanwar5@gmail.com

Where Y is crop yield,

$$Z_{ij} = \sum_{w=n_1}^{n_2} r_{iw}^{j} X_{iw} \Big/ \sum_{w=n_1}^{n_2} r_{iw}^{j} \qquad j = 0,1$$

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

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, n1 & n2 are the initial and final weeks considered in developing the indices respectively and p is number of weather variables used. T is trend variable (T=1,2,3,……,n).e is error term assumed to follow independently normal distribution with mean 0 and variance2s. The step-Wise regression analysis was employed to develop the forecast model.

Model-2: Using weekly data on six weather variables for 12 weeks, weighted average (indices) of each weather variable has been computed following the procedure given in Model-1. The six indices have been used for carrying out discriminant function analysis and finally we get two sets of discriminant scores denoted as ds1 and ds2. Using ds1 and ds2 and time trend variable 'T' as a regressor and yield as regressand, the following model is fitted by ordinary least square technique.

$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + e$

where y is crop yield, $\beta_i^{'s}$ (i=0,1,2,3) are model parameters, and T is trend variable (T=1,2,3,……,n) and is error term assumed to follow independently normal distribution with mean 0 and variance .

Model-3: Discriminant function analysis has been carried out using 1^{st} week data of six weather variables and we get finally two sets of discriminant scores. Using these two sets of discriminant scores and the data on 2^{nd} week of weather variables have been used for discriminant function analysis. We again get two sets of discriminant scores and this process is continued for consecutive weeks of data of weather variables till 12^{th} week. Finally at the 12^{th} week we get two sets of discriminant scores and the model similar to the Model-2 is fitted.

Model-4: A total of 42 weather indices (weighted and unweighted averages and weighted and unweighted interactions) have been computed following the procedure given in Model-1. Using these 42 weather indices as the discriminating factors, discriminant function analysis has been carried out and two sets of discriminant scores were computed. Using the two sets of discriminant scores and time trend variable as regressor and yield as a regressand variable the model similar to the Model-2 has been fitted.

Model-5: Using 12 weather indices (six weighted and six unweighted averages from six weather variables) as the discriminating factors, discriminant function analysis has been carried out and finally we get two sets of discriminant scores. Using these two sets of discriminant scores and time trend variable as regressor and yield as a regressand variable the model similar to the Model-2 has been fitted.

The proposed two models (Model 6 and Model 7) are due to Sisodia *et al*. (2014) and are described below.

Model-6: Two weather indices (weighted and unweighted) from 1^{st} weather variable have been used as discriminating factors. Discriminant function analysis has been carried out with these two indices, and finally we get two sets of discriminant scores. Using these two sets of scores and two indices of the 2^{nd} weather variable, discriminant function analysis have been carried out and again we get two sets of discriminant scores. This process is carried out upto 6^{th} weather variable and finally we get two sets of discriminant scores ds1 and ds2. Using these two sets of discriminant scores and time trend variable as regressor and yield as a regressand variable the model similar to the Model-2 has been fitted.

Model-7: Using 12 weeks of data on 1^{st} weather variable have been used as discriminating factors, discriminant function analysis has been carried out, and we get two sets of discriminant scores. Using these two sets of discriminant scores and 12 weeks of data on 2^{nd} weather variable have been again used in discriminant function analysis and subsequently we get two sets of discriminant scores. This process is continued upto the 6^{th} weather variable and ultimately we get two sets of discriminant scores ds1 and ds2. Using these two sets of discriminant scores and time trend variable as regressor and yield as a regressand variable the model similar to the Model-2 has been fitted.

Majors for validation and comparison of Models

Different majors for validation and comparison of models have been used as described in the paper by Mohd Azfar *et al*. (2014) and Sisodia *et al*. (2014).

Results and Discussion

The models have been developed by utilizing 17 years data of potato yield (1990-91 to 2006-07) and remaining three years were left for the validation of the model. The fitted models are described below in Table 1.

The potato yields were forecasted for the years 2007-08, 2008-09 and 2009-10 using the forecast models described in Table 1. Various measures of validation and comparison of the models were also computed. The results are presented in the Table 2.

The validation of model based on coefficient of variation (R^2) alone may be misleading (Montgomery and Peck, 1982). Therefore, some other measures of validation such as percent deviation of forecast from actual yield, RMSE, percent standard error (PSE) of forecast together with R^2 have been used to judge the best model. Observing the results presented in the Table-2, we find that the model-7 has comparatively lower values of percent deviation, RMSE and PSE, and considerably high value of R^2 (82.3%) in comparison to that of the other models. It may be noted that R^2 was found to quite high (94.9%) for the Model-1 but values of other measures were not reasonably lower. Therefore, on the basis of these four measures of validation and comparison, the Model-7 is judged the best one followed by Model-4 and Model-3. Hence, preferably these Models-7, 4 and 3 can be used in practice for pre-harvest forecast of potato yield about one month before the harvest.

References

Agrawal R, Jain RC and Jha MP (1981). Effect of weather and Agriculture inputs on Rice yield and its forecast (Raipur District). Division of crop yield forecasting methodology, IASRI (ICAR) Library Avenue, New Delhi.

Agrawal R, Jain RC and Jha MP (1983). Joint effects of weather variables on rice yields. *Mausam*, **32(2):** 177-181.

Agrawal R, Jain RC and Jha MP (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.

Azfar M, Sisodia BVS, Rai VN and Devi M (2014). Pre-harvest forecast of Rapeseed and Mustard yield based on weather variable: An application of Discriminant function analysis. *International Journal of Agricultural and Statistical Sciences*, **10(2):** 497-502.

Jain RC, Agrawal R and Jha MP (1980). Effect of climatic variables on rice yield and its forecast. *Mausam*, **31(4):** 591-596.

Montgomery DC and Peck EA (1982). Introduction to linear regression analysis, (John Wiley & Sons, Canada).

Rai T and Chandrahas (2000). Use of discriminant function of weather parameters for developing forecast model of rice crop, (IASRI Publication).

Sisodia BVS, Yadav RR, Kumar Sunil and Sharma MK (2014). Forecasting of Pre-harvest Crop Yield Using Discriminant Function Analysis of Meteorological Parameters. *Journal of Agrometeorology*, **16(1):** 121-125.

Table 1: Forecast models along with value of R^2 (coefficient of determination)

Models		R^2
1	Y = 449.882 – 8.430 Z21 +0.674 Z41 - 0.145Q91 -0.195Q130	94.9
2	Y = 162.473 – 9.057 ds1 -10.861 ds2 -0.759 T	76.7
3	Y = 160.126 – 2.457 ds1 – 2.499 ds2 – 0.308 T	80.8
4	Y = 155.980 + 2.321 ds1 – 5.859 ds2 + 0.099T	76.9
5	Y = 150.889 – 4.526 ds1 – 9.771 ds2 + 0.637 T	73.6
6	Y = 151.844 – 11.200ds1 - 0.195ds2 + 0.489T	74.6
7	Y = 158.294 – 0.189 ds1 + 1.685 ds2 -0.138T	82.3

*NB:Z21 is unweighted average of maximum temperature, Z41 is weighted average of maximum relative humidity, Q91 is unweighted interaction between maximum temperature and wind-velocity and Q130 is weighted interaction between maximum relative humidity and sunshine hour.

Table 2: Forecast of potato yield along with various measures of validation and comparison

Model	Year	Actual potato yield (Q/ha)	Predicted potato yield (Q/ha)	Percent deviation	Percent Standard Error	RMSE	R^2 (%)
1	2007-08	131.51	160.23	21.84	10.35	43.12	94.9
	2008-09	119.74	194.43	62.37	7.94		
	2009-10	195.77	177.10	9.53	3.33		
2	2007-08	131.51	138.96	5.66	9.00	9.28	76.7
	2008-09	119.74	132.37	10.54	5.55		
	2009-10	195.77	162.31	17.08	5.00		
3	2007-08	131.51	156.39	18.92	4.35	6.78	80.8
	2008-09	119.74	114.20	4.62	7.15		
	2009-10	195.77	190.76	2.55	4.36		
4	2007-08	131.51	143.80	9.34	5.54	7.37	76.9
	2008-09	119.74	122.91	2.65	7.57		
	2009-10	195.77	197.27	0.76	5.08		
5	2007-08	131.51	139.61	6.16	5.89	12.61	73.6
	2008-09	119.74	136.55	14.03	6.33		
	2009-10	195.77	198.99	1.64	5.54		
6	2007-08	131.51	137.05	4.21	4.94	8.82	74.6
	2008-09	119.74	137.24	14.61	5.31		
	2009-10	195.77	188.81	3.55	4.80		
7	2007-08	131.51	144.54	9.90	4.06	5.57	82.3
	2008-09	119.74	116.73	7.52	2.51		
	2009-10	195.77	195.09	0.34	4.15		