Different approaches on pre harvest forecasting of wheat yield

Y. A. Garde[1]*B. S. Dhekale[2] and S. Singh[3]

[1]Department of Agricultural Statistics, College of Agriculture, N.A.U. Waghai (Dangs) - 394730 (Gujarat), INDIA
[2]Department of Agricultural Statistics, Faculty of Agriculture, BCKV, Mohanpur, Nadia - 741252 (West Bengal),INDIA
[3]Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi- 221005, (Uttar Pradesh), INDIA
*Corresponding author. E-mail: y.garde@yahoo.co.in

Abstract: Agriculture is backbone of Indian economy, contributing about 40 per cent towards the Gross National Product and provide livelihood to about 70 per cent of the population. According to the national income published in Economic survey 2014-15, by the CSO, the share of agriculture in total GDP is 18 percent in 2013-14. The Rabi crops data released by the Directorate of Economics and Statistics recently indicates that the total area coverage has declined; area under wheat has gone down by 2.9 per cent. Therefore needs to be do research to study weather situation and effect on crop production. Pre harvest forecasting is true essence, is a branch of anticipatory sciences used for identifying and foretelling alternative feasible future. Crop yield forecast provided useful information to farmers, marketers, government agencies and other agencies. In this paper Multiple Linear Regression (MLR) Technique and discriminant function analysis were derived for estimating wheat productivity for the district of Varanasi in eastern Uttar Pradesh. The value of Adj. R^2 varied from 0.63 to 0.94 in different models. It is observed that high value of Adj. R^2 in the Model-2 which indicated that it is appropriate forecast model than other models, also the value of RMSE varied from minimum 1.17 to maximum 2.47. The study revealed that MLR techniques with incorporating technical and statistical indicators (Model 2) was found to be better for forecasting of wheat crop yield on the basis of both Adjusted R^2 and RMSE values.

Key words: Discriminant function analysis, MLR techniques, Weather indices, Weather score, Wheat yield

INTRODUCTION

Wheat (*Triticum aestivum* L.) is an important cereal crop in India occupying second place, next to rice in production and which plays a critical role in food security. India raises almost exclusively winter wheat. The major wheat growing areas in India are located in the northern regions of the country. The state of Uttar Pradesh produces the most wheat in India, accounting for 31.19 percent of India's total wheat production. The Varanasi division comprises of four districts viz. Varanasi, Ghazipur, Chandauli and Jaunpur. Varanasi has about 69399 ha of the total area under wheat crop and has productivity is about 27.23 q/ha. (Dacnet 2015).

The government systems, agricultural businesses and farmers depending on agriculture for sustenance, may all be significantly responsive to fluctuations in weather, largely through the impacts on production and associated management intervention. The pre-harvest forecast of crop yield is likely to provide valuable information in regard to sale, storage, export, industries and government for advanced planning. The use of statistical models in forecasting food production and prices for agriculture and livestock sectors holds great significance. Similar work have been done by many

scientist viz. Jain *et al*. (1980) developed pre harvest model which was reliable to forecast rice yield only after about two months of sowing. Agrawal *et al*. (2001) developed forecasting model for wheat in Vindhyanchal Plateau zone of Madhya Pradesh. It was reported that reliable forecasting yield could be obtained when the crops were 12 weeks old i.e. about 2 month before harvest. Lee *et al*. (2011) developed wheat regression models that account for the effect of weather are developed to forecast wheat yield and quality. Rai and Chandrahas (2000) developed pre harvest forecast model for rice based on weather variables using techniques of discriminant function analysis in Raipur district. It was found that forecast of rice crop can be made about two months before harvest. Agrawal *et al*. (2012) used discriminant function analysis for developing wheat yield forecast model for Kanpur. The approach provided reliable yield forecast about two months before harvest.Skill in pre harvest forecasting offers considerable opportunities to crop managers through the potential to provide improvements in the overall farming system involved. The objective of the present study was to develop a simple approach for forecasting the wheat yield before harvesting using weather parameters.

MATERIALS AND METHODS

The study was carried at Varanasi district of eastern Uttar Pradesh. It is located in the middle Ganga valley of North India, in the eastern part of the state of Uttar Pradesh. The study was carried out in the year 2012-13. It utilized secondary yearly yield data of wheat (q/ha) for 27 years (1982-83 to 2008-09) which were collected from different issues of "Uttar Pradesh Ke Krishi Aankre" published by Directorate of Agricultural Statistics and Crop Insurance, Krishi Bhavan, Lucknow (U.P.), and weekly weather data for same period of the years were collected from All India Coordinated Research Project on Dry Land Agriculture, Department of Agronomy, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi. Seven weather parameters were included in the study namely Maximum Temperature (X_1), Minimum Temperature (X_2), Total Rainfall (X_3), Pan Evaporation (X_4), Relative Humidity at 7.00 hrs (X_5), Relative Humidity at 14.00 hrs (X_6), and Sunshine hrs (X_7). However, weekly weather data related to Rabi crop season starting from a fortnight before sowing up to last of reproductive stage were utilized for the development of statistical models. Therefore, the weather data for wheat crop, from October 15 (42nd standard meteorological week, SMW) to February 18 (7th standard meteorological week, SMW) in each year were utilized. In the study 42nd SMW correspond to 1st week, 43rd SMW correspond to 2nd week to 52nd SMW correspond to 11th week and 7th SMW correspond to 18th week. The wheat yield distribution over the year is shown in Fig.1

Statistical techniques: The Multiple linear regression techniques (MLR) was used for developing pre-harvest forecast models using predictors as appropriate unweighted and weighted weather indices. This methodology further improves by incorporating statistical & technical indicators as regressors along with weather indices/score. Another approach used for obtaining pre-harvest forecast models was discriminant function analysis. The detailed methodology is explained below:

Pre harvest forecast models based on weather indices: In this method, weekly data on weather variables of 18 weeks have been utilized for constructing
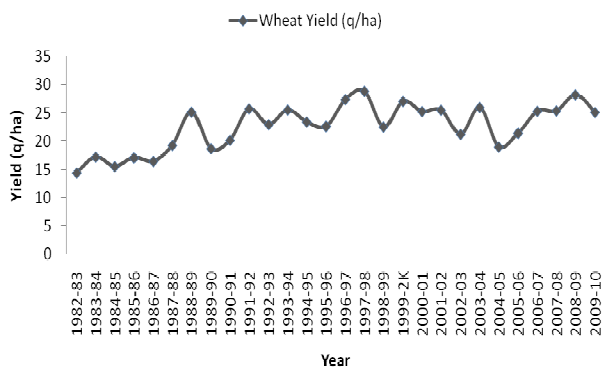

Fig. 1. *Wheat yield (q/ha) during 1982-83 to 2009-10 for Varanasi.*

weather indices (weighted & un-weighted along with their interactions). The weighted indices are weighted accumulations of the weather variables over weeks, where weight being correlation coefficient between de-trended yearly crop yields and weather parameters with respective weeks (Agrawal *et al.* 2001, 2007). The forms of indices are given as below:

$$Z_{i,j} = \sum_{w=1}^{m} r_{iw}^{j} X_{iw} \quad \text{and} \quad Z_{ii',j} = \sum_{w=1}^{m} r_{ii'w}^{j} X_{i'w}$$

Where,

$j = 0, 1$ (where, '0' represents un-weighted indices and '1' represents weighted indices)

m = Week up to forecast ($m=18^{th}$)

w = week number (1, 2, ..., m)

r_{iw} = Correlation coefficient between adjusted crop yield and i^{th} weather variable in wth week

$r_{ii'w}$ = Correlation coefficient between adjusted crop yield and the product of i and i'^{th} weather variable in wth week

X_{iw} and $X_{i'w}$ are the i and i'^{th} weather variable in wth week respectively

The pre-harvest forecast models were obtained by applying the MLR Techniques by taking predictors as appropriate un-weighted and weighted weather indices. Stepwise regression analysis was used for selecting significant variables (Draper and Smith 1981; Gomez and Gomez 1984). The regression model was as follows:

Model-1

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

Where,

and are the weather indices

i, i' = 1, 2, ...p

p = Number of weather variables under study

Y = District total crop yield (q/ha)

T = Year number

A_0 is the intercept

$a_{ij}, a_{ii'j}, c$ are the regression coefficient

e is error term normally distributed with mean zero and constant variance

Pre harvest forecast models based on weather indices with incorporating technical and statistical indicators: The Model-1 was further made to order by incorporating technical and statistical indicators along with weather indices as predictors. Technical and statistical indicators provide hidden pattern of the time series data (Refens *et al.* 1994). The fitted model is:

Model-2

$$Y = A_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{i,j} Z_{i,j} + \sum_{i \neq i'=1}^{p} \sum_{j=0}^{1} a_{i,i',j} Z_{i,i',j} + \alpha_1 MA + \alpha_2 EMA + \alpha_3 OSC + \alpha_4 ROC + \alpha_5 RAN$$

$$+ \alpha_6 VAR + \alpha_7 M3 + \alpha_8 M4 + \alpha_9 CV + \alpha_{10} MD + \alpha_{11} \beta1 + \alpha_{12} \beta2 + cT + e$$

Where, A_0 is the intercept, $\alpha_1, \alpha_2, ..., \alpha_{12}$ are the regression coefficients, *MA, EMA, OSC,* and *ROC* are technical indicators and *RAN, VAR, M3, M4, CV, MD, β1* and *β2* are statistical indicators which were produced through time series yield data (Garde *et al.* 2012). Where, i) MA= Four Year's Moving Averages, ii)

EMA= Exponential Moving Average, iii) OSC= Oscillator, iv) ROC= Rate of Change, v) RAN= Range, vi) VAR= Second Moment about Mean, μ_2, vii) M3= Third Moment about Mean, μ_3, viii) M4= Fourth Moment about Mean, μ_4, ix) CV= Coefficient of Variance, x) MD= Mean Deviation, xi) $\beta 1$= Coefficient of skewness, xii) $\beta 2$= Coefficient of kurtosis.

$z_{i,j}$ and $z_{i,i',j}$ are the weather indices

$i,i' = 1, 2, …$p

p = Number of weather variables under study

Y = District total crop yield (q/ha)

T = Year number

A_0 is the intercept

a_{ij}, $a_{ii'j}$, c are the regression coefficient e is error term normally distributed with mean zero and constant variance

Pre harvest forecast models using Discriminant function analysis: The statistical technique of discriminant function is a technique which discriminates between various groups of objects on the basis of characters which are considered to be relevant. The procedure start with crop years have been divided in to three groups namely congenial, normal and adverse on the basis of crop yield adjusted for trend effect.

Method-1

In this method, weekly data on weather variables of 18 weeks have been utilized for constructing weighted indices along with their interactions as $z_{i,i',j} = \sum_{w=1}^{m} r_{ii'w}^{j} x_{i'w}$. Total twenty eight weighted indices were considered and next two discriminant score have been obtained from these indices. For quantitative forecast, regression models were fitted by taking the discriminant scores and trend variable as the regressors and crop yield as the regress and (Agrawal *et al.* 2012). The form of the developed model is as follows:

Model-3

$$Y = A_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + e$$

Where,

A_0= intercept of model

β_i 's (i=1,2,3) = the regression coefficients

ds_1 and ds_2 are the two discriminant scores

T = trend variable

e is error term normally distributed with mean zero and constant variance

Method-2

In this method, discriminant scores have been computed using seven weather parameters for the first week (42nd SMW). At second week (43rd SMW), seven weather parameters along with the scores computed at the first week (42nd SMW) have been used as discriminator variables. Based on these variables the discriminant analysis has been done and two new discriminant score have been obtained. The same calculations have been repeated at third week (44th SMW). This procedure continued up to 18th week (7th SMW) and final two discriminant score have been obtained. The regression model has been fitted by these two scores obtained at the 18th week, along with trend variable as the predictors and crop yield as the dependent variable. The form of the fitted model is as follows:

Model-4

$$Y = A_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + e$$

Where,

ds_1 and ds_2 are the two discriminant scores on 18th week

A_0, β_i 's (i=1,2,3), and T are as defined in model 3

Method-3

In this method, average of total of weighted indices from first week to 18th weeks was done. This average computed indices utilized for calculating discriminant scores using discriminant function analysis and two discriminant score have been obtained. This model considered complete weather indices of relative different weeks. The data on these two discriminant scores along with trend variable were used to developed the model through stepwise regression technique. The fitted model is given below:

Model-5

$$Y = A_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + e$$

Where,

ds_1 and ds_2 are are the two discriminant scores on average weighted indices.

A_0, β_i 's (i=1,2,3), and T are as defined in model 3

Comparison and validation of forecast models: The coefficient of determination (Adjusted R^2) is general used to for checking adequacy of the model. It is preferable to used, when models to be compared are based on different number of independent variables. On adding a independent variables in the model, Adjusted R^2 increases only if the addition of regressor reduces residual mean square. Therefore it is important in evaluating and comparing the developed models.

Another method was taken under consideration for the comparison of developed pre harvest forecasting models is the Forecast Error percentage and it is calculated using following formula:

$$Forecast\ Error\ (\%) = \frac{Observe\ Yield - Forecast\ Yield}{Observe\ Yield} \times 100$$

In addition to Forecast errors (%), Root Mean Square Error (RMSE) was calculated as a measure of comparing two models. The formula of RMSE is given below:

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

Where, n is the number of years for which forecasting has been done.

RESULTS AND DISCUSSION

Wheat yield forecast models were developed using stepwise MLR techniques, therefore all included vari-

ables in the model equations are significant. The details of yield forecast models discussed hereunder-
The data obtained for analysis was done in Microsoft Excel -07 and SPSS statistical packages. The results obtained are tabulated here. The forecast models equation obtained under five methods discussed above are given in Table 1 along with Adjusted (Adj.) R^2. It is found that except Model-5 in all models, trend variable T was significant. In model-1, weather indices $Z_{2,7,1}$ (Min. Temp & Sunshine hrs), $Z_{4,6,1}$ (Pan Evaporation & Relative Humidity at 14 hrs), $Z_{4,6,0}$ (Pan Evaporation & Relative Humidity at 14 hrs and $Z_{2,0}$ (Min. Temp.) were found significant. In model-2, weather indices $Z_{2,7,1}$ (Min. Temp & Sunshine hrs), $Z_{4,6,1}$ (Pan Evaporation & Relative Humidity at 14 hrs), $Z_{4,6,0}$ (Pan Evaporation & Relative Humidity at 14 hrs), $Z_{5,7,0}$ (Relative Humidity at 14 hrs & Sunshine hrs), $Z_{5,7,1}$ (Relative Humidity at 14 hrs & Sunshine hrs), and EMA were found significant. Apart from these, other significant variables were found as discriminant score ds_1 in Model-3, ds_2 in Model-4 and ds_1 & ds_2 in Model-5. The Adj. R^2 varied from 0.63 to 0.94 in different models. It is observed that high value of Adj. R^2 in the Model-2 which indicated that it is appropriate than others.
The percent deviation and the RMSE were computed on the basis of observed and forecast yield (q/ha) for the years 2006-07 to 2009-10. The results (Table 2) revealed that the percent deviation of forecast varied from 0.39 to 10.95 in the Model 1, 0.44 to 8.86 in the Model 2, 1.52 to 5.63 in the Model 3, 5.44 to 11.9 in

the Model 4 and 1.02 to 14.26 in the Model 5 over the four years. The RMSE (Table 2) varied from minimum 1.17 to maximum 2.47 in Model 3 and Model 4 respectively.
The result based on Table1 and Table 2, revealed that Model 2 showed high value of R^2 as compare to others and second minimum value of RMSE. So it is concluded that Model 2 is the most suitable model among the models considered for wheat yield forecasting in Varanasi district of Uttar Pradesh. The model provides reliable forecast around two months before harvest. Similar model has been found more reliable for forecast of rice yield using weather indices in Pantnagar given by Garde *et al.* (2012).

Conclusion

This study concluded that stepwise techniques MLR can be successfully used for pre-harvest wheat crop yield forecasting. This model was most consistent and can be applied on zone or state level. The study also revealed that incorporating of technical and statistical indicators can increase the efficiency of the model. The technique of Discriminant function was found useful in classifying the crop year in to congenial, normal and adverse year with respect to crop yield further weather score developed through Discriminant function analysis method. The proposed model by incorporating technical and statistical indicators along with weather indices was found to be best as compared to

Table 1. Wheat yield forecast models.

Model No.	Forecast model equation	Adjusted R^2
1	$Y = 31.512 + 0.047Z_{2,7,1} + 0.293T + 0.047Z_{4,6,1} - 0.010Z_{4,6,0} - 0.052Z_{2,0}$	0.90
2	$Y = 14.413 + 0.065Z_{2,7,1} + 0.183T + 0.051Z_{4,6,1} - 0.012Z_{4,6,0} + 0.166\,EMA$ $+ 0.001Z_{5,7,0} - 0.007Z_{5,7,1}$	0.94
3	$Y = 19.791 + 1.146\,ds_1 + 0.193\,T$	0.81
4	$Y = 20.150 - 1.221\,ds_2 + 0.182\,T$	0.63
5	$Y = 22.041 + 1.039\,ds_1 + 0.510\,ds_2$	0.89

Table 2. Observed and forecasts of wheat yield.

Forecast Year	Observed yield (q/ha)	Forecast ield (q/ha)				
		Model I	Model II	Model III	Model IV	Model V
2006-07	25.33	25.43 (0.39)	25.33 (0.44)	25.72 (1.52)	28.18 (11.24)	25.07 (1.02)
2007-08	25.40	28.18 (10.95)	25.40 (8.27)	26.81 (5.57)	26.78 (5.44)	27.22 (7.17)
2008-09	28.23	25.93 (8.15)	28.23 (8.82)	26.64 (5.63)	25.88 (8.32)	24.20 (14.26)
2009-10	25.13	27.34 (8.81)	25.13 (8.86)	26.03 (3.59)	28.12 (11.90)	25.49 (1.45)
RMSE		2.12	1.97	1.17	2.47	2.22

figures in parenthesis denotes forecast error (%)

model based on only weather indices for the crop.

ACKNOWLEDGEMENT

REFERENCES

Agrawal R., Chandrahas and Kaustav A. (2012). Use of discriminant function analysis for forecasting crop yield, *Mausam,* 63, 3 (July 2012), 455-458

Agrawal, R. and Mehta, S.C. (2007). Weather based forecasting of crop yields, pest and diseases – IASRI Models. *J. Ind. Soc. of Agril. Stat.* 61 (2):255-263.

Agrawal, R. Jain, R.C. and Mehta, S.C. (2001). Yield forecast based on weather variables and agricultural inputs on agro-climatic zone basis. *Ind. J. Agri. Sci.,* 71 (7), 487-490.

Dacnet (2015). District wise crop production statistics retrieved July, 2015 from http://apy.dacnet.nic.in/cps.aspx

Draper, N.R. and Smith, H. (1981). Applied regression analysis, second edition, John Wiley and sons, New York.

Garde, Y.A.; Shukla, A.K. and Singh, S. (2012). Pre-Harvest Forecasting of Rice Yield Using Weather Indices in Pantnagar (Uttarakhand), *Int. J. Agricult. Stat. Sci.,* Vol. 8, No. 1, pp. 233-241.

Gomez, K.A. and Gomez, A.A. (1984). Statistical Procedures in Agricultural Research. 2nd edition, New York, Chichester, etc.: Wiley paperback, pp. 680

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

Lee, B.H., Kenkel, P. and Brorsen, B.W. (2011). Pre-harvest forecasting of county wheat yield and wheat quality conditional on weather information. Paper prepared for presentation at *the Southern Agricultural Economics Association Annual Meeting, Corpus Christi Texas,* February 5-8, 2011.

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

Refens, A.P.M.; Zapranis, A. and Francis, G. (1994). Stock performance modeling using neural networks: A comparative study with regression models, *Neural Networks*, 7 (2): 375-388