MAUSAM

Different statistical models based on weather parameters in Navsari district of Gujarat

K. B. BANAKARA, Y. A. GARDE* and H. R. PANDYA*

Mahalanobis National Crop Forecast Centre (MNCFC), Ministry of Agriculture and Farmers Welfare,

GoI, Pusa Campus, New Delhi – 110 012

*Department of Agricultural Statistics, Navsari Agricultural University, Navsari, Gujarat – 396 450

(*Received 26 October 2021, Accepted 12 December 2022*)

e mail : y.garde@yahoo.co.in

सार — कृषि देश के विकास में बहुत महत्वपूर्ण भूमिका निभाती है। चावल दुनिया की आधी से ज्यादा आबादी का मुख्य भोजन है। समयोचित और विश्वसनीय पूर्वानुमान महत्वपूर्ण व उचित इनपुट, दूरदर्शिता और भविष्य की सुनिश्चित योजना प्रदान करता है। वर्तमान जांच खरीफ चावल की उपज का पूर्वानुमान लगाने के लिए दो अलग-अलग सांख्यिकीय तकनीकों नामतः विभेदक कार्य विश्लेषण और सुप्रचालन समाश्रयण विश्लेषण का उपयोग करके की गई। 1990 से 2012 तक के डेटा का उपयोग करके सांख्यिकीय मॉडल विकसित किए गए और विकसित मॉडल का सत्यापन शेष डेटा यानी 2013 से 2016 का उपयोग करके किया गया। यह देखा गया कि विभिन्न मॉडलों में समायोजित R^2 का मान 73.00 प्रतिशत से 93.30 प्रतिशत तक भिन्न था। समायोजित R^2, पूर्वानुमान त्रुटि और RMSE के उच्च मान के आधार परसर्वोत्तम पूर्वानुमान मॉडल का चयन किया गया। नवसारी जिले से प्राप्त परिणामों के आधार पर, चावल की फसल की उपज कीकटाई से पहले के पूर्वानुमान के लिए सुप्रचालन समाश्रयण विश्लेषण (मॉडल-12) की तुलना में विभेदक कार्य विश्लेषण तकनीक (मॉडल-5) को बेहतर पाया गया। परिणामों से पता चलता है कि मॉडल-5 ने Adj. R^2 के उच्चतम मान के साथ तुलनात्मक रूप से कम पूर्वानुमान त्रुटि (%) दिखाई। R^2 (93.30) और RMSE का न्मनतम मान (120.07)। इसके अलावा मॉडल-12 (40$^{वां}$ एसएमडब्ल्यू) मॉडल-5 की तुलना में एक सप्ताह पहले (39$^{वां}$ एसएमडब्ल्यू) उपज को पूर्वानुमान प्रदान करने में सक्षम है।

ABSTRACT. Agriculture plays very important role in development of country. Rice is a staple food for more than half of world's population. Timely and reliable forecasting provides vital and appropriate input, foresight and informed planning. The present investigation was carried out to forecast *Kharif* rice yield using two different statistical techniques, *viz*., discriminant function analysis and logistic regression analysis. The statistical models were developed using data from 1990 to 2012 and validation of developed models was done by using remaining data, *i.e*., 2013 to 2016. It was observed that value of adjusted R^2 varied from 73.00 per cent to 93.30 per cent in different models. The best forecast model was selected based on high value of adjusted R^2, Forecast error and RMSE. Based on obtained results in Navsari district, the discriminant function analysis technique (Model-5) was found better than logistic regression analysis (Model-12) for pre-harvest forecasting of rice crop yield. The results revealed that Model-5 showed comparatively low forecast error (%) along with highest value of Adj. R^2 (93.30) and lowest value of RMSE (120.07). Also Model-5 is able to generate yield forecast a week earlier (39thSMW) than Model-12 (40thSMW).

Key words – Discriminant function analysis, Logistic regression, Forecast, Weather indices.

1. Introduction

The different statistical techniques have been used for development of pre-harvest models which foretell yield before harvest of the crop yield. The dependable and timely forecasts provide vital and applicable suitable input, foresighted and informed planning. Rice, (*Oryza sativa* L.) is the staple food for more than half of the world's total population and hence a key pillar for food security. More than 90 per cent of the world's rice is grown and consumed in Asia where 60 per cent of world's population lives. In India, major share of rice is cultivated during *Kharif* season. A small share of rice is grown in *rabi* /summer season with assured irrigation.

The crop weather relationship has been studied by Fisher (1924) and Hendricks and Scholl (1943) and developed models which required small number of
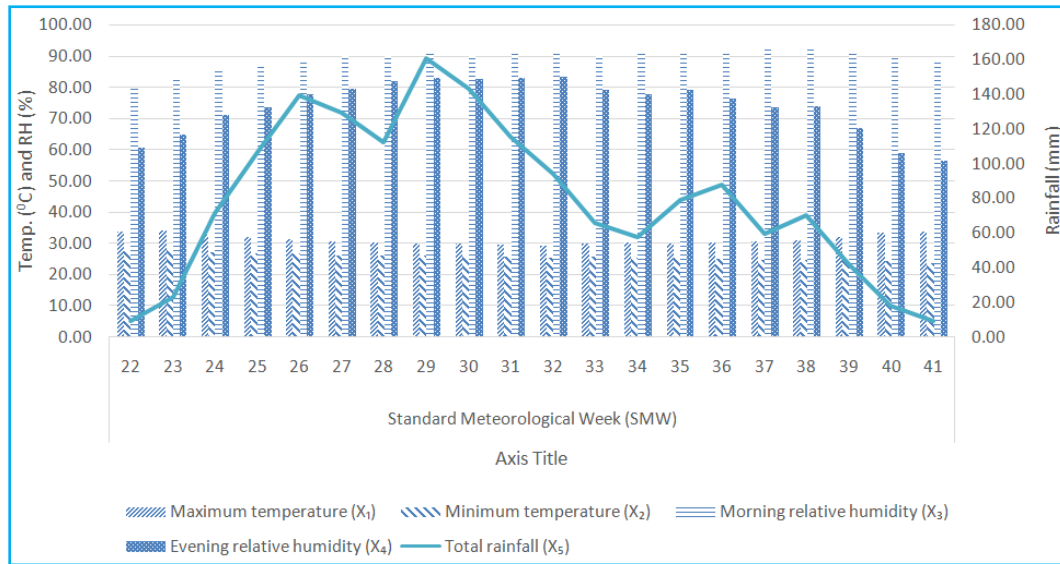
Fig. 1. Average weekly weather variables over the study period 1990 to 2016

parameters to be estimated while taking care of distribution pattern of weather over the crop season. Agrawal *et al.* (1986) and Jain *et al.* (1980) modified this model by expressing effects of changes in weather parameters on yield in the particular week as second degree polynomial in respective correlation coefficients between yield and weather parameters. This model was further modified (Agrawal *et al.* 1986) by explaining the effects of changes in weather parameters on yield in particular week using correlation as weight using linear function. Patel *et al.* (2007); Chauhan *et al.* (2009); Garde *et al.* (2012); Mahdi *et al.* (2013); Ghosh *et al.* (2014); Singh *et al.* (2014) and Garde *et al.* (2020) studied the relationship of weather parameters and rice crop yield in different regions of the world. Varmola *et al.* (2004); Agrawal *et al.* (2012); Sisodia *et al.* (2014) and Garde *et al.* (2015) developed forecast models for Wheat crop in different regions of India. Similarly, for pigeon pea, Kumar *et al.* (1999) and Sarika *et al.* (2011), for Sugarcane Priya and Suresh (2009) and for Groundnut Dhekale *et al.* (2014) developed models.

Rice is confined to south and middle Gujarat which occupies about 8.41 lakh hectares of cropped area of the Gujarat state, accounts for around 1883.61 thousand tonnes of food grain production with productivity of about 2238 kg per hectare (Anonymous, 2021). South Gujarat consists of seven districts, *viz.*, Navsari, Valsad, Surat, Bharuch, Dang, Tapi and Narmada.

The development of pre-harvest statistical models can play important role in policy decision regarding export and import, food procurement and distribution, price policies and exercising several administrative measures for storage and marketing of agricultural commodities. Therefore, forecasting food production and prices for agriculture hold great significance. Although no statistical model can help in forecasting the values exactly but, knowing even approximate values can help in formulating future plans. The present study was undertaken with the main objective of comparing different forecasting techniques of *Kharif* rice yield utilizing weather parameters in Navsari district of south Gujarat, as it is one of the most important districts in Gujarat.

2. Data and methodology

Considering the specific objective of the study, *Kharif* rice yield data were collected from the Directorate of Economics and Statistics, Government of Gujarat, Gandhinagar, Gujarat from the year 1990 to 2016.

Advancement of monsoon normally takes place in Navsari district during 1st week of June. Rainfall is maximum in the month of July followed by August, September and June. In Navsari, sowing of *Kharif* rice is carried out immediately after onset of monsoon. Duration of *Kharif* rice is 100-120 days depending on varieties.

The present study utilized weekly weather data which were collected from the Dept. of Agril. Engineering, N. M. College of Agriculture, Navsari Agricultural University, Navsari. Different weather parameters were taken under consideration, *viz.*, Maximum temperature (X_1), Minimum temperature (X_2), Morning relative humidity (X_3), Evening relative humidity (X_4) and Total rainfall (X_5) to study the effect on *Kharif* rice yield. The weekly weather data related to *Kharif* crop

season starting from a first fortnight before sowing to last of reproductive stage were utilized for the development of statistical models (Agarwal *et al.*, 1986). Accordingly, for each year, weather data from the month May-June [23rd Standard Meteorological Week (SMW)] to the month of October (41st SMW) were utilized for development of statistical models in *Kharif* rice. The graphical representation of average weekly weather variables over the year is presented in Fig. 1.

2.1. *Development of weather indices using correlation coefficient as weight*

As yield of crops is highly influenced by weather, the models were developed for studying the effects of weather on production of *Kharif* rice. Two weather indices ($Z_{i,j}$) were developed for each weather variable, one as simple accumulation of weather variable and other one as weighted accumulation of weather variable, where weights being correlation coefficient between weather variable in respective weeks and de-trended yield (Agarwal *et al.*, 1980; Garde *et al.*, 2012). Similarly, for interaction effect of weather variables, the indices ($Z_{i,i'j}$) were generated by taking the products of weekly weather variables, two at a time. The forms of indices are given below:

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

where,

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

m = Week of forecast

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

r_{iw} = Correlation coefficient between de-trended crop yield and i[th] weather variable in wth week

$r_{ii'w}$ = Correlation coefficient between de-trended 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

2.2. *Statistical approaches*

Different statistical approaches were adopted for development of forecast models using weather indices. In present investigation data analysis was carried out by using following statistical tools.

2.2.1. *Discriminant Function Analysis*

Discriminant function analysis is an appropriate statistical technique when the dependent variable is categorical and the independent variables are metric. The aim of discriminant analysis is to develop discriminant functions, *i.e.*, the linear combination of independent variables that will discriminate between the categories of the dependent variable. It is also an appropriate statistical technique for testing the hypothesis that the group means of a set of independent variables for two or more groups are equal. In the present study, the year have been divided into two and three groups on the basis of de-trended rice crop yield. The grouping was done by arranging de-trended yield in ascending order and were divided into two equal groups [adverse (0) and normal (1)]. Similarly, crop years were grouped into three equal groups namely adverse (0), normal (1) and congenial (2). The discriminant scores were obtained by discriminant function analysis for development of crop yield forecast model by using weather indices (Garde *et al.*, 2020).

2.2.1.1. *Development of Models Based on Two groups*

Method-1

Five un-weighted weather indices were utilized to extract discriminant scores by using discriminant function analysis. One discriminant score was obtained for each year. The forecasting model was fitted taking the *Kharif* rice yield as the regress and discriminant score (ds_1) along with time trend T as the regressors. The form of the model is given below;

Model-1

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 T + \varepsilon$$

where,

Y is crop yield, 'βi's (i = 0,1,2) are regression coefficients, ds_1 is the discriminant score based on un-weighted weather indices, T is the time trend variable and ε is error term assumed to follow NID ~ (0, σ^2).

Method-2

Similar to method-1, five weighted weather indices were used instead of un-weighted weather indices to extract discriminant scores using discriminant function analysis. The form of the model given below;

Model-2

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 T + \varepsilon$$

where,

ds_1 is the discriminant score based on weighted weather indices and Y, 'βi's and T areas mentioned Model-1.

Method-3

The model was developed by dividing cropping period, starting from four weeks before transplanting up to the time of forecast (*i.e.*, 17 weeks starting from 23rd SMW) into three phases (Rai and Chandrahas, 2000). For each phase simple average of the individual weather variables was obtained. The phase-wise discriminant scores were obtained through discriminant function analysis. Thus, in all, three scores were obtained for each year and model was fitted using these three discriminant scores along with time trend (T) as independent variables and *Kharif* rice yield as dependent variable. The form of regression equation as mentioned below;

Model-3

$$Y = \beta_0 + \sum_{l=1}^{1} \sum_{m=1}^{3} \beta_{lm} \, ds_{lm} + \beta_{11}T + \varepsilon$$

where,

Y is the rice crop yield, β_0 is intercept of the model, 'β_{lm}'s ($l = 1$, $m = 1$, 2, 3) and β_{11} are the regression coefficients, ds_{lm} is the lth discriminant score in m^{th} phase, T is the time trend variable (year) and ε is error NID ~ (0, σ^2).

2.2.1.2. *Development of Models Based on Three groups*

The model development was carried out with three groups namely congenial, normal and adverse on the basis of crop yield adjusted for trend effect. The details of developed methods are discussed here under:

Method-4

The model was developed by using discriminant scores which were extracted through five un-weighted weather indices. Two discriminant scores were obtained for each year and used as regressors along with the time trend T in development of the forecasting model. The form of model is given below;

Model-4

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where,

Y is rice crop yield, 'βi's ($i = 0,1,2,3$) are regression coefficients, ds_1 and ds_2 are two sets of discriminant scores based on un-weighted weather indices, T is the time trend variable and ε is error term assumed to follow NID ~ (0, σ^2).

Method-5

Similar to Model-4, five weighted weather indices were used to extract discriminant scores using discriminant function analysis. The forecasting model was fitted taking the *Kharif* rice yield as the regress and two sets of scores (ds_1 and ds_2) and the time trend T as the regressors. The form of the model equation is given below;

Model-5

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where,

ds_1 and ds_2 are two sets of discriminant scores based on weighted weather indices, Y, 'βi's ($i = 0$, 1, 2, 3) and T areas mentioned in Model-4.

Method-6

The method followed same procedure as mentioned in Method-3, for each phase up to obtaining simple average of the individual weather variables. The phase-wise discriminant scores were obtained through discriminant function analysis. Thus, in all, six scores were obtained for each year and model was fitted using these six discriminant scores along with time trend (T) as regressors and *Kharif* rice yield as regress and. The form of regression equation is mentioned below;

Model-6

$$Y = \beta_0 + \sum_{l=1}^{2} \sum_{m=1}^{3} \beta_{lm} \, ds_{lm} + \beta_{11} T + \varepsilon$$

where,

Y is the rice crop yield, β_0 is the intercept of the model, 'β_{lm}'s ($l = 1$, 2; $m = 1$, 2, 3) and β_{11} are the regression coefficients, ds_{lm} is the lth discriminant score in m^{th} phase, T is the time trend variable (year) and ε is error NID ~ (0, σ^2).

2.2.2. *Logistic regression*

Logistic regression is mathematical modeling approach that can be used to describe the relationship of

several variables to a binary/dichotomous dependent variable. Cox (1958) and Walker and Duncan (1967) are pioneer to logistic regression analysis. The year have been divided into two and three groups on the basis of de-trended rice crop yield as explained above in section 2.2.1.

2.2.2.1. *Development of Models Based on Two groups*

Three different models, *viz.*, Model-7, 8 and 9 were developed considering data of weather indices as mentioned in method-1, 2 and 3 respectively. The log it probabilities were worked out by weather indices using logistic regression analysis. The forecasting model was fitted taking the *Kharif* rice yield as the explained variable and the log it probability (Ps_1) and the time trend T as the predicting variable. The form of models are given as follows;

Model-7

$$Y = \beta_0 + \beta_1 Ps_1 + \beta_2 T + \varepsilon$$

where,

Y is un-trended crop yield, 'βi's ($i = 0, 1, 2$) are regression coefficients, Ps_1 is the log it probability based on un-weighted weather indices, T is the time trend variable and ε is error term assumed to follow NID $\sim (0, \sigma^2)$.

Model-8

$$Y = \beta_0 + \beta_1 Ps_1 + \beta_2 T + \varepsilon$$

where,

Ps_1 is the log it probability based on weighted weather indices, remaining symbols are same as Model-7.

Model-9

$$Y = \beta_0 + \sum_{l=1}^{1} \sum_{m=1}^{3} \beta_{lm} Ps_{lm} + \beta_{11} T + \varepsilon$$

where,

Y is the rice crop yield, β_0 is the intercept of the model, 'β_{lm}'s ($l=1$; $m=1, 2, 3$) and β_{11} are the regression coefficients, Ps_{lm} is the lth log it probabilities in m^{th} phase, T is the time trend variable (year) and ε is error NID $\sim (0, \sigma^2)$.

2.2.2.2. *Development of Models Based on Three groups*

Three different models, *viz.*, Model-10, 11 and 12 were developed using similar data of weather indices as mentioned in method -4, 5 and 6 respectively. The log it

probabilities were calculated using data of weather indices by logistic regression analysis. Two log it probabilities were obtained for each year. The forecasting model was fitted taking the *Kharif* rice yield as the regress and the two sets of scores (Ps_1 and Ps_2) along with the time trend T as the regressors. The form of models are given as follows;

Model-10

$$Y = \beta_0 + \beta_1 Ps_1 + \beta_2 Ps_2 + \beta_3 T + \varepsilon$$

where,

Y is un-trended crop yield, 'βi's ($i = 0, 1, 2, 3$) are regression coefficients, Ps_1 and Ps_2 are log it probabilities based on un-weighted weather indices, T is the time trend variable and ε is error term assumed to follow NID $\sim (0, \sigma^2)$.

Model-11

$$Y = \beta_0 + \beta_1 Ps_1 + \beta_2 Ps_2 + \beta_3 T + \varepsilon$$

where,

Ps_1 and Ps_2 are two sets of log it probabilities based on weighted weather indices, remaining symbols are same as Model-10.

Model-12

$$Y = \beta_0 + \sum_{l=1}^{2} \sum_{m=1}^{3} \beta_{lm} Ps_{lm} + \beta_{11} T + \varepsilon$$

where,

Y is the rice crop yield, β_0 is the intercept of the model, 'β_{lm}'s ($l = 1, 2$; $m = 1, 2, 3$) and β_{11} are the regression coefficients, Ps_{lm} is the lth log it probabilities in m^{th} phase, T is the time trend variable (year) and ε is error NID $\sim (0, \sigma^2)$.

2.3. *Comparison and validation of models*

The comparisons and validation of models were done using following approaches.

2.3.1. *Forecast error (%)*

The validation of the model using observed yield (O_i) and forecasted yield (E_i) was computed using below formula,

$$\text{Forecast error}(\%) = \left[\frac{(O_i - E_i)}{O_i} \right] \times 100$$

TABLE 1

Forecast models using two group discriminant function analysis

Model	SMW	Model equations	Adj. R^2
Model-1	37	$Y=1797.42+30.84T*+110.35ds_1*$	73.80
Model-2	40	$Y=1786.73+31.73T*+104.84ds_1*$	87.60
Model-3	40	$Y=1741.98+35.45T*+55.66ds_1*+45.76ds_2-64.91ds_3$	75.40

*Significant at P≤0.05, **Significant at P≤0.01

TABLE2

Forecast models using three group discriminant function analysis

Model	SMW	Model equations	Adj. R^2
Model-4	37	$Y=1844.01+26.96T*+93.65ds_1*-60.16ds_2$	75.20
Model-5	39	$Y=1837.99+27.46T*+102.27ds_1*+12.21ds_2$	93.30
Model-6	40	$Y=1829.70+28.15T*+98.01ds_1*-68.05ds_5*$	76.80

*Significant at P≤0.05, **Significant at P≤0.01

TABLE3

Comparison of Forecast models developed using discriminant function analysis

| Model Name	SMW No.	Year	Observed Yield (Kg/ha)	Forecasted Yield (Kg/ha)	Forecast Error (%)	RMSE(% RMSE)	Adj. R^2
Model-1	37	2013	2432	2714	-11.61	133.82 (5.11)	73.80
		2014	2740	2715	0.94		
		2015	2727	2643	3.05		
		2016	2573	2621	-1.85		
Model-2	40	2013	2432	2788	-14.64	161.39 (6.16)	87.60
		2014	2740	2683	2.11		
		2015	2727	2714	0.46		
		2016	2573	2571	0.09		
Model-3	40	2013	2432	2765	-13.68	167.24 (6.38)	75.40
		2014	2740	2678	2.28		
		2015	2727	2697	1.10		
		2016	2573	2729	-6.06		
Model-4	37	2013	2432	2648	-8.89	105.31 (4.02)	75.20
		2014	2740	2695	1.66		
		2015	2727	2661	2.41		
		2016	2573	2622	-1.90		
Model-5	39	2013	2432	2692	-10.70	120.07 (4.59)	93.30
		2014	2740	2724	0.60		
		2015	2727	2682	1.64		
		2016	2573	2619	-1.76		
Model-6	40	2013	2432	2576	-5.92	98.96 (3.78)	76.80
		2014	2740	2677	2.32		
		2015	2727	2785	-2.13		
		2016	2573	2718	-5.61		
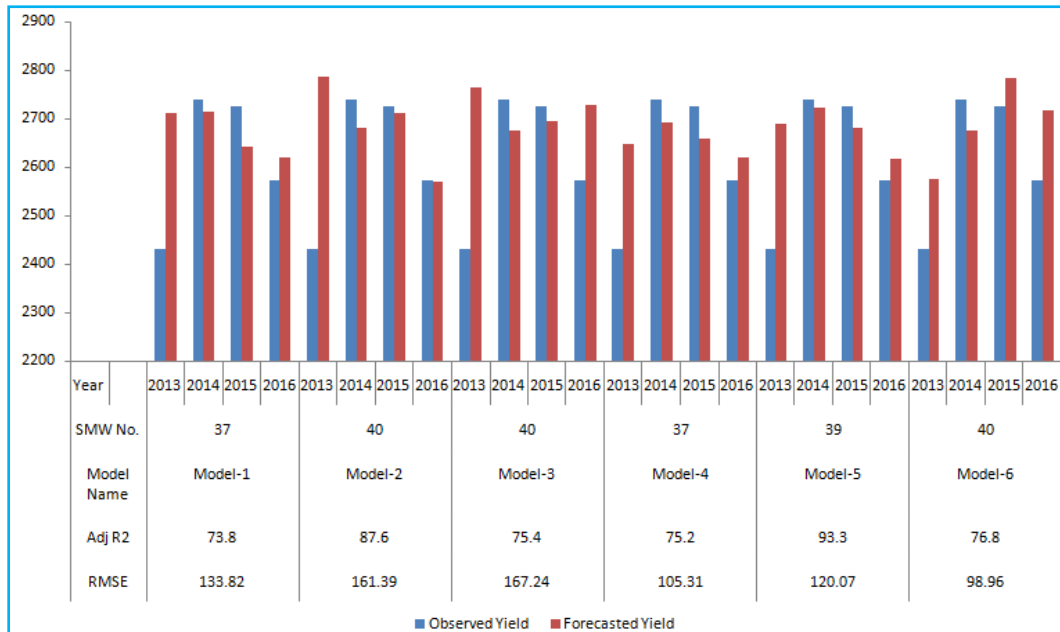
Fig. 2. Graphical representation of Observed and Forecast yield using discriminant function analysis models

TABLE 4

Forecast models using two group logistic regression

Model	SMW	Model equations	Adj. R^2
Model-7	35	$Y=2022.37+29.39T*-397.68Ps_1*$	73.00
Model-8	35	$Y=1942.90+32.06T*-306.91Ps_1*$	80.30
Model-9	40	$Y=2120.22+35.33T*-237.26Ps_1-173.04Ps_2-311.82Ps_3*$	79.10

*Significant at P≤0.05, **Significant at P≤0.01

2.3.2. Coefficient of multiple determination (adj. R^2)

The best fitted model among developed models were decided based on highest value of Adjusted (Adj.) R^2

$$\text{Adj } R^2 = 1 - \frac{\dfrac{SS_{res}}{(n-p)}}{\dfrac{SS_t}{(n-1)}}$$

where,

$ss_{res}/(n-p)$ is the residual mean square.

$ss_t/(n-1)$ is the total mean sum of square.

3. Root Mean Squared Error (RMSE)

The cross validation of the model was done using RMSE, for the year 2013 to 2016 using observed yield

(O_i) and forecasted yield (E_i), computed using below formula,

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

4. Results and discussion

The model development was carried out for each week starting from 35th SMW to 40th SMW by using all twelve models as discussed in above section. The best fit model was selected based on highest value of adjusted R^2 among all twelve models and is discussed hereunder. The equations of the best fit models using discriminant function analysis (sections 2.2.1.1 and 2.2.1.2) are presented in Table 1 and Table 2. It is observed that value of Adj. R^2 varies from 73.80 per cent to 93.30 per cent. The Model-2 (two group) and Model-5 (three group) showed highest Adj. R^2, *i.e.*, 87.60 per cent and 93.30 per cent, respectively. Further validation of selected best fit

TABLE 5

Forecast models using three group logistic regression

Model	SMW	Model equations	Adj. R^2
Model-10	40	$Y=2061.05+29.02T*-480.88Ps_1*-244.75Ps_2$	79.80
Model-11	35	$Y=2025.75+25.41T*-367.58Ps_1*-115.97Ps_2$	83.30
Model-12	40	$Y=2139.21+25.99T*-428.38Ps_1*-442.02Ps_6*$	83.60

*Significant at P≤0.05, **Significant at P≤0.01

TABLE 6

Comparison of Forecast models developed using logistic regression

| Model Name	SMW No.	Year	Observed Yield (Kg/ha)	Forecasted Yield (Kg/ha)	Forecast Error (%)	RMSE (% RMSE)	Adj. R^2
Model-7	35	2013	2432	2704	-11.18	126.08 (4.81)	73.00
		2014	2740	2709	1.15		
		2015	2727	2671	2.03		
		2016	2573	2612	-1.49		
Model-8	35	2013	2432	2712	-11.53	131.32 (5.02)	80.30
		2014	2740	2744	-0.15		
		2015	2727	2776	-1.82		
		2016	2573	2502	2.79		
Model-9	40	2013	2432	2688	-10.51	153.46 (5.86)	79.10
		2014	2740	2652	3.24		
		2015	2727	2725	0.07		
		2016	2573	2784	-8.20		
Model-10	40	2013	2432	2725	-12.06	155.46 (5.94)	79.80
		2014	2740	2627	4.15		
		2015	2727	2582	5.30		
		2016	2573	2605	-1.22		
Model-11	35	2013	2432	2614	-7.47	117.31 (4.48)	83.30
		2014	2740	2607	4.86		
		2015	2727	2612	4.21		
		2016	2573	2643	-2.73		
Model-12	40	2013	2432	2663	-9.52	122.23 (4.67)	83.60
		2014	2740	2694	1.69		
		2015	2727	2711	0.57		
		2016	2573	2710	-5.32		
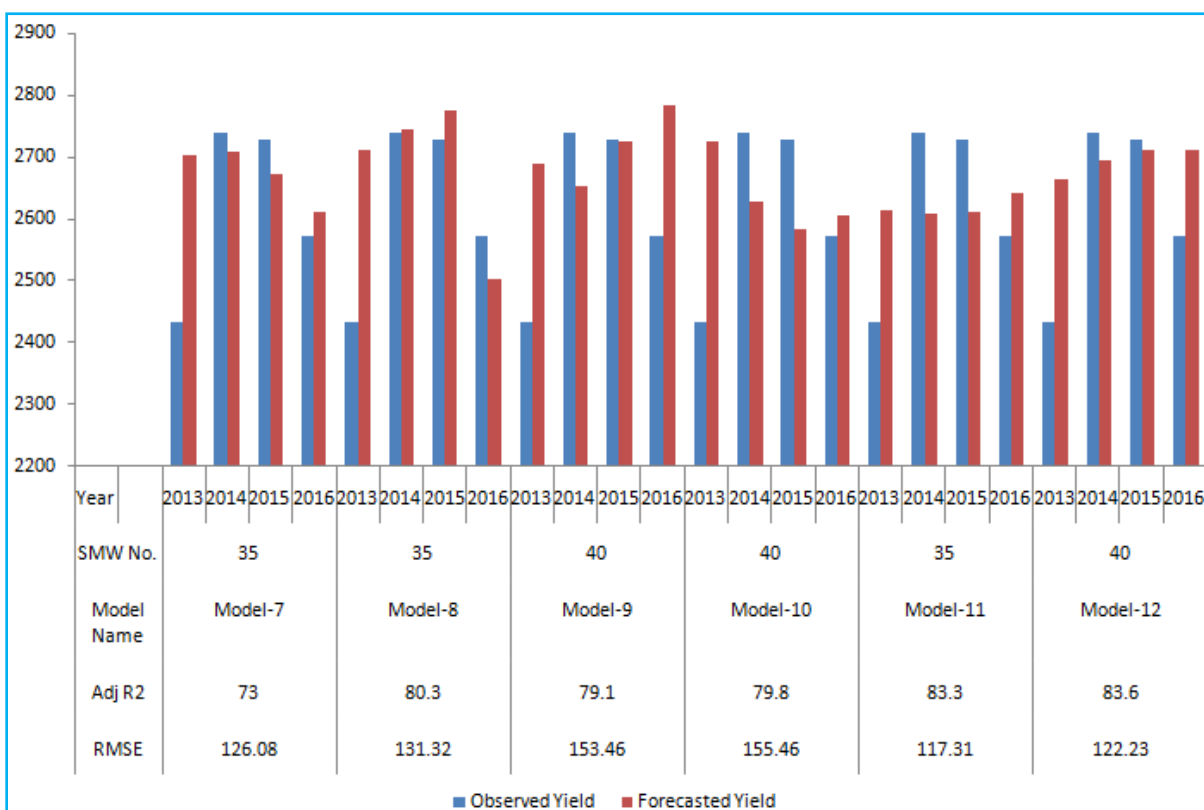
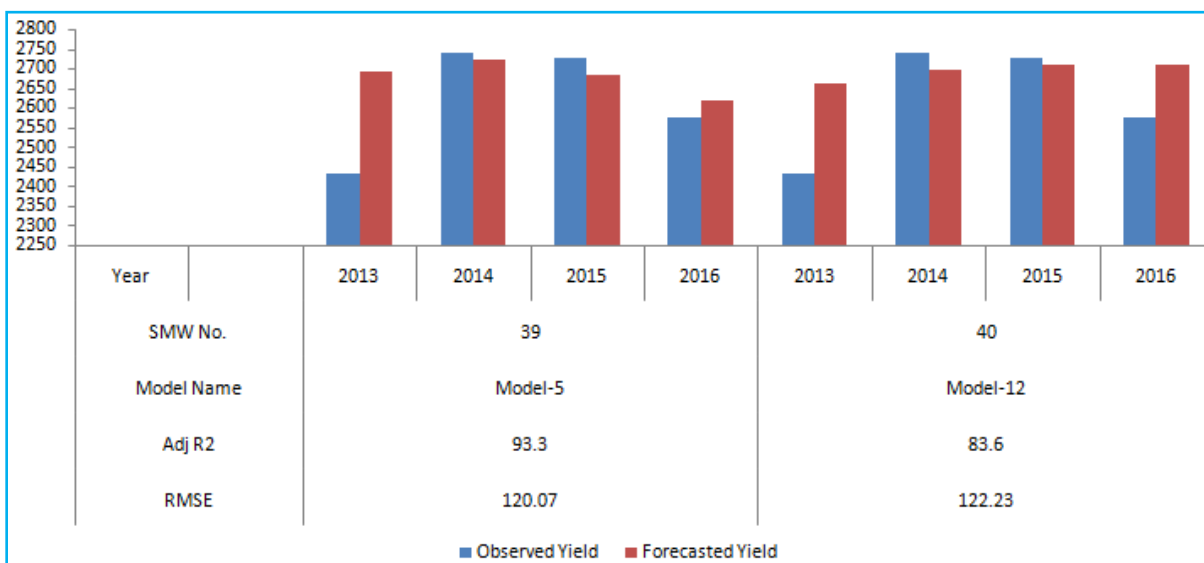**Fig. 3.** Graphical representation of Observed and Forecast yield using Logistic regression models

Fig. 4. Comparison between best fitted models of discriminant function analysis and logistic regression analysis approach

models, developed using discriminant function analysis, was performed utilising data off our years from the year 2013 to 2016 and results are presented in Table 3. The results of validation showed considerably low forecast error (%). Model-5 at 39th SMW showed highest value of Adj. R^2 (93.30) with fairly low value of RMSE (120.07).

Therefore, study revealed that Model-5 is the best model for pre-harvest rice yield forecasting at 39th SMW using discriminant function analysis. The comparison of observed and forecast yield along with RMSE and Adj. R^2 values for models using discriminant function analysis is graphically presented in Fig. 2.

Similarly, the best fitted models, which were developed by using logistic regression as discussed in sections 2.2.2.1 and 2.2.2.2, are presented in Table 4 and Table 5, respectively. It is observed that value of Adj. R^2 varies from 73.00 per cent to 83.60 per cent. The Model-8 (two group) and Model-12(three group) showed highest Adj. R^2, *i.e.*, 80.30 per cent and 83.60 per cent, respectively. Results of further validation of best fit models, developed using logistic regression, indicated relatively low forecast error (%) (Table 6). Low value of RMSE (122.23) and highest value of Adj. R^2 (83.60 per cent) revealed that Model-12 at 40th SMW is the best model for pre-harvest rice yield forecasting among the models developed using logistic regression, even though performance of Model-11 at 35th SMW is at par with that of Model-12. Accordingly, the comparison of observed and forecast yield along with RMSE and Adj. R^2 values is graphically presented in Fig. 3.

The present investigation was undertaken to forecast *Kharif* rice yield well in advance based on categorical dependent variables. Two different statistical approaches were applied for development of pre-harvest forecast models, *viz.*, discriminant function analysis and logistic regression. Similar method, *i.e.*, discriminant function analysis was utilized to develop pre-harvest forecasting models and found significant by Agrawal *et al.* (2012), Sisodia *et al.* (2014), Garde *et al.* (2015) and Goyal (2016). Similarly Kumari *et al.* (2016), Sudesh *et al.* (2016) and Garde *et al.* (2020) utilised logistic regression approach for development of forecasting models with substantial results.

The comparison of different statistical models helps to investigate best statistical approach for pre harvest yield forecast. In the current study comparison between statistical approaches of discriminant function analysis and logistic regression was made by using validated values of best fit models. It is revealed from Fig. 4 that Model-5 showed comparatively low forecast error (%) with highest value of Adj. R^2 (93.30) and lowest value of RMSE (120.07) as compared to Model-12. It is also observed that Model-5 is able to generate yield forecast one week prior (39th SMW) to Model-12 (40th SMW). Therefore, it can be concluded that discriminant function analysis (Model-5) is found better in predicting pre-harvest rice yield than logistic regression technique (Model-12). Further discriminant loadings

TABLE 7

Correlation between discriminant score and weather variable (discriminant loadings)

Weather Indices	ds_1	ds_2
Z_{11}	0.79**	-0.19
Z_{21}	0.40	0.40
Z_{31}	0.83**	-0.03
Z_{41}	0.88**	0.14
Z_{51}	0.82**	0.40

[*]Significant at P≤0.05, [**]Significant at P≤0.01

under discriminant function analysis were obtained due to significant weather indices through correlation study (Table 7). It indicated that weighted maximum temperature, rainfall, morning and evening relative humidity played significant role in classifying the data and extracting significant discriminant scores. The comparison of observed and forecast yield along with RMSE and Adj. R^2 values is graphically presented in Fig. 4.

Similar results were reported by Kumar *et al.* (2016) regarding weighted maximum temperature along with time trend in rice yield prediction in Navsari, whereas, maximum and minimum temperature in Surat district and weighted minimum temperature, morning relative humidity and rainfall in Valsad, Narmada, Tapi and Dangs. Pandey *et al.* (2015) found weighted rainfall as an important weather variable in rice yield prediction in Faizabad district of Uttar Pradesh. Similarly, Biswas *et al.* (2017) noticed significance of maximum temperature and weighted relative humidity in rice yield prediction. The entry of time trend in model explained about technological change, *viz.*, change in pesticide or fertilizer application or release of new varieties leading to drastic increase in yield of *kharif* rice and results were confirmed with Agnihotri and Sridhara (2014), Pandey *et al.* (2015) and Kumar *et al.* (2016). This result confirmed that maximum temperature, rainfall, morning and evening relative humidity have important role in predicting *Kharif* rice yield.

4. Conclusion

In this study, different statistical forecast models were developed at different SMW (from 35th SMW to 40th SMW) keeping the necessity of obtaining the forecast atleast one month before the harvest of rice crop in view. The study revealed that statistical approach of discriminant function analysis was found superior as compared to logistic regression technique. The Model-5

was found competent to forecast rice yield six weeks before (39th SMW) actual harvest of the crop, *i.e.*, during heading stage of the crop growth period. It was also observed that maximum temperature, rainfall, morning and evening relative humidity played important role in predicting *Kharif* rice yield along with time trend. The study also revealed from the obtained outcome that there would be a wide scope for using alternative approaches for developing predictors/indices that may be used in developing the statistical models for reliable and consistent forecast. Such approaches can be applicable in many crops, *viz.*, cereals, pulses, oil seeds, sugarcane etc. and obtained forecast through models will have significant value in agricultural planning and policy making.

Acknowledgement

Disclaimer : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Agnihotri, G. and Sridhara, S., 2014, "Pre -harvest forecasting models for *kharif* rice yield in coastal Karnataka using weather indices", *Journal of Agrometeorology*, **16**, 2, 207-209.

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

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

Anonymous, 2021, Report on "Normal Estimates of Area, Production and Yield of Selected Principal Crops, May 2021", Directorate of Economics & Statistics Agricultural Statistics Division New Delhi.

Biwas, R., Bhattacharyya, B. and Banerjee, S., 2017, "Predicting rice yield from weather variable through detrended production index", *Contemporary Research in India*, **7**, 2, 462-466.

Chauhan, V. S., Shekh, A. M., Dixit, S. K., Mishra, A. P. and Kumar, S., 2009, "Yield prediction model of rice in Bulsar district of Gujarat", *Journal of Agrometeorology*, **11**, 2, 162-168.

Cox, D. R., 1958, "The regression analysis of binary sequences (with discussion)", *Journal of the Royal Statistical Society* B., **20**, 215-242.

Dhekale, B. S., Mahdi, S. and Sawant, P. K., 2014, "Forecast models for groundnut using meteorological variables in Kolhapur, Maharashtra", *Journal of Agrometeorology*, **16**, 2, 238-239.

Fisher, R. A., 1924, "The influence of rainfall on yield of wheat at Rothamsted", *Philosophical Transaction of Royal Society of London*, Series B, **213**, 89-142.

Garde, Y. A., Dhekale, B. S. and Singh, S., 2015, "Different approaches on pre harvest forecasting of wheat yield", *Journal of Applied and Natural Science*, **7**, 2, 839-843.

Garde, Y. A., Shukla, A. K. and Singh, S., 2012, "Pre-harvest forecasting of rice yield using weather indices in Pantnagar", *International Journal of Agricultural Statistical Science*, **8**, 1, 233-241.

Garde**,** Y. A. Thorat, V. S. Pisal R. R. and Shinde V. T., 2020, "Pre harvest forecasting of kharif rice yield using weather parameters for strategic decision making in agriculture", *International Journal of Environment and Climate Change*, **10**, 12, 162-170.

Ghosh, K., Balasubramanian, R., Bandopadhyay, S., Chattopadhyay, N., Singh, K. K. and Rathore, L. S., 2014, "Development of crop yield forecast models under FASAL-a case study of kharif rice in West Bengal", *Journal of Agrometeorology*, **16**, 1, 1-8.

Goyal, M., 2016, "se of different multivariate techniques for pre harvest wheat yield estimation in Hisar (Hariyana)", *International Journal of Computer & Mathematical Sciences*, **5**, 12, 06-11.

Hendricks, W. A. and Scholl, J. C., 1943, "Technique in measuring joint relationship. The joint effect of temperature and precipitation on corn yield", *N. C. Staff Agricultural Experimental Techniques Bulletin*, p74.

Jain, R. C., Agrawal, R. and Jha, M. P., 1980, "Effect of climatic variables on rice yield and its forecast", *MAUSAM*, **31**, 3, 591-596.

Kumar, N., Pisal, R. R., Shukla, S. P. and Pandey, K. K., 2016, "Crop yield forecasting of paddy and sugarcane through modified Hendrick and Scholl technique for south Gujarat", *MAUSAM*, **67**, 2, 405-410.

Kumar, R., Gupta, B. R. D., Athiyaman, B., Singh, K. K. and Shukla, R. K., 1999, "Stepwise regression technique to predict Pigeon pea yield in Varanasi district", *Journal of Agrometeorology,* **1**, 2, 183-186.

Kumari, P., Mishra, G. C. and Srivastava, C. P., 2016, "Statistical models for forecasting pigeon pea yield in Varanasi region", *Journal of Agrometeorology*, **18**, 2, 306-310.

Mahdi, S. S., Lotus, S., Singh, G., Ahmad, L., Singh K. N., Dar, L. A. and Bhat, A., 2013, "Forecast of rice (*Oryza sativa* L.) yield based on climatic parameters in Srinagar district of Kashmir Valley", *Journal of Agrometeorology*, **15**, 1, 89-90.

Pandey, K. K., Rai, V. N., Sisodia, B. V. S. and Singh, S. K., 2015, "Effect of Weather Variables on Rice Crop in Eastern Uttar Pradesh, India", *Plant Archives*, **15**, 1, 575-579.

Patel, G. B., Vaishnav, P. R., Patel, J. S. and Dixit, S. K., 2007, "Pre-harvest forecasting of rice (*Oryza Sativa* L.) yield based on weather variables and technological trend", *Journal of Agrometeorology*, **9**, 2, 167-173.

Priya, S. R. K. and Suresh, K. K., 2009, "A study on pre-harvest forecast of sugarcane yield using climatic variables", *Statistics and Applications*, **7&8**, 1&2, 1-8.

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

Sarika, Iquebal, M. A. and Chattopadhyay, 2011, "Modelling and forecasting of pigeonpea (*Cajanuscajan*) production using autoregressive integrated moving average methodology", *Indian Journal of Agricultural Sciences*, **81**, 6, 520-523.

Singh, R. S., Patel, C., Yadav, M. K. and Singh, K. K., 2014, "Yield forecasting of rice and wheat crops for eastern Uttar Pradesh", *Journal of* Agrometeorology, **16**, 2, 199-202.

Sisodia, B. V. S., Yadav, R. R., Kumar, S. 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.

Sudesh, Verma P. and Verma U., 2016 "Using logistic regression to predict wheat yield in western zone of Haryana", *International Journal of Computer & Mathematical Sciences*, **5**, 12, 33-37.

Varmola, S. L., Dixit, S. K., Patel, J. S. and Bhatt, H. M., 2004, "Forecasting of wheat yield on the basis of weather variables", *Journal of Agrometeorology*, **6**, 2, 223-228.

Walker, S. H. and Duncan, D. B., 1967, "Estimation of the probability of an event as a function of several independent variables", *Biometrika*, **54**, 167-178.