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, Gol, 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)

DOI : https://doi.org/10.54302/mausam.v74i3.3495
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

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
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 ($Z_{i'j}$) were generated by taking the products of weekly weather variables, two at a time. The forms of indices are given below:

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

where,

\begin{align*}
    j &= 0, 1 \text{ (where, ‘0’ represents un-weighted indices and ‘1’ represents weighted indices)} \\
    m &= \text{Week of forecast} \\
    w &= \text{week number (1, 2, ..., m)} \\
    r_{iw} &= \text{Correlation coefficient between de-trended crop yield and } i^{th} \text{ weather variable in } w^{th} \text{ week} \\
    r_{i'w} &= \text{Correlation coefficient between de-trended crop yield and the product of } i \text{ and } i' \text{ weather variable in } w^{th} \text{ week} \\
    X_{iw} \text{ and } X_{i'w} \text{ are the } i \text{ and } i' \text{ weather variable in } w^{th} \text{ week respectively}
\end{align*}

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;

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

where,

$Y$ is crop yield, $\beta_i$'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 $\epsilon$ is error term assumed to follow NID ~ (0, $\sigma^2$).

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;

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 T + \epsilon$$
where,

$ds_1$ is the discriminant score based on weighted weather indices and $Y$, $\beta_l's$ and $T$ areas mentioned in 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 Chandrasah, 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

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

where,

$Y$ is the rice crop yield, $\beta_0$ is intercept of the model, $\beta_{lm}$'s ($l = 1, m = 1, 2, 3$) and $\beta_{11}$ are the regression coefficients, $ds_{lm}$ is the $l^{th}$ discriminant score in $m^{th}$ phase, $T$ is the time trend variable (year) and $\varepsilon$ is error term assumed to follow NID $\sim (0, \sigma^2)$.

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

$$Y = \beta_0 + \beta_{11} ds_1 + \beta_{22} ds_2 + \beta_{11} T + \varepsilon$$

where,

$Y$ is rice crop yield, $\beta_l's$ ($i = 0, 1, 2, 3$) 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 $\varepsilon$ is error term assumed to follow NID $\sim (0, \sigma^2)$.

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

$$Y = \beta_0 + \beta_{11} ds_1 + \beta_{22} ds_2 + \beta_{11} T + \varepsilon$$

where,

$ds_1$ and $ds_2$ are two sets of discriminant scores based on weighted weather indices, $Y$, $\beta_l's$ ($i = 0, 1, 2, 3$) 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

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

where,

$Y$ is the rice crop yield, $\beta_0$ is the intercept of the model, $\beta_{lm}$'s ($l = 1, 2; m = 1, 2, 3$) and $\beta_{11}$ are the regression coefficients, $ds_{lm}$ is the $l^{th}$ discriminant score in $m^{th}$ phase, $T$ is the time trend variable (year) and $\varepsilon$ is error term assumed to follow NID $\sim (0, \sigma^2)$.

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 ($P_s$) and the time trend $T$ as the regressors. The form of models are given as follows;

**Model-7**

\[
Y = \beta_0 + \beta_1 P_s + \beta_2 T + \varepsilon
\]

where,

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

**Model-8**

\[
Y = \beta_0 + \beta_1 P_s + \beta_2 T + \varepsilon
\]

where,

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

**Model-9**

\[
Y = \beta_0 + \sum_{l=1}^{3} \sum_{m=1}^{3} \beta_{lm} P_{slm} + \beta_{11} T + \varepsilon
\]

where,

$Y$ is the rice crop yield, $\beta_0$ is the intercept of the model, \( \beta_{lm} \)'s (l = 1, 2; m = 1, 2, 3) and $\beta_{11}$ are the regression coefficients, $P_{slm}$ is the $l^{th}$ log it probabilities in $m^{th}$ phase, $T$ is the time trend variable (year) and $\varepsilon$ 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 ($P_{s1}$ and $P_{s2}$) along with the time trend $T$ as the regressors. The form of models are given as follows;

**Model-10**

\[
Y = \beta_0 + \beta_1 P_{s1} + \beta_2 P_{s2} + \beta_3 T + \varepsilon
\]

where,

$Y$ is un-trended crop yield, \( \beta_i \)'s (i = 0, 1, 2, 3) are regression coefficients, $P_{s1}$ and $P_{s2}$ are log it probabilities based on un-weighted weather indices, $T$ is the time trend variable and $\varepsilon$ is error term assumed to follow $NID \sim (0, \sigma^2)$.

**Model-11**

\[
Y = \beta_0 + \beta_1 P_{s1} + \beta_2 P_{s2} + \beta_3 T + \varepsilon
\]

where,

$P_{s1}$ and $P_{s2}$ are two sets of log it probabilities based on weighted weather indices, remaining symbols are same as Model-10.

**Model-12**

\[
Y = \beta_0 + \sum_{l=1}^{3} \sum_{m=1}^{3} \beta_{lm} P_{slm} + \beta_{11} T + \varepsilon
\]

where,

$Y$ is the rice crop yield, $\beta_0$ is the intercept of the model, \( \beta_{lm} \)'s (l = 1, 2; m = 1, 2, 3) and $\beta_{11}$ are the regression coefficients, $P_{slm}$ is the $l^{th}$ log it probabilities in $m^{th}$ phase, $T$ is the time trend variable (year) and $\varepsilon$ 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

<table>
<thead>
<tr>
<th>Model</th>
<th>SMW</th>
<th>Model equations</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model-1</td>
<td>37</td>
<td>$Y=1797.42+30.84T^<em>+110.35ds_1^</em>$</td>
<td>73.80</td>
</tr>
<tr>
<td>Model-2</td>
<td>40</td>
<td>$Y=1786.73+31.73T^<em>+104.84ds_1^</em>$</td>
<td>87.60</td>
</tr>
<tr>
<td>Model-3</td>
<td>40</td>
<td>$Y=1741.98+35.45T^<em>+55.66ds_1^</em>+45.76ds_2^<em>-64.91ds_3^</em>$</td>
<td>75.40</td>
</tr>
</tbody>
</table>

*Significant at P≤0.05; **Significant at P≤0.01

### Table 2
Forecast models using three group discriminant function analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>SMW</th>
<th>Model equations</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model-4</td>
<td>37</td>
<td>$Y=1844.01+26.96T^<em>+93.65ds_1^</em>-60.16ds_2^*$</td>
<td>75.20</td>
</tr>
<tr>
<td>Model-5</td>
<td>39</td>
<td>$Y=1837.99+27.46T^<em>+102.27ds_1^</em>+12.21ds_2^*$</td>
<td>93.30</td>
</tr>
<tr>
<td>Model-6</td>
<td>40</td>
<td>$Y=1829.70+28.15T^<em>+98.01ds_1^</em>-68.05ds_5^*$</td>
<td>76.80</td>
</tr>
</tbody>
</table>

*Significant at P≤0.05; **Significant at P≤0.01

### Table 3
Comparison of Forecast models developed using discriminant function analysis

<table>
<thead>
<tr>
<th>Model No.</th>
<th>SMW No.</th>
<th>Observed Yield (Kg/ha)</th>
<th>Forecasted Yield (Kg/ha)</th>
<th>Forecast Error (%)</th>
<th>RMSE(%) RMSE Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model-1</td>
<td>37</td>
<td>2013: 2432</td>
<td>2714</td>
<td>-11.61</td>
<td>133.82(5.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014: 2740</td>
<td>2715</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015: 2727</td>
<td>2643</td>
<td>3.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016: 2573</td>
<td>2621</td>
<td>-1.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013: 2432</td>
<td>2788</td>
<td>-14.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014: 2740</td>
<td>2683</td>
<td>2.11</td>
<td>161.39(6.16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015: 2727</td>
<td>2714</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016: 2573</td>
<td>2571</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013: 2432</td>
<td>2765</td>
<td>-13.68</td>
<td>167.24(6.38)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014: 2740</td>
<td>2678</td>
<td>2.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015: 2727</td>
<td>2697</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016: 2573</td>
<td>2729</td>
<td>-6.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013: 2432</td>
<td>2648</td>
<td>-8.89</td>
<td>105.31(4.02)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014: 2740</td>
<td>2695</td>
<td>1.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015: 2727</td>
<td>2661</td>
<td>2.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016: 2573</td>
<td>2622</td>
<td>-1.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013: 2432</td>
<td>2692</td>
<td>-10.70</td>
<td>120.07(4.59)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014: 2740</td>
<td>2724</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015: 2727</td>
<td>2682</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016: 2573</td>
<td>2619</td>
<td>-1.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013: 2432</td>
<td>2576</td>
<td>-5.92</td>
<td>98.96(3.78)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014: 2740</td>
<td>2677</td>
<td>2.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015: 2727</td>
<td>2785</td>
<td>-2.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016: 2573</td>
<td>2718</td>
<td>-5.61</td>
<td></td>
</tr>
</tbody>
</table>
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.) $R^2$.

\[
\text{Adj } R^2 = 1 - \frac{SS_{res}}{SS_T} \frac{(n-p)}{(n-1)}
\]

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

<table>
<thead>
<tr>
<th>Model</th>
<th>SMW</th>
<th>Model equations</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model-10</td>
<td>40</td>
<td>$Y=2061.05+29.02T*-480.88P_{s1}*-244.75P_{s2}$</td>
<td>79.80</td>
</tr>
<tr>
<td>Model-11</td>
<td>35</td>
<td>$Y=2025.75+25.41T*-367.58P_{s1}*-115.97P_{s2}$</td>
<td>83.30</td>
</tr>
<tr>
<td>Model-12</td>
<td>40</td>
<td>$Y=2139.21+25.99T*-428.38P_{s1}*-442.02P_{s6}$</td>
<td>83.60</td>
</tr>
</tbody>
</table>

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

### TABLE 6
Comparison of Forecast models developed using logistic regression

<table>
<thead>
<tr>
<th>Model Name</th>
<th>SMW No.</th>
<th>Year</th>
<th>Observed Yield (Kg/ha)</th>
<th>Forecasted Yield (Kg/ha)</th>
<th>Forecast Error (%)</th>
<th>RMSE (%)</th>
<th>Adj. R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model-7</td>
<td>35</td>
<td>2013</td>
<td>2432</td>
<td>2704</td>
<td>-11.18</td>
<td>126.08 (4.81)</td>
<td>73.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>2740</td>
<td>2709</td>
<td>1.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td>2727</td>
<td>2671</td>
<td>2.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016</td>
<td>2573</td>
<td>2612</td>
<td>-1.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-8</td>
<td>35</td>
<td>2013</td>
<td>2432</td>
<td>2712</td>
<td>-11.53</td>
<td>131.32 (5.02)</td>
<td>80.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>2740</td>
<td>2744</td>
<td>-0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td>2727</td>
<td>2776</td>
<td>-1.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016</td>
<td>2573</td>
<td>2502</td>
<td>2.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-9</td>
<td>40</td>
<td>2013</td>
<td>2432</td>
<td>2688</td>
<td>-10.51</td>
<td>153.46 (5.86)</td>
<td>79.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>2740</td>
<td>2652</td>
<td>3.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td>2727</td>
<td>2725</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016</td>
<td>2573</td>
<td>2784</td>
<td>-8.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-10</td>
<td>40</td>
<td>2013</td>
<td>2432</td>
<td>2725</td>
<td>-12.06</td>
<td>155.46 (5.94)</td>
<td>79.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>2740</td>
<td>2627</td>
<td>4.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td>2727</td>
<td>2582</td>
<td>5.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016</td>
<td>2573</td>
<td>2605</td>
<td>-1.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-11</td>
<td>35</td>
<td>2013</td>
<td>2432</td>
<td>2614</td>
<td>-7.47</td>
<td>117.31 (4.48)</td>
<td>83.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>2740</td>
<td>2607</td>
<td>4.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td>2727</td>
<td>2612</td>
<td>4.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016</td>
<td>2573</td>
<td>2643</td>
<td>-2.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-12</td>
<td>40</td>
<td>2013</td>
<td>2432</td>
<td>2663</td>
<td>-9.52</td>
<td>122.23 (4.67)</td>
<td>83.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>2740</td>
<td>2694</td>
<td>1.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2015</td>
<td>2727</td>
<td>2711</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2016</td>
<td>2573</td>
<td>2710</td>
<td>-5.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 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  

<table>
<thead>
<tr>
<th>Weather Indices</th>
<th>$d_{s1}$</th>
<th>$d_{s2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_{s1}$</td>
<td>0.79**</td>
<td>-0.19</td>
</tr>
<tr>
<td>$Z_{s2}$</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>$Z_{s3}$</td>
<td>0.83**</td>
<td>-0.03</td>
</tr>
<tr>
<td>$Z_{s4}$</td>
<td>0.88**</td>
<td>0.14</td>
</tr>
<tr>
<td>$Z_{s5}$</td>
<td>0.82**</td>
<td>0.40</td>
</tr>
</tbody>
</table>

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

The author acknowledges Dept. of Agril. Engineering, N. M. College of Agriculture, Navsari Agricultural University, Navsari for providing weather data for present research work.

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


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.


