

## Pre-harvest forecast of *kharif* rice yield using PCA and MLR technique in Navsari district of Gujarat

K.B. BANAKARA<sup>1\*</sup>, H.R. PANDYA<sup>1</sup> and Y.A. GARDE<sup>2</sup>

<sup>1</sup>Department of Agricultural Statistics, Navsari Agricultural University, Navsari, Gujarat – 396 450

<sup>2</sup>Department of Agricultural Statistics, College of Agriculture, Navsari Agricultural University, Waghai, Dang, Gujarat – 394 730

\*Corresponding author: kantheshbanakara@gmail.com

### ABSTRACT

In this paper Principal Components (PC) and Multiple Linear Regression (MLR) Technique were used for development of pre-harvest model for rice yield in the Navsari district of south Gujarat. The weather indices were developed and utilized for development of pre-harvest forecast models. The data of rice yield and weather parameters from 1990 to 2012 were utilized. The cross validation of the developed forecast model were confirmed using data of the years 2013 to 2016. It was observed that value of  $Adj. R^2$  varied from 89 to 96. The appropriate forecast model was selected based on high value of  $Adj. R^2$ . Based on the outcomes in Navsari district, MLR techniques found to be better than PCA for pre harvest forecasting of rice crop yield. The Model-2 found competent to forecast rice yield in Navsari district before eight weeks of actual harvest of crop (37<sup>th</sup> SMW) *i.e* during reproductive stage of the crop growth period.

**Keywords:** Weather indices, MLR techniques, PCA, forecast

Indian rural economy mainly depends on Agriculture. Rice is the most important staple food in Asia. More than 90 per cent of the world's rice is grown and consumed in Asia, where 60 per cent of the world's population lives. India ranks second with 154.6 million tonnes of paddy (FAO, 2015). In the Gujarat state, rice occupies about 7 per cent of the gross cropped area of the state and accounts for around 14 per cent of the total food grain production. About 90 per cent of area under rice is confined to South and middle Gujarat.

The forecast of yield information is a mandatory for government organizations, agro-based industries, traders and agriculturists. Such information is utilized by government organization in policy decisions in regards to food security and distribution, buffer stocking, import-export, price fixation and marketing of agriculture commodities. Pre-harvest forecast is a statistical approach to foretell yield before harvest of the crop. The various organizations in India and also in foreign country are engaged in developing methodology for pre-harvest forecast of crop yield using various approaches which include inputs, weather parameters, plant characters and remote sensed data. Mallick *et al.* (2007) developed a modified statistical model to forecast the crop yield based on weather variables. Some other investigators (Patel *et al.*, 2007; Chauhan *et al.*, 2009; Mahdi *et al.*, 2013; Singh *et al.*, 2014, studied the relationship of weather parameters and rice crop

yield in different regions of world. Varmola *et al.* (2004), Bal *et al.* (2004) and Sisodia *et al.* (2014) developed forecast models for wheat crop in different regions of India.

Although the state of Gujarat has some small area under rice in some pocket but its importance of food basket cannot be ignored in the state. The use of statistical models in forecasting food production and prices for agriculture hold great significance. Therefore an effort is made in the present paper to develop statistical models for pre-harvest forecast of the rice yield based on PCA and MLR.

### MATERIALS AND METHODS

The present study was carried out in the Navsari district of South Gujarat. Considering the specific objectives of the study, *kharif* rice yield data were collected from the Directorate of Economics and Statistics, Government of Gujarat, Gandhinagar, Gujarat from 1990 to 2016. The study utilized weekly weather data which were collected from the Department of Agrometeorology, Navsari Agricultural University, Navsari. The maximum temperature ( $X_1$ ), minimum temperature ( $X_2$ ), Morning relative humidity ( $X_3$ ), Evening relative humidity ( $X_4$ ), and total rain fall ( $X_5$ ) considered for studying the effect on *kharif* rice grain yield. The weekly weather data related to *kharif* rice crop season starting from a first fortnight before sowing to last of

reproductive stage were utilized for the development of statistical models. Therefore for the each year weather data, from May-June (23<sup>rd</sup> standard meteorological week, SMW) to October (40<sup>th</sup> SMW) were utilized for *kharif* rice crop.

**Development of weather indices for yield forecasting**

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

Where,

$j=0, 1$  (where, '0' represents un-weighted indices, '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  $w^{th}$  week,  $r_{ii^*w}$ =Correlation coefficient between adjusted crop yield and the product of  $i$  and  $i^*$  weather variable in  $w^{th}$  week,  $X_{iw}$  and  $X_{i^*w}$  are the  $i$  and  $i^*$  weather variable in  $w^{th}$  week respectively

**Principal component analysis**

Principal component analysis (PCA) is a multivariate statistical technique which reduces data with large number of correlated variables into substantially smaller set of new variables through linear combination of the variables that accounts for total variation present in the original variables. The linear combination of variables is called principal components and estimated using correlation or covariance matrix. When the variables measured with different units, scale effects can influence the composition of derived components. In this case it is necessary to standardize the original variables. Therefore, correlation matrix is considered to better as it does not require standardization.

Let  $X=[x_1, x_2, \dots, x_p]'$  be the  $p$ -dimension random vector. Assuming that mean of  $X$  is  $\theta$  and variance covariance matrix  $\Sigma$  which is real positive definite matrix. The non-zero eigen values of  $\Sigma$  are  $\lambda_1 > \lambda_2 > \dots > \lambda_p$  and corresponding eigenvectors  $a_1, a_2, \dots, a_p$ . For distinct  $\lambda_i$ 's ( $i=1,2,\dots,p$ ); a  $p \times p$  orthogonal matrix  $\Gamma$  can be formed

The orthogonal transformation of  $X$  vector to  $Y$  vector by

$$Y = \Gamma'X$$

$$y_1 = a_1'X = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$$

$$y_2 = a_2'X = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p$$

$$\dots$$

$$y_p = a_p'X = a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p$$

Where,  $Y = [y_1, y_2, \dots, y_p]$  are the principal components, the transformation have

$$E(Y) = 0$$

$$V(Y) = E(Y Y') = E(\Gamma' X X' \Gamma) = \Gamma' \Sigma \Gamma = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_p]$$

$$\sum_{i=1}^p \text{Var}(X_i) = \text{tr}(\Sigma) = \text{tr}(\Lambda) = \sum_{i=1}^p \text{Var}(Y_i)$$

Where,  $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_p]$

This shows that the total variation of vector  $X$  remains same even after transformation of  $X$  vector to a vector of principal components  $Y$ .

Because of positive definite nature of matrix  $\Sigma$ , the extracted  $\lambda_i$ 's are surely positive. The first principal component ( $y_1$ ) has maximum variance equal to  $\lambda_1$ , followed by  $y_2$  with  $\lambda_2$  and so on. The percentage of variation of  $X$  explained by  $i^{th}$  principal component was given below.

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \times 100$$

**Number of principal components to be retained**

Principal component analysis extract as many as components equal to the number of original variables. It is necessary to retain some of important components for further analysis. Some of the approaches discussed here

**Kaiser's criterion**

In principal component analysis, Kaiser's criterion is a powerful and promising component retention method, because it is based on distribution theory of eigenvalues, shows good performance, easily visualized and computed. With this approach retention and interpretation of any component with eigenvalues greater than unity. Any component that displays an eigenvalue greater than unity, accounts for a greater amount of variation in PCs.

**The scree test**

Scree test is carried out by plotting eigenvalues associated with each components and look for a break between the components with relatively large eigenvalues and those with small eigenvalues. The components that appear before the break are assumed to be meaningful and retained for rotation, those appearing after the break are assumed to be unimportant and are not retained.

In this model 30 weather indices (unweighted and

**Table 1:** Week wise correlation coefficient between rice yield and weather parameters for Navsari district

SMW	$T_{\max}(X_1)$	$T_{\min}(X_2)$	MRH( $X_3$ )	ERH( $X_4$ )	RAINFALL ( $X_5$ )
23	0.29	0.16	0.18	-0.11	0.01
24	0.34	0.31	0.01	-0.18	-0.25
25	0.54**	0.33	0.10	-0.29	-0.18
26	0.04	0.11	0.18	-0.02	0.17
27	0.07	0.10	0.21	0.07	0.18
28	-0.05	0.06	0.21	0.00	-0.29
29	0.11	0.11	0.52*	0.12	0.15
30	-0.24	0.05	0.72**	0.41*	0.40
31	-0.31	0.04	0.45*	0.43*	0.19
32	-0.13	0.13	0.25	0.28	0.15
33	0.09	0.21	0.16	-0.19	-0.13
34	0.28	0.03	0.05	-0.31	-0.21
35	-0.10	-0.08	0.21	0.21	0.57**
36	0.09	0.13	0.45*	0.27	0.36
37	0.12	0.27	0.25	0.17	0.28
38	0.10	0.28	0.19	0.21	0.26
39	-0.09	-0.03	0.22	0.04	-0.14
40	-0.04	0.02	-0.19	-0.11	-0.01
41	0.17	-0.18	-0.24	-0.40	-0.13
42	0.30	-0.09	-0.02	-0.13	-0.14

**Table 2:** Per cent variation and cumulative variation of PCA, model-2

PCs	% VAR	CUM VAR	% VAR	CUMVAR	% VAR	CUM VAR
<b>SMW</b>	<b>35</b>		<b>36</b>		<b>37</b>	
<b>PC1</b>	35.72	35.72	36.44	36.44	36.92	39.91
<b>PC2</b>	26.97	62.69	28.53	64.97	29.45	66.36
<b>PC3</b>	17.01	79.70	14.85	79.82	14.19	80.55
<b>PC4</b>	9.98	89.68	10.53	90.36	10.04	90.58
<b>PC5</b>	4.02	93.70	3.85	94.20	3.70	94.28
<b>SMW</b>	<b>38</b>		<b>39</b>		<b>40</b>	
<b>PC1</b>	37.03	37.03	37.24	37.24	38.18	38.18
<b>PC2</b>	29.85	66.88	29.48	66.71	27.89	66.07
<b>PC3</b>	14.53	81.41	14.43	81.14	15.12	81.19
<b>PC4</b>	9.42	90.83	9.41	90.55	9.46	90.66
<b>PC5</b>	3.57	94.40	3.79	94.34	3.51	94.17

weighted,  $j=0,1$ ) were utilized to develop principal components. The number of components retained using scree plot and Kaiser's criterion. The developed components were utilized for model development using regression analysis.

#### Model-1

$$Y = \beta_0 + \beta_1 T + \beta_2 PC_1 + \beta_3 PC_2 + \beta_4 PC_3 + \dots + \beta_{k+1} PC_k + \varepsilon$$

Where,

$Y$  is the rice crop yield,  $T$  is the trend variable,  $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_{k+1}$  are the regression coefficients,  $PC_1, PC_2, PC_3, \dots, PC_k$  are extracted principal components with maximum variance,  $\varepsilon$  is the error term

#### Multiple linear regression models (MLR):

Multiple linear regression models are the functional relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. The pre-harvest forecast models were obtained by applying the MLR techniques by taking predictors as appropriate unweighted and weighted weather indices. Stepwise regression analysis was used for selecting significant variables (Draper and Smith, 1981). The regression model was as follows:

#### Model-2

The model was developed using 30 weather indices (15 unweighted and 15 weighted indices) as a independent variable and crop yield as dependent variable. The developed model is given as

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

Where,

$Z_{ij}$  and  $Z_{i'i'}$  are the weather indices,  $i, i'=1, 2, \dots, p$ ,  $p$ =Number of weather variables under study,  $Y$ = District total crop yield ( $\text{kg ha}^{-1}$ ),  $T$ =Year number (trend parameter),  $A_0$  is the intercept,  $a_{ij}$  and  $a_{i'i'}$ ,  $c$  are the regression coefficient,  $\varepsilon$  is error term normally distributed with mean zero and constant variance

#### Comparison and validation of models

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

#### Forecast error (%)

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

$$\text{Forecast Error} = \left[ \frac{O_i - E_i}{O_i} \right] \times 100$$

#### Coefficient of multiple determination (Adjusted $R^2$ )

The best fitted model among developed models were decided based on highest value of Adjusted  $R^2$

$$R_{adj}^2 = 1 - \frac{SS_{res}/(n-p)}{SS_t/(n-1)}$$

Where,

$ss_{res}/(n-p)$  is the residual mean square,  $ss_t/(n-1)$  is the total mean sum of square.

The best fitted models were selected based on lower RMSE, MAE and MAPE cross validation was done for the year 2013 to 2016 using observed yield ( $O_i$ ) and forecasted yield ( $E_i$ ) was computed using below formula,

#### Root mean square error (RMSE)

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

#### Mean absolute error

$$MAE = \frac{\sum_{i=1}^n |O_i - E_i|}{n}$$

#### Mean absolute percentage error

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{O_i - E_i}{O_i} \right|}{n}$$

## RESULTS AND DISCUSSION:

#### Association between rice yield and weather parameters

The standard method of Karl Pearson correlation coefficient was adopted to study week wise association between rice yield and weather parameters and its computed values presented in Table 1. It was observed that positive and significant correlations between rice yield and some of the weekly weather parameters *viz.* maximum temperature (25<sup>th</sup> SMW *i.e.* vegetative phase of the crop growth stage), Morning Relative Humidity (29, 31, 30 *i.e.* vegetative phase and 36<sup>th</sup> SMW *i.e.* reproductive phase of the crop growth stage), Evening Relative Humidity (30 and 31<sup>st</sup> SMW *i.e.* vegetative phase of the crop growth stage) and Rainfall (35<sup>th</sup> SMW *i.e.*

**Table 3:** Rice yield forecasting model-3 equations

Model	Week	Model	Adj. R <sup>2</sup>
A <sub>1</sub>	35	$Y=1842.01+27.13T+162.11PC_1+88.95PC_2$	91.1
A <sub>2</sub>	36	$Y=1849.84+26.47T+182.08PC_1$	89.9
A <sub>3</sub>	37	$Y=1863.07+25.37T+184.44PC_1$	90.3
A <sub>4</sub>	38	$Y=1861.18+25.53T+185.06PC_1$	90.7
A <sub>5</sub>	39	$Y=1859.15+25.70T+186.32PC_1$	91.4
A <sub>6</sub>	40	$Y=1839.51+27.34T+183.35PC_1$	90.9

**Table 4:** Rice yield forecasting model-1 equations

Model	Week	Model	Adj. R <sup>2</sup>
B <sub>1</sub>	35	$Y=-535.78+27.93T+0.12Z_{341}+0.57Z_{231}$	88.9
B <sub>2</sub>	36	$Y=-857.72+26.42T+0.10Z_{341}+0.58Z_{231}$	89.4
B <sub>3</sub>	37	$Y=-4676.18+24.75T+0.06Z_{351}+20.57Z_{31}-0.16Z_{151}+4.13Z_{20}$	96.4
B <sub>4</sub>	38	$Y=-2433.02+26.61T+0.06Z_{351}+17.57Z_{31}-0.18Z_{151}$	96.0
B <sub>5</sub>	39	$Y=604.67+27.93T+0.06Z_{341}+1.69Z_{121}-5.24Z_{10}+8.82Z_{31}$	95.6
B <sub>6</sub>	40	$Y=-1373.94+25.80T+0.10Z_{341}+1.13Z_{121}+0.05Z_{230}$	94.3

vegetative phase of the crop growth stage). The remaining week wise correlation coefficient between the yield and weather parameters found non-significant. The value of 'r' varies from -0.40 to 0.72, indicating that individual character does not explain more than 51.84 per cent variation in the yield. This suggests that simple regression using single weather parameter is not adequate to forecast the yield. It is necessary to utilize all weather parameters simultaneously. Therefore, study utilized the constructing un-weighted indices and weighted indices which have constructed by taking weighted accumulation of correlation coefficient.

#### **Principal component Analysis**

Principal component regression model was developed using stepwise regression in which extracted principal components were utilized as explanatory variables along with time trend variable and *khari* rice yield as explained variable.

#### **Number of principal components retained**

The number of principal components was retained using Kaiser's criteria and scree plot methods. The number of

principal components along with the per cent of variation and cumulative variation are presented in the Table 2. From the table it was observed that five principal components were retained based on eigen value greater than unity. Cumulative variation of five components was found about than 93 per cent in all weeks. Further scree plot technique was utilized to retain number of principal components by visualizing process and found similar to Kaiser's criterion (Fig 1).

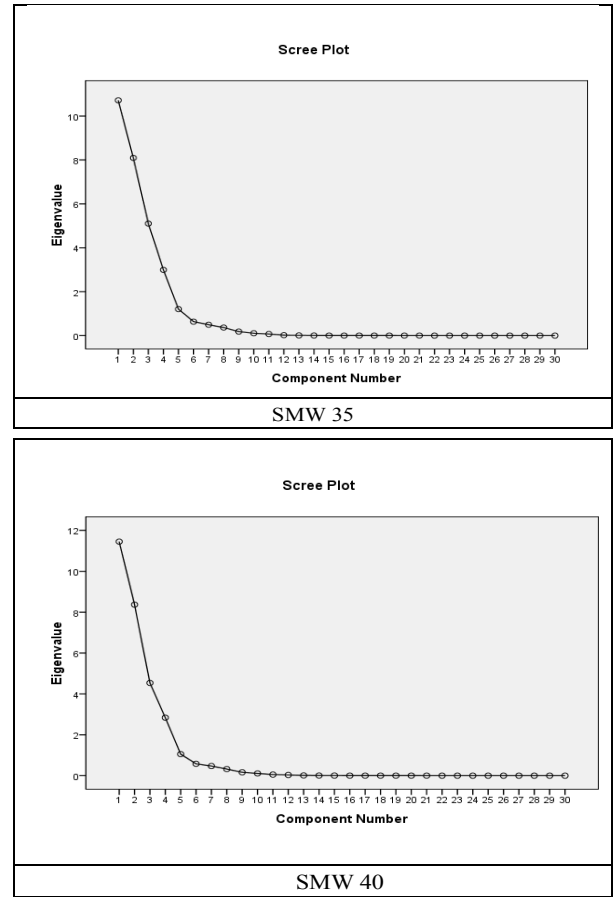
#### **Pre-harvest forecast model-1**

The model was developed using principal components extracted from 15 un-weighted and 15 weighted weather indices. The details of developed model equation are given in Table 3.

The forecast model equations were obtained for each SMW from 35<sup>th</sup> to 40<sup>th</sup> SMW to get early forecast. It was observed that the value of adjusted R<sup>2</sup> in different models vary from 89.90 per cent in model A<sub>2</sub> to 91.40 per cent in model A<sub>5</sub>. The performance of the model was checked using RMSE, MAE and MAPE. The lowest was found RMSE (80.01), MAE (61.25) and MAPE (0.03) in model A<sub>5</sub> (Table 5). The model A<sub>5</sub>

**Table 5:** MLR and PCR model performance for *Kharif* rice yield of Navsari in training data set

SMW	Criteria	MLR	PCR
35	RMSE	88.64	81.17
	MAE	71.79	62.87
	MAPE	0.03	0.03
36	RMSE	86.38	86.75
	MAE	66.79	65.26
	MAPE	0.03	0.03
37	RMSE	<b>47.59</b>	85.07
	MAE	<b>36.48</b>	66.45
	MAPE	<b>0.02</b>	0.03
38	RMSE	51.89	83.27
	MAE	41.56	63.27
	MAPE	0.02	0.03
39	RMSE	52.59	<b>80.01</b>
	MAE	36.66	<b>61.25</b>
	MAPE	0.02	<b>0.03</b>
40	RMSE	61.83	82.09
	MAE	44.44	63.04
	MAPE	0.02	0.03



**Fig. 1:** Scree plot for percent of variation in principal components

**Table 7:** Comparison of MLR and PCA models

Model	SMW	Year	Observed yield	Forecast Yield	forecast Error (%)	RMSE	Adj. $R^2$
A <sub>5</sub>	39	2013	2432	2694	-10.80	136.55	91.4
		2014	2740	2733	0.26		
		2015	2727	2671	2.04		
		2016	2573	2622	-1.91		
B <sub>3</sub>	37	2013	2432	2616	-7.60	160.44	96.4
		2014	2740	2761	-0.76		
		2015	2727	2528	7.27		
		2016	2573	2744	-6.62		

has high adjusted  $R^2$  (91.40 %) considered to be the best fit model. It indicated 91.40 per cent of variation in *kharif* rice yield explained by first principal component which shows maximum amount percent of variation (91.40 %) along with  $T$  (time trend). The model A<sub>5</sub> has been developed at 39<sup>th</sup> SMW (ripening phase of the *kharif* rice) *i.e.* six weeks before the harvest.

**Multiple linear regression models (MLR)**

The model was developed by incorporating 15 un-weighted and 15 weighted weather indices. The details of developed model equation are given in Table 4.

The forecast model equations were obtained for each SMW from 35<sup>th</sup> to 40<sup>th</sup> SMW to get early forecast. It was observed that



**Table 6:** MLR and PCR model performance for *Kharif* rice yield of Navsari in testing set

SWM	Year	Actual yield	Predicted yield		SWM	Year	Actual yield	Predicted yield	
			MLR	PCR				MLR	PCR
35	2013	2432	2654	2713	38	2013	2432	2596	2715
	2014	2740	2774	2698		2014	2740	2600	2728
	2015	2727	2618	2696		2015	2727	2552	2664
	2016	2573	2776	2578		2016	2573	2759	2613
	RMSE		161.03	143.28		RMSE		167.15	146.42
	MAE		141.96	89.77		MAE		166.32	99.39
	MAPE		5.56	3.61		MAPE		6.38	3.98
36	2013	2432	2460	2741	39	2013	2432	2479	2694
	2014	2740	2580	2717		2014	2740	2721	2733
	2015	2727	2385	2656		2015	2727	2351	2671
	2016	2573	2551	2618		2016	2573	2920	2622
	RMSE		189.73	160.71		RMSE		256.94	<b>136.55</b>
	MAE		<b>138.45</b>	112.23		MAE		197.34	<b>93.65</b>
	MAPE		<b>5.11</b>	4.48		MAPE		7.48	<b>3.75</b>
37	2013	2432	2616	2700	40	2013	2432	2407	2676
	2014	2740	2761	2746		2014	2740	2739	2757
	2015	2727	2528	2674		2015	2727	2300	2696
	2016	2573	2744	2618		2016	2573	2736	2624
	RMSE		<b>160.44</b>	138.50		RMSE		228.54	<b>125.98</b>
	MAE		<b>143.60</b>	92.76		MAE		153.79	<b>85.64</b>
	MAPE		<b>5.56</b>	3.72		MAPE		5.76	<b>3.44</b>

the value of adjusted  $R^2$  in different models vary from 88.90 per cent in model B<sub>1</sub> to 96.40 per cent in model B<sub>3</sub>. The performance of the model was checked using RMSE, MAE and MAPE. The lowest was found RMSE (47.59), MAE (36.48) and MAPE (0.02) in model B<sub>3</sub>(Table 5). The model B<sub>3</sub> has high adjusted  $R^2$  value (96.40 %) was considered to be the best fit model. It indicated 96.40 per cent of variation in *kharif* rice yield explained by weather indices *viz.*  $Z_{351}$  (weighted average interaction between morning relative humidity and total rainfall),  $Z_{31}$  (weighted average morning relative humidity),  $Z_{151}$  (weighted average interaction between maximum temperature and total rainfall) and  $Z_{20}$  (weighted average minimum temperature) along with  $T$  (time trend). The model B<sub>3</sub> has been developed at 37<sup>th</sup> SMW (reproductive phase of the *kharif* rice) *i.e.* eight weeks before the harvest.

#### Comparison of MLR and PCA models

The comparison of developed models of MLR and PCA were done. The best suited models were chosen based on Adj.  $R^2$  along with RMSE, MAE and MAPE. It was found that model-1 (A<sub>3</sub>) and model-2 (B<sub>3</sub>) were best fitted models (Table 3,4 and 5). The further comparison between models (MLR and PCA) is presented in Table 6 and 7.

It was observed from the Table 6 that in testing set the value of RMSE (125.98), MAE (85.64) and MAPE (3.44) found lower in 40<sup>th</sup> SMW followed by 39<sup>th</sup> SMW (136.55, 93.65 and 3.75, respectively) of model-1. Similarly, in model-2 the value of RMSE (160.44) was found lower in 37<sup>th</sup> SMW. The value of MAE (138.45) and MAPE (5.11) found lower in 36<sup>th</sup> SMW followed by 37<sup>th</sup> SMW (143.60 and 5.56,

respectively). The Table 7 showed that forecast error individual model varied from -10.80 to 2.04 in model-1 ( $A_3$ ) and -7.60 to 7.27 in model-2 ( $B_3$ ). Further. It was observed that the forecast error (%) for the year 2013 found higher as compared to the years 2014, 2015 and 2016 for both the models.

The based on above discussion it suggested that the model-2 ( $B_3$ ) was suitable for pre-harvest forecasting of *kharif* rice yield as compared model-1 based on higher value of adjusted  $R^2$  (96.40), lower value of RMSE (47.59), MAE (36.48) and MAPE (0.02) in training set which utilizes maximum amount (85 per cent) of data for analysis. The model-2 ( $B_3$ ) can estimate the *Kharif* rice yield in 37<sup>th</sup> SMW *i.e.*, eight weeks before the harvest which is earlier as compared to model-1( $B_3$ ) in 39<sup>th</sup> SMW *i.e.*, six weeks before the harvest.

## CONCLUSION

The present investigation was undertaken to forecast *kharif* rice yield well in advance. The study concluded that pre-harvest rice yield forecasting with MLR found superior as compared to principal component analysis. The Model-2 found competent to forecast rice crop yield before eight weeks of actual harvest of the crop (in the 37<sup>th</sup> SMW) *i.e.* during reproductive stage of the crop growth period. There is a wide scope for using MLR approaches to develop pre-harvest forecast models. However principal component analysis also provide reasonably reliable and consistent forecast. Therefore, it is important to continue research on these aspects for various other crops also on a continuous basis. This methodology can be applicable in many crops *viz.* rice, pulses, oil seeds, sugarcane etc. to develop pre-harvest forecasting models and these forecasts have significant value in agricultural planning and policy making.

## REFERENCES

- Bal, S.K., Mukherjee, J., Mallick, K. and Hundal, S.S. (2004) Wheat yield forecasting models for Ludhiana district of Punjab state. *J. Agrometeorol.*, 6 (spl.issue): 161-165.
- 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. *J. Agrometeorol.*, 11(2): 162-168.
- Dhekale, B.S., Mahdi, S. and Sawant, P.K. (2014). Forecast models for groundnut using meteorological variables in Kolhapur, Maharashtra. *J. Agrometeorol.*, 16(2): 238-239.
- Draper, N.R. and Smith, H. (1988). Applied Regression Analysis. 3<sup>rd</sup> edn, John Wiley & Sons Inc, pp. 1-15.
- Food and Agriculture Organization (2015). Rice market monitor, report 18(2): 2-6.
- 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. *J. Agrometeorol.*, 15(1): 89-90.
- Mallick, K., Mukherjee, J., Bal, S.K., Bhalla, S.S and Hundal, S.S. (2007). Real time rice yield forecasting over central Punjab region using crop weather regression model. *J. Agrometeorol.*, 9(2): 158-166.
- 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. *J. Agrometeorol.*, 9(2): 167-173.
- Singh, R.S., Patel, C., Yadav, M.K. and Singh, K.K. (2014). Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. *J. Agrometeorol.*, 16(2): 199-202.
- Sisodia, B.V.S., Yadav, R.R., Kumar, S. and M.K. Sharma (2014). Forecasting of pre-harvest crop yield using discriminant function analysis of meteorological parameters. *J. Agrometeorol.*, 16(1): 121-125.
- Varmora, S.L., Dixit, S.K., Patel, J.S. and Bhatt, H.M. (2004). Forecasting of wheat yield on the basis of weather variables. *J. Agrometeorol.*, 6(2): 223-228.