Full Length Research Article

© 2025, Centre for Advanced Research in Agricultural Sciences
Research Journal of Agricultural Sciences
Volume 16; Issue 02 (Mar–Apr 2025); pp 190–198

ISSN: 0976-1675 (P)
ISSN: 2249-4538 (E)

Multiple Linear Regression Modelling for Mustard Yield Forecasting in Haryana

Ajay Kumar[1], Raj Kumar*[2], Satish Manda[3] and Shikha Bhukal[4]

[1-2] Assistant Professor (Agricultural Economics), College of Horticulture, Maharana Pratap Horticultural University, Karnal - 132 001, Haryana, India
[3] Assistant Professor, SCPK RRC, Maharana Pratap Horticultural University, Karnal - 132 001, Haryana, India
[4] Assistant Professor, Department of Extension Education and Communication Management, CCS, Haryana Agricultural University, Hisar - 125 004, Haryana

Received: 12 Feb 2025; Revised accepted: 02 Apr 2025

Abstract

Agriculture nowadays has become highly input and cost-intensive. Under the changed scenario today, forecasting of various aspects relating to agriculture are becoming more essential. However, despite a strong need for reliable and timely forecasts, the current status is far from satisfactory. The present study compares the efficacy of Multiple Linear Regression Models in quantifying the pre-harvest mustard yield in Hisar, Bhiwani, Sirsa, Fatehabad, Mahendragarh, Rewari, Jhajjar and Gurugram districts of Haryana. The objective of this study was to assess the forecast accuracy of the contending models for district-level mustard yield forecasts in Haryana. The fortnightly weather data on rainfall, minimum temperature and maximum temperature over the crop growth period (September-October to February-March) have been utilized from 1980-81 to 2010-11 for the models' building. The weather-yield data from 2011-12 to 2015-16 have been used to check the post-sample validity of the fitted models for mustard yield forecasts in comparison to those obtained from State Department of Agriculture crop yield(s) estimates.

***Key words*:** Mustard, Forecasting, Yield, Multiple Linear regression, Relative deviation

Regression analysis is the most frequently used statistical technique for investigating and modeling the relationship between variables. Building a regression model is an iterative process. Usually, several analyses are required as improvement in the model structure and flaws in the data are discovered [1]. The use and interpretation of multiple linear regression model often depends on the estimates of individual regression coefficients. However, in some situations, the problem of multicollinearity exists when there are near linear dependencies between/among the regressors. Some applications of regression involve regressor and response variables that have a natural sequential order over time and then the need of time series (TS) modeling arises for the analysis of such dependence.

Multiple regression analysis plays an important role in forecasting a variable's unknown value from the known values of one or more variables. It is in common use to forecast crop production and analyze the effect of weather variables on crop yield. Khatri *et al*. [2] used regression analysis for groundnut yield estimation based on rainfall data for developing forecast models in Surat district of Gujarat, India. Jitendra *et al.* [3] have worked on Indian mustard for assessing outbreak of *Lipaphis erysimi* under present and future climate scenarios. Verma *et al.* [4] have applied stratified random sampling for the generation of vegetation indices to zonal level wheat crop area and production estimation in Haryana. For this purpose, Indian Remote Sensing Satellite Digital IRS-ID, LISS-I, LISS-II, and LISS-III sensor data were used. The acreage estimate was based on crop classification applying on digital multi-spectral remote sensing data and a supervised pattern recognition algorithm. With the use of zonal spectral-trend-agrometeorological yield models, the yield forecasts at the district level were improved significantly. Kumar and Bhar [5] used multiple linear regression to forecast Indian mustard production in Hisar district of Haryana. The time of growth was split into four stages and the regression models were developed for each phase. The earliest and latest forecast was done 4-5 weeks before harvesting. Wissmann *et al.* [6] used categorical variables in linear regression model. It exposed the diagnostic tool condition number to regression model with categorical explanatory variables and analyzed how the dummy variables could affect the degree of multicollinearity. Verma *et al*. [7] developed weather-crop yield models using regression, time-series and multivariate approaches under Crop Acreage Production Estimation (CAPE) project for effective use of the fitted models to forecast the yields of different crops at district as well as agro-climatic zone level in Haryana. Adrian [8] attempted model-based approach for forecasting maize and soybean yields. The model incorporated the direct survey estimates as well as external sources of information and produced a measure of statistical efficiency.

Keong and Keng [9] applied stepwise multiple linear regression (MLR) model with monthly oil palm yield as

***Correspondence to:** Raj Kumar, E-mail: rajk.apeco@mhu.ac.in; Tel: +91 9034688638

Citation: Kumar A, Kumar R, Manda S, Bhukal S. 2025. Multiple linear regression modelling for mustard yield forecasting in Haryana. *Res. Jr. Agril. Sci.* 16(2): 190-198.

dependent variable and weather variables in cumulated time-lag period prior to harvest as independent variables. The MLR model displayed an acceptable performance with multiple coefficients of determination (R^2) explaining 68% yield, with palm age and available water holding capacity as predictor variables. Rao *et al.* [10] worked on Indian mustard at six northern locations viz., New Delhi, Udaipur, Mohanpur, Rakh Dhiansar, Palampur and Bharatpur of India for assessing outbreak of *Lipaphis erysimi* and aphid infestation under present and future climate scenarios respectively. Shabnam *et al*. [11] used time series data of temperature and yield to assess the impact of climate change on mustard crop yield in Haryana. It was derived that an increase of 1 °C in the temperature during the crop growth period will increase the mustard yield in the state by around 140 kg ha^{-1}. Verma *et al*. [12] developed zonal-yield models for district level cotton yield forecasts in Haryana, by incorporating a dummy regressor in the form of crop condition term along with the weather variables to improve the predictive accuracy of the agromet cotton yield models. Bhatt *et al*. [13] developed three statistical models using regression analysis for forecasting the yield of mustard crop over central Punjab using different weekly weather variables (1974-2011) viz., minimum and maximum temperature, morning and evening relative humidity, sunshine hours, rainfall and number of rainy days. Three years meteorological data (2012-14) were used to check the models' validation.

Chaudhari *et al*. [14] applied stepwise regression analysis to examine the three types of models using meteorological weather variables and mustard crop yield in Gandhinagar district of Gujarat and suggested a suitable pre-harvest forecast model which provided the earliest forecast with high coefficient of determination and minimum percentage deviation from the observed mustard yield. Ravita and Verma [15] applied multivariate statistical technique to achieve rapeseed-mustard yield estimation on district-level in Haryana State. The weather-yield models having crop condition term as dummy regressor had the desired forecast accuracy by showing 5-10 percent mean deviations in most of the mustard-growing districts. Niedbała *et al*. [16] developed a suitable model based on meteorological data (air temperature and precipitation) and information about mineral fertilization (2008-2017) for prediction and simulation of winter rapeseed yield using multiple regression method. Sharma *et al.* [17] developed weather- yield forecast models for soybean and wheat crops in eight districts of the Malwa agro-climatic zone during 2017-18, using stepwise regression method and the accuracy of the models were tested with coefficient of determination (R^2). Daka *et al*. [18] forecasted mustard crop productivity for Banaskantha district of Gujarat state using 32 years (1982-83 to 2013-14) weather data (weekly average of maximum and minimum temperature, morning and evening relative humidity, bright sunshine hours/day and rainfall). They applied step-wise regression procedure including time trend as an independent variable and observed the positive and significant effect of rainfall while the effect of time trend was not significant for mustard yield. They provided a suitable pre-harvest model predicting mustard yield about 4 weeks in advance of the actual harvest. Das and Kumar [19] investigated the impact of climatic factors on crop productivity of wheat in Banaskantha district of Gujarat using weekly weather data (1982-83 to 2011-12) of temperature, relative humidity, bright sunshine hours and rainfall using the stepwise regression equation. The results concluded that the model has significant R^2 (0.67) with the standard error of estimation. Furthermore, the selected model was also tested with the next three years.

MATERIALS AND METHODS

Agriculture is the mainstay of more than 65 per cent population in Haryana with the second largest contribution to the food bowl of the country. The Haryana state comprised of 22 districts is situated between 74° 25' to 77° 38' E longitude and 27° 40' to 30° 55' N latitude. The total geographical area of the state is 44212 sq. km. The present study dealt with modeling the yield of mustard crop in Hisar, Bhiwani, Sirsa, Fatehabad, Mahendragarh, Rewari, Jhajjar and Gurugram districts of Haryana.

The state Department of Agriculture and Farmers Welfare mustard yield data compiled for the period 1980-81 to 2015-16 of Hisar, Bhiwani, Sirsa, Mahendragarh and Gurugram, 1989-90 to 2015-16 of Rewari and 1997-98 to 2015-16 of Jhajjar and Fatehabad districts were utilized for the purpose. The mustard yield data from 1980-81 to 2010-11 along with weather data (collected from IMD, Delhi and different meteorological stations in Haryana) of the same period were used for the training set. The weather-yield data of post-sample period, i.e., 2011-12 to 2015-16 have been used for validity testing of the developed mustard yield forecast models.

Computation of weather parameters

Rainfall and temperature are the main weather parameters which affect the crop growth through different physiological processes and rate of phenological development. The fortnightly weather data were prepared from daily data as shown below:

$$\text{Average Maximum Temperature (Tmx)} = \frac{\sum_{i=1}^{15} Tmx_i}{15}$$

$$\text{Average Minimum Temperature (Tmn)} = \frac{\sum_{j=1}^{15} Tmn_j}{15}$$

$$\text{Accumulated Rainfall (Arf)} = \sum_{k=1}^{15} Arf_k$$

Where;
Tmx_i- i^{th} day maximum temperature
Tmn_j- j^{th} day minimum temperature
Arf_k- k^{th} day rainfall
i, j, k - 1, 2, ...,15 (daily weather data)

Distribution of weather variables

The distribution of weather variables spreads over 10 fortnights for maximum and minimum temperature over the crop growth period (1st fortnight of October to 2nd fortnight of February) and 12 fortnights for accumulated rainfall during 1st fortnight of September to 2nd fortnight of February. The fortnightly weather data base was prepared from daily weather data. The average weather value over 1^{st} to 15^{th} October gave 1^{st} fortnight for maximum and minimum temperature while 1^{st} to 15^{th} September provided 1^{st} fortnight for accumulated rainfall and proceeding in the similar manner; 32 weather parameters were obtained over the total growth period viz., Tmx_1, Tmx_2,..., Tmx_{10}, Tmn_1, Tmn_2,…., Tmn_{10} and Arf_1, Arf_2,…, Arf_{12}.

Statistical methodology

Crop productivity is affected by technological change and weather variability. It can be assumed that the technological factors will increase crop yield smoothly through time and therefore, year or some other parameter of time can be used to study the overall effect of technology on yield. The linear time-trend based model(s) i.e.

$$T_r = a + bt$$

Where;

T_r = Yield (q/ha), a = Intercept, b = Slope and t= Year were fitted on the basis of crop yield data for all the districts. Predictions T_r based on this model yielded a predictor variable that is denoted as 'trend yield' in this study. Trend yield may be considered as an indication of technological advancement, qualitative/quantitative changes in fertilizer/insecticide/ pesticide/weedicide use and increased use of high yielding varieties over time.

Regression analysis

The purpose of standard regression model is to explore an association between dependent and independent variables, to identify the impact of these covariates on the response and further helps in predicting the future values of the dependent variable. The regression analysis has different forms according to the nature of relationship between dependent and independent variables. The main two types are linear regression and non-linear regression analyses. In linear regression, the relationship is modelled by the functions, which are linear combinations of variables. In non-linear regression, the relationship is modelled by the functions of non-linear combinations of variables.

Simple linear regression

Simple regression assesses the relationship between one dependent variable and only one independent variable. It simply has an equation of the form:

$$Y = a + b\,X + e$$

Where;
Y is the value of dependent variable to be predicted
a is the constant term that equals the value of Y when X = 0
b is the coefficient of X, which is the slope of regression line that explains how much Y changes for one unit change in X variable
X is the value of independent variable
e is the error term with assumption NID (0, σ^2)

Multiple linear regression

Weather variability both within and between seasons is major uncontrollable source of variability in yield. Weather variables affect the crop differently during different stages of development. This increases the number of variables in the model and in turn, a large number of parameters are to be evaluated from the time series data for precise estimation. Thus, a technique based on relatively smaller number of manageable parameters and at the same time, taking care of entire weather distribution is always preferred to solve the problem.

For quantitative forecasting, regression models have been fitted by taking weather variables and trend yield as regressors and crop yield as the regressand.

The multiple linear regression model considered may be expressed as follows:

$$Y = a + cT_r + \sum_{i=1}^{10} b_i\, Tmx_i + \sum_{j=1}^{10} b_j\, Tmn_j + \sum_{k=1}^{12} b_k\, Arf_k + e$$

Where;
Y - Mustard yield (q/ha)
T_r - Trend yield (q/ha)
A- Overall mean effect
c - Regression coefficient of trend yield
b_i, b_j, b_k - Regression coefficients of weather variables
(i, j, k - weather fortnights, i.e. 1,2,3…10/11,12)
e - Error term

The multiple linear regression analysis was carried out for the development of weather-yield models. The weather-yield models have been fitted to relate crop yield to average maximum temperature, average minimum temperature calculated for 10 fortnights covering the crop growth period i.e. 1st fortnight of October to 2nd fortnight of February, and accumulated rainfall obtained for 12 fortnights over the period 1st fortnight of September to 2nd fortnight of February.

The best subsets of weather variables are selected using the stepwise regression method [1] in which all the variables were first included in the model and eliminated one at a time with decisions at any particular step conditioned by the results of previous step. The best supported weather variables in the model are retained if they had the highest adjusted R^2 and the lowest standard error at a given stage. Once a regression model has been constructed, it may be important to confirm the goodness of fit of the model and the statistical significance of estimated parameters. Commonly used checks of goodness of fit include R^2, analysis of the pattern of residuals and hypothesis testing. Statistical significance is checked by an F-test of the overall fit, followed by t-test of individual parameters. The weather-yield models based on regression analysis were compared on the basis of per cent relative deviations and root mean square errors to obtain pre-harvest mustard yield forecasts.

To further enhance the predictive performance, the weather-yield models were again fitted by taking crop condition term (CCT) as categorical/dummy variable along with weather variables as regressors and DOA mustard yield as regressand. The CCT being an indicator variable is generated by splitting the trend yield data into different non-overlapping classes. Also, the reason behind this step was presence of multicollinearity among weather variables and thus analysed to see whether dummy variables could affect the degree of multicollinearity [6].

Comparison and validation of the fitted models

Reliability in numerical models are quantified by verification and validation. The process of determining the degree to which a model is an accurate representation of the real world, is known as validation. The forecasting performance of the developed model is compared in relation to the state Department of Agriculture yield estimates using different statistical measures as mentioned below:

Percent relative deviation (RD%)

This measure the deviation (in percentage) of forecast yield from the observed yield and is measured as:

$$\text{Percent Relative Deviation} = \{(\text{observed yield} - \text{forecasted yield})/\text{observed yield}\} \times 100$$

Coefficient of determination (R^2)

It is in general use for checking the adequacy of the model. R^2 is given by the following formula:

$$R^2 = 1 - \frac{SS_{res}}{SS_t}$$

Where;
SS_{res} and SS_t are the residual sum of square and the total sum of square respectively.

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

Where;
$ss_{res}/(n\text{-}p)$ is the residual mean square and $ss_t/(n\text{-}1)$ is the total mean square.

Root mean square error (RMSE)

It is used as a measure of comparing alternatives models and its formula is given as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - E_i)^2}$$

Where;
O_i and E_i are the observed and forecast values and
n is the numbers of years for which forecasting has been done.

RESULTS AND DISCUSSION

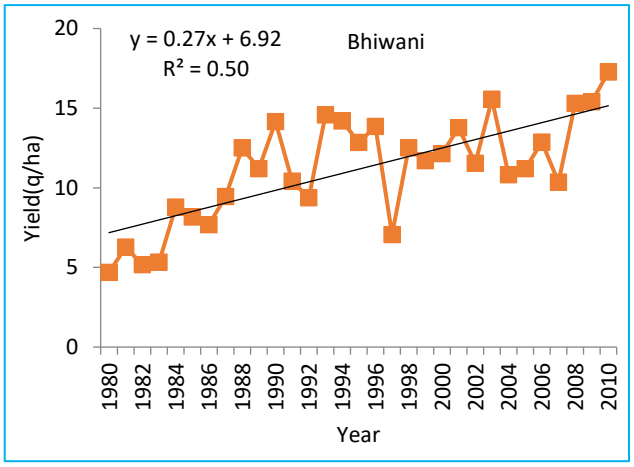
Crop productivity is influenced by technological changes and weather variability. Technological variables can be expected to impact crop yields significantly over a long period of time, so the overall effect of technology on yield can be calculated by using year or some other time parameters as explanatory variable.

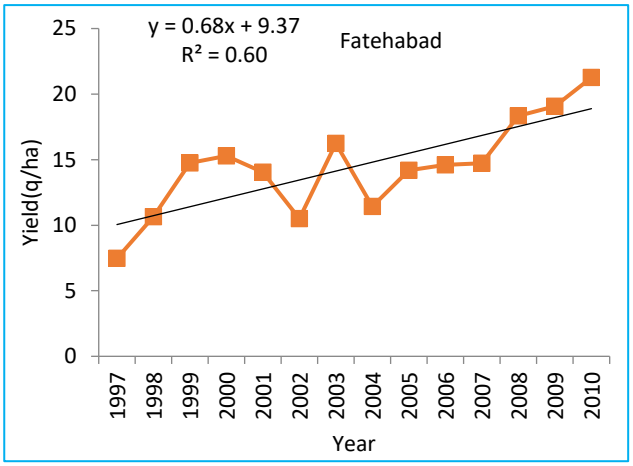
Time-trend analysis

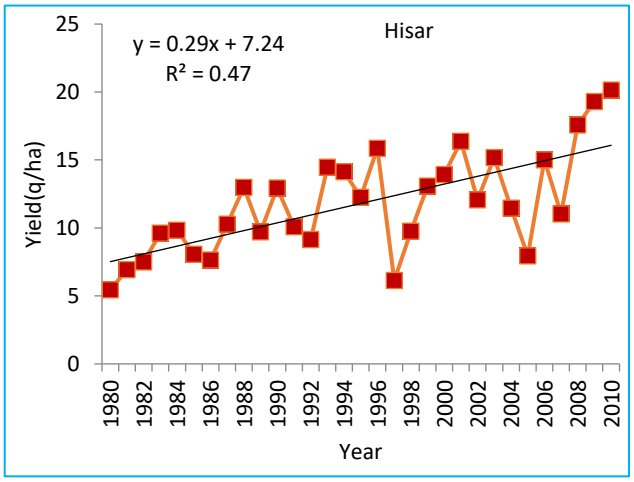
Time-trend analysis often reflects an underlying pattern/behaviour in a time series which would otherwise be partly or nearly completely hidden by noise. The following time versus yield graphs (Figure 1) are showing overall increasing trend for mustard crop in all the districts. The linear time-trend based model(s), i.e.,

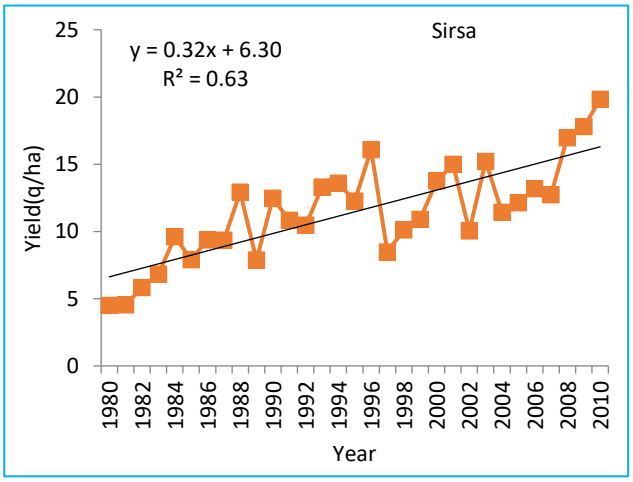
$$T_r = a + bt$$

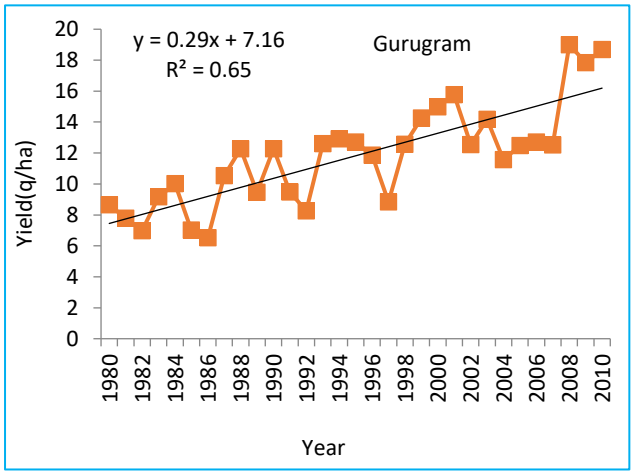
Where;
T_r = Yield (q/ha), a = Intercept, b= Slope and t = Year, have been fitted and predictions T_r, based on this model, yielded a predictor variable i.e. 'trend yield'

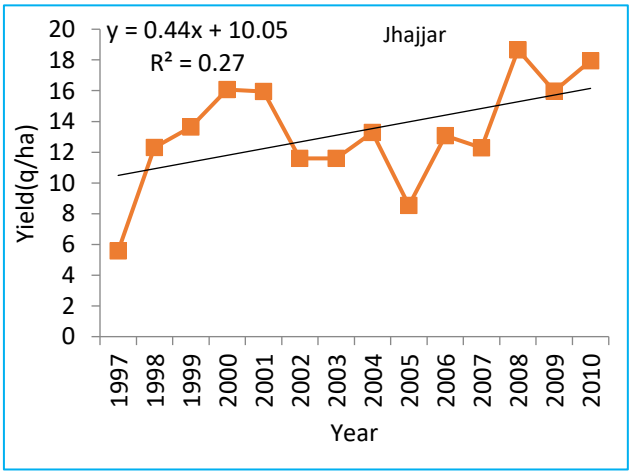












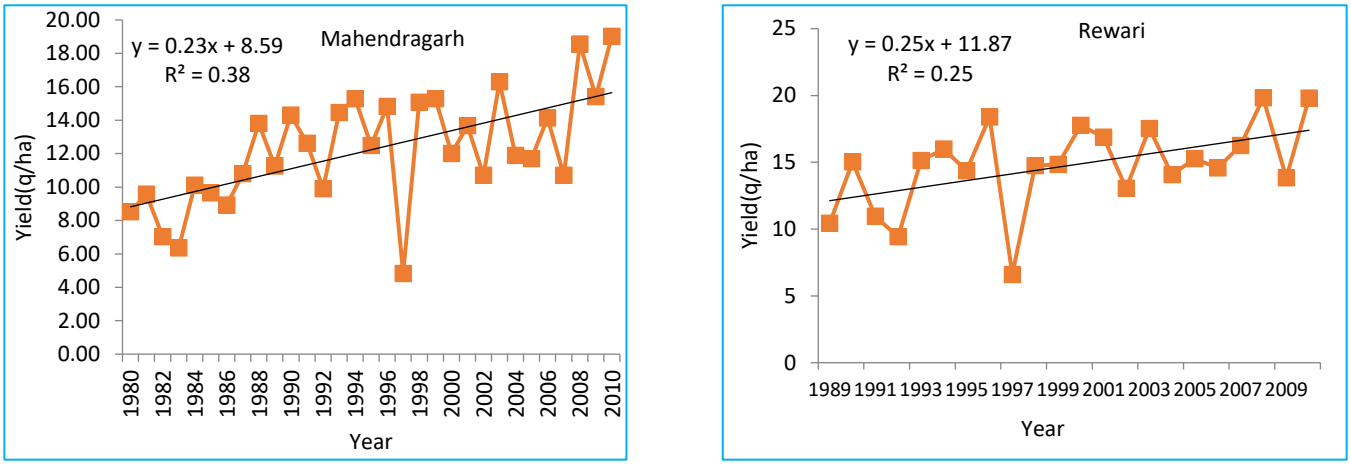


Fig 1 Time versus yield graph(s) of mustard crop for all the districts

Regression based weather-yield forecast models

The DOA mustard yield data were used by considering time (year) as an independent variable and regressed against yield to get the trend equation of the form as have been shown in (Fig 1). The weather-yield models were developed by stepwise regression method using trend-based yield and weather parameters (rainfall, minimum temperature and maximum temperature) computed over different fortnights of crop growth period as regressors. The best supported weather variables in the model were retained if they had the highest adjusted R^2 and lowest standard error of estimate (Table 1). The yield forecasts based on finally selected weather-yield models had higher percent relative deviations from real-time mustard yield(s) in most of the districts, sometimes even too high than to be acceptable. Consequent upon, the selected models were seemed unsuitable for operational yield forecasting purpose. Adding linear time-trend to the model, along with the selected weather variables couldn't significantly improve the forecast accuracy of the mustard yield.

Further, an effort was made to improve the predictive accuracy of weather-yield models by identifying and adding additional covariate to the model, along with the selected weather variables. In particular, CCT/dummy variable was incorporated in weather-yield models by repeating the stepwise regression analysis and that substantially improved the predictive accuracy of the models. The CCT is a categorical covariate obtained by dividing the trend predicted yield series into three non-overlapping classes reflecting high, normal and low yield. Incorporating CCT as dummy variable along with weather parameters enhanced the predictive accuracy of weather-yield models. The suitable models along with Adj. R^2 and standard error are shown in (Table 1).

Table 1 Regression based weather-yield forecast models for all the districts

Types	Fitted models	Adj. R^2	SE
Model 1	Y_{est} = -18.60 +1.02 T_r + 0.01 Arf_2 – 0.04 Arf_3 + 0.49 Tmx_2 + 0.31 Tmn_8 + 0.20 Tmx_9 – 0.43 Tmn_{10}	0.71	1.90
Model 1*	Y_{est} = -19.45 +1.02 T_r + 0.35 Tmx_2 + 0.29 Tmx_5 + 0.01 Arf_2 – 0.04 Arf_3 + 0.14 Tmx_9 – 0.25 Tmn_{10}	0.71	1.93
Model 2	Y_{est} = -16.72 + 2.70 CCT + 0.01 Arf_1 + 0.02 Arf_2 + 0.48 Tmx_3 – 0.41 Tmn_4 + 0.59 Tmx_2	0.62	2.18
Model 2*	Y_{est} = -18.54 + 2.85 CCT + 0.01 Arf_1 + 0.02 Arf_2 – 0.29 Tmn_{10} + 0.55 Tmn_3 + 0.42 Tmx5	0.61	2.20
Model 3	Y_{est} = -9.54 – 5.71 D_1 - 2.62 D_2 + 0.02 Arf_2 + 0.54 Tmx_3 + 0.41 Tmx_5 – 0.29 Tmn_4 + 0.01 Arf_1	0.61	2.20
Model 3*	Y_{est} = -1.33 - 5.13 D_1 - 2.21 D_2 + 0.01 Arf_2 + 0.77 Tmx_3 – 0.29 Tmx_4 – 0.04 Arf_{12} – 0.01 Arf_1	0.58	2.30

Where;
Y_{est} - Model predicted yield (q/ha)
T_r - Trend yield (q/ha)
Tmx - Maximum temperature
Tmn - Minimum temperature
Arf- Accumulated rain fall
CCT- Crop condition term
D - Dummy variable
R^2 - Coefficient of Determination
SE - Standard error

Model 1, 1* - Weather parameters and trend yield as regressors
Model 2, 2* - Weather parameters and CCT as regressors
Model 3, 3* - Weather parameters and dummy variables as regressors

The forecast performance(s) of best suited weather-yield models have been observed in terms of per cent relative deviations of yield forecasts from the real time yield(s) and root mean square errors (RMSEs). Model predicted yield(s) of all the districts along with observed yield(s) and per cent relative deviations are given in (Table 2).

Table 2 Post sample mustard yield forecasts based on finally selected models for all the districts

District /Forecast years Bhiwani	Observed yield (q/ha)	Model 1		Model 2		Model 3	
		Fitted yield (q/ha)	RD (%)	Fitted yield (q/ha)	RD (%)	Fitted yield (q/ha)	RD (%)
2011-12	12.00	16.66	-38.86	16.46	-37.19	17.40	-44.96

2012-13	16.40	13.95	14.92	14.35	12.47	15.30	6.69
2013-14	15.16	18.26	-20.46	15.73	-3.76	15.35	-1.27
2014-15	13.98	13.24	5.25	15.10	-8.00	15.89	-13.65
2015-16	14.61	17.26	-18.20	14.96	-2.42	14.12	3.33
Av. Abs. percent dev.			19.54		12.77		13.98

Fatehabad

2011-12	18.66	18.58	0.38	16.46	11.78	17.40	6.78
2012-13	15.99	15.92	0.42	14.35	10.23	15.30	4.29
2013-14	18.53	20.27	-9.41	15.72	15.11	15.35	17.15
2014-15	15.37	15.30	0.44	15.09	1.77	15.89	-3.37
2015-16	13.55	19.37	-42.95	14.96	-10.44	14.12	-4.23
Av. Abs. percent dev.			10.72		9.86		7.17

Hisar

2011-12	17.07	17.37	-1.76	16.46	3.56	17.40	-1.90
2012-13	16.78	14.66	12.62	14.35	14.45	15.30	8.80
2013-14	16.26	18.97	-16.68	15.73	3.26	15.35	5.59
2014-15	14.17	13.96	1.48	15.10	-6.55	15.89	-12.13
2015-16	18.16	17.98	0.97	14.96	17.60	14.12	22.23
Av. Abs. percent dev.			6.70		9.08		10.13

Sirsa

2011-12	16.78	17.64	-5.10	16.46	1.89	17.40	-3.66
2012-13	16.47	14.96	9.14	14.35	12.84	15.30	7.08
2013-14	17.37	19.31	-11.17	15.73	9.44	15.35	11.62
2014-15	15.00	14.33	4.44	15.10	-0.65	15.89	-5.92
2015-16	17.09	18.39	-7.64	14.96	12.44	14.12	17.36
Av. Abs. percent dev.			7.50		7.45		9.13

Gurugram

2011-12	20.25	17.16	15.25	18.59	8.21	19.05	5.91
2012-13	20.26	15.92	21.42	15.80	22.03	16.24	19.86
2013-14	15.94	14.73	7.57	16.28	-2.15	16.40	-2.86
2014-15	12.69	15.55	-22.55	16.28	-28.27	17.07	-34.52
2015-16	16.52	17.80	-7.78	16.32	1.22	16.36	0.95
Av. Abs. percent dev.			14.91		12.38		12.82

Jhajjar

2011-12	15.95	16.93	-6.12	18.58	-16.54	19.05	-19.46
2012-13	16.86	15.68	6.98	15.80	6.30	16.23	3.70
2013-14	15.49	14.50	6.41	16.28	-5.12	16.39	-5.85
2014-15	13.66	15.32	-12.12	16.28	-19.16	17.07	-24.97
2015-16	15.80	17.57	-11.20	16.32	-3.28	16.36	-3.57
Av. Abs. percent dev.			8.56		10.08		11.51

Mahendragarh

2011-12	18.27	17.93	1.88	18.59	-1.74	19.05	-4.29
2012-13	16.99	16.72	1.59	15.80	7.02	16.24	4.43
2013-14	16.99	15.57	8.37	16.28	4.16	16.40	3.50
2014-15	14.99	16.42	-9.54	16.28	-8.59	17.07	-13.88
2015-16	15.73	18.71	-18.94	16.32	-3.74	16.36	-4.03
Av. Abs. percent dev.			8.07		5.05		6.03

Rewari

2011-12	18.40	19.09	-3.77	18.59	-1.02	19.05	-3.55
2012-13	21.57	17.85	17.22	15.80	26.76	16.24	24.73
2013-14	18.64	16.67	10.55	16.28	12.65	16.40	12.04
2014-15	15.41	17.50	-13.53	16.28	-5.63	17.07	-10.77
2015-16	23.34	19.75	15.37	16.32	30.08	16.36	29.89
Av. Abs. percent dev.			12.09		15.23		16.20

Regression diagnostics of the fitted models

Residual Diagnostics are concerned with testing the goodness of fit of a model and suggesting appropriate modifications if required. Thus, residual histogram and normal-probability plots for the alternative models were prepared for examining normality assumptions of the residuals. Histograms show approximate behaviour with slight deviation from normality. The P-P plots also infer the same. Standardized residual plots appear fine. On the whole, the following plots (Fig 2-4) do not exhibit serious violations of the model assumptions [20]. The regression diagnostics of the fitted models reveal no major violations of model assumptions.

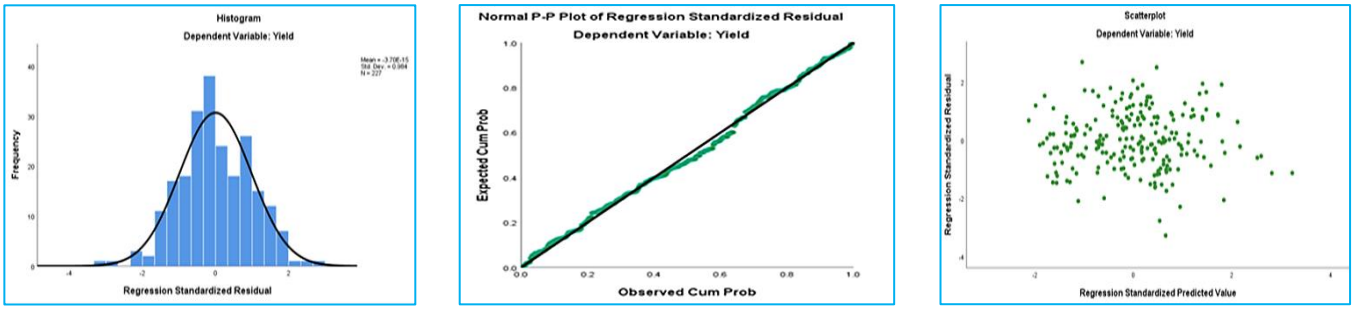


Fig 2 Regression diagnostics of the fitted model (Trend yield + weather variables)

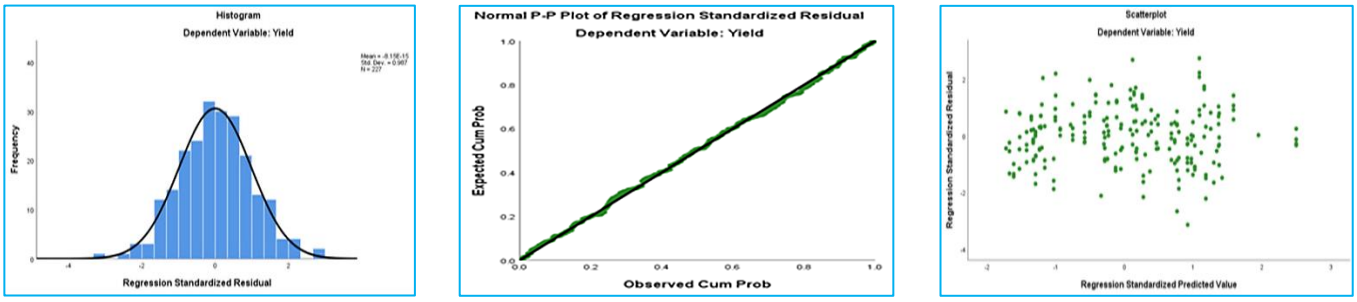


Fig 3 Regression diagnostics of the fitted model (CCT + weather variables)

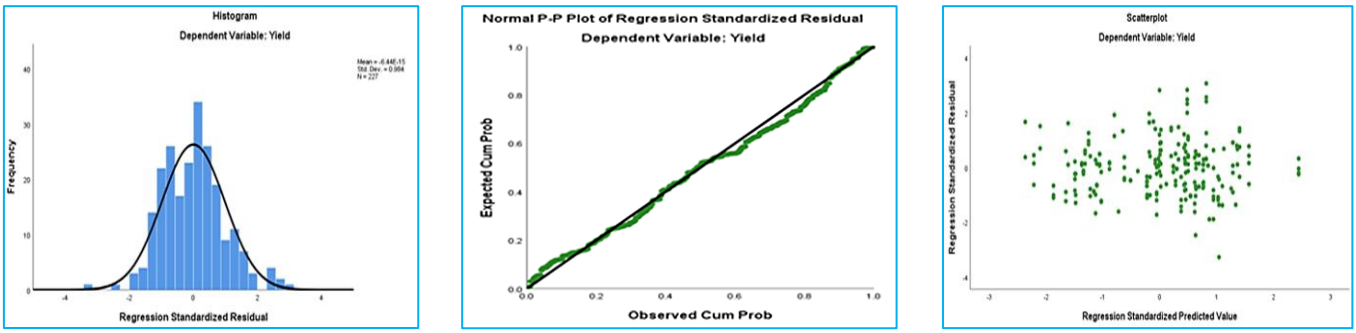


Fig 4 Regression diagnostics of the fitted model (Dummy + weather variables)

Comparison of the fitted models

Mustard yield forecasts for the post-sample years 2011-12, 2012-13, 2013-14, 2014-15 and 2015-16 have been obtained on the basis of multiple linear regression models. The performance(s) of the compending models were examined in terms of average absolute percent deviations and RMSEs (Root mean square errors) of mustard yield forecasts in relation to real-time yield(s) [21]. Comparative view in terms of per cent relative deviations and root mean square errors has been presented in (Table 2-3).

Table 2 Comparative view in terms of average absolute percent deviations of mustard yield forecasts with real time yield(s) for all the districts

District(s)	Average absolute percent deviations		
	Regression based models		
	Weather variables and trend yield	Weather variables and CCT/Dummy variables	
Bhiwani	19.54	12.77	13.98
Fatehabad	10.72	9.86	7.17
Hisar	6.70	9.08	10.13
Sirsa	7.50	7.45	9.13
Gurugram	14.91	12.38	12.82
Jhajjar	8.56	10.08	11.51
Mahendragarh	8.07	5.05	6.03
Rewari	12.09	15.23	16.20

Table 3 Comparative view in terms of root mean square error(s) of mustard yield forecasts with real time yield(s) for all the districts

District(s)	Root mean square error(s)		
	Regression based Models		
	Weather variables and trend yield	Weather variables and CCT/Dummy variables	
Bhiwani	3.00	2.27	2.62
Fatehabad	2.72	1.87	1.60
Hisar	1.55	1.88	2.11
Sirsa	1.34	1.54	1.76
Gurugram	2.82	2.67	2.72
Jhajjar	1.36	1.78	2.13
Mahendragarh	1.62	0.90	1.12
Rewari	2.66	4.22	4.13

In spite of seeming everything fine related to the above fitted models, it has been observed that adding linear time-trend/CCT to the model, along with the selected weather variables; though significantly improved the forecast accuracy of mustard yield. The overall results indicate the preference of using linear time-trend/CCT models for obtaining mustard yield forecasts in the districts under study. Linear time-trend/CCT models performed well in most of the time-regimes and consistently showed the superiority over compending models in capturing lower percent relative deviations pertaining to mustard yield forecasts in Haryana [22].

CONCLUSION

In conclusion, crop productivity is intricately influenced by both technological advancements and weather variability, with each factor playing a critical role in shaping yield outcomes. The time-trend analysis of mustard crops across multiple districts highlighted a consistent upward trend in yields, underscoring the potential long-term impact of technological changes on crop productivity. While the regression-based weather-yield forecast models showed promise, they initially struggled with forecast accuracy, particularly when used in isolation. Incorporating linear time-trend and weather variables significantly enhanced the predictive capability of the models. The addition of Crop Condition Terms further improved accuracy by classifying yield trends into high, normal, and low categories, allowing for more precise yield forecasts. The results from the regression diagnostics confirmed that the models, despite slight deviations from normality, adhered to the underlying assumptions, making them reliable for forecasting purposes. Among the different model configurations, the combination of linear time-trend and CCT with weather variables emerged as the most robust for predicting mustard yields, particularly in Haryana's varied districts. These models consistently outperformed others in minimizing forecast errors and relative deviations, highlighting their superiority in capturing yield dynamics. Thus, for operational yield forecasting, models incorporating time-trend and CCT offer a promising approach, showing both practical and statistical advantages in predicting mustard crop yields with greater accuracy and reliability.

LITERATURE CITED

1. Draper NR, Smith H. 2003. *Applied Regression Analysis*. 3rd Edition. John Wiley & Sons, New York.
2. Khatri TJ, Patel RM, Mistry RM. 1983. Crop weather analysis for pre-harvest forecasting of groundnut yield in Surat at Bular district of Gujarat state. *Gujarat Agriculture University Research Journal* 9(1): 29-32.
3. Jitender K, Malik YP, Singh SV. 1999 Forecasting models for outbreak of *Lipaphiserysimi* on some cultivars of mustard, *Brassica juncea*. *Indian Jr. Entomology* 61(1): 59-64.
4. Verma U, Ruhal DS, Yadav M, Khera AP, Hooda RS, Singh CP, Kalubarme MH, Hooda IS. 2003 Wheat production forecasting using remote sensing and agromet variables in Haryana state. *Photonirvachak Journal of Indian Society Remote Sensing* 31(2): 141-144.
5. Kumar A, Bhar L. 2005 Forecasting model for yield of Indian mustard using weather parameter. *Indian Journal of Agricultural Sciences* 75(10): 688-690.
6. Wissman M, Toutenburg H, Shalabh. 2007. Role of categorical variables in multicollinearity in linear regression model. *Journal of Applied Science* 19(1): 62-67.
7. Verma U, Dabas DS, Hooda RS, Kalubarme MH, Yadav M, Sharma MP. 2011. Remote sensing-based wheat acreage and spectral-trend-agrometeorological yield forecasting: Factor analysis approach. *Statistics and Applications* 9(1/2): 1-13.
8. Adrian DW. 2012. A model-based approach to forecasting corn and soybean yields. Paper presented at the Fourth International Conference on Establishment Surveys, Montreal, Quebec, Canada. pp 11-14.
9. Keong YK, Keng WM. 2012 Statistical modeling of weather-based yield forecasting for young mature oil palm. *APCBEE Procedia* 4: 58-65.
10. Rao B, Rao V, Nair L, Prasad YG, Ramaraj AP, Chattopadhyay C. 2013 Assessing aphid infestation in Indian mustard (*Brassica Juncea* L.) under present and future climate scenarios. *Bangladesh Journal of Agricultural Research* 38(3): 373-387.
11. Shabnam Bansal SK, Dabas DS. 2013 Use of time-series data of temperature and yield to assess the impact of climate change on crop yield using mustard in Haryana as example. *International Research Journal of Social Science* 2(4): 31-33.
12. Verma U, Piepho HP, Ogutu JO, Kalubarme MH, Goyal M. 2014. Development of agromet models for district-level cotton yield forecasts in Haryana state. *International Journal of Agricultural Statistical Science* 10(1): 59-65.
13. Bhatt K, Gill KK, Sandhu SS. 2015. Comparison of different regression models to predict mustard yield in Central Punjab. *Vayu Mandal* 41: 28-38.
14. Chaudhari CM, Thaker MB, Patel NV. 2016. Pre-harvest forecasting model of mustard crop yield based on weather parameters in Gandhi Nagar District of Gujarat. *International Journal of Science and Research* 5(9): 1421-1427.

15. Ravita, Verma U. 2017. Use of crop condition based dummy regressor and weather input for parameter estimation of mustard yield forecast models. *Journal of Applied and Natural Science* 9(3): 1703-1709.
16. Niedbała G, Piekutowska M, Adamski M. 2018. Multiple regression analysis model to predict and simulate winter rapeseed yield. *Journal of Research and Applied Agricultural Engineering* 63(4): 139-144.
17. Sharma SK, Bhagat DV, Ranjeet, Dubey P, Khapedia HL, Mirdha IS, Sikarwar RS. 2018. Soybean and wheat crop yield forecasting based on statistical model in Malwa agroclimatic zone. *International Journal of Chemical Studies* 6(4): 1070-1073.
18. Daka SR, Chaudhary GK, Marviya PB. 2019. Pre-harvest forecasting of mustard yield on the basis of weather variables in Banaskantha district of Gujarat. *International Journal of Agricultural Science* 8(11): 8325-8328.
19. Das S, Kumar M. 2019. Development of yield forecasting model for wheat by using weather variables. *Research Journal of Agricultural Sciences* 10(3): 534-536.
20. Feng C, Li L, Sadeghpour A. 2020. A comparison of residual diagnosis tools for diagnosing regression models for count data. *BMC Medical Research Methodology* 20: 175.
21. Kakati N, Deka RL, Das P. 2022. Forecasting yield of rapeseed and mustard using multiple linear regression and ANN techniques in the Brahmaputra valley of Assam, North East India. *Theoretical and Applied Climatology* 150: 1201-1215.
22. Braide J, Leung B. 2016. A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *Journal of Biogeography* 44(6): 1344-1361.