Environment and Ecology 40 (3D) : 1880—1885, July—September 2022
ISSN 0970-0420

Prediction of Wheat Yield using Ordinal Logistic Regression Based on Weather Parameters

Ajay Sharma, Joginder Kumar, Narender Kumar, Nitin Sharma, Chetna, Pawan Kumar

Received 18 May 2022, Accepted 11 June 2022, Published on 17 September 2022

ABSTRACT

In the present study, an attempt has made to develop suitable statistical models for forecasting wheat yield of Haryana using ordinal logistic regression (OLR) based on meteorological parameters (fortnightly). For this, time series data of wheat yield for 39 years (1980 to 2019) have been obtained from Statistical Abstract of Haryana. The data from 1980 to 2016 has been used for the development of the model and rest subsequent three years i.e. from 2017 to 2019 data has been used to check the accuracy/validation of the fitted models. The performance of the developed models for different fortnights has also been compared using predicted error sum of squares (PRESS), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute deviation (MAD). The findings of the present study revealed that OLR method based on various comparative measures. From the results of the study, it has also been observed that appropriate time for forecasting the wheat yield is one month before harvesting the wheat crop. Findings of the study may also be helpful in advance planning, formulation and implementation of policies related to food procurement, import-export decisions, price structure and distributions.

Keywords Forecast model, Meteorological parameters, OLR, Wheat yield.

Ajay Sharma*[1], Joginder Kumar[2], Chetna[3]
[1]PhD Research Scholar, [2]Assistant Scientist, [3]PhD Research Scholar
Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar, India

Narender Kumar
District Extension Specialist
Department of Soil and Water Conservation, CCS Haryana Agricultural University, Hisar, India

Nitin Sharma
PhD Research Scholar
Department of Agriculture Economics, CCS Haryana Agricultural University, Hisar, India

Pawan Kumar
PhD Research Scholar
Department of Horticulture, CCS Haryana Agricultural University, Hisar, India

Email: ajaystatician03@hau.ac.in
*Corresponding author

INTRODUCTION

Wheat is one of the most important food grain crops in India as it forms a major constituent of the staple food of a world population. *kharif* and *rabi* are two major crop seasons in Haryana. The major *rabi* crops are wheat and mustard (Source: www.mapsofindia.com/India agriculture). In India, wheat is grown on an area of 30.37 (approx) million hectares which produces 90.78 million tons of wheat with a productivity of 2.99 tons per hectare. India is the second largest producer among wheat growing countries of the World. Haryana is the third largest producer of wheat in India and it contributes more than 11.63 million tonnes wheat grown in 2.5 mha. Haryana and Punjab are known as the "*Grain Bowl*" of India (Anonymous 2015).

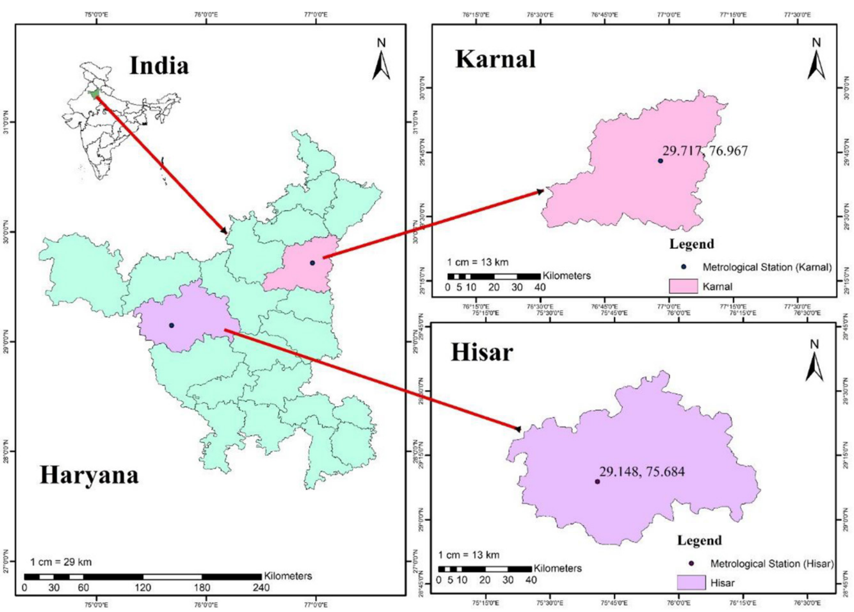
Further, weather and climatic variability has direct impact on yield of wheat crop by affecting soil quality, water availability and leads vulnerability of crops to pests and diseases. The warming rate for minimum temperature is 0.198ºC/decade and cooling rate for maximum temperature is 0.046ºC. The mean temperature is rising with rate of 0.075ºC/decade. The average annual rainfall is increasing with rate of 22.5 mm/decade in Haryana. Thus, forecasting of the crop production is required when crop is still standing in the field. For this, various researchers have been tried to develop weather based forecasting models using different statistical techniques such as stepwise multiple regression and ordinal logistic regression. OLR techniques can be used to handle either two or multiple (three or more) groups.

Various researchers have been made efforts to develop statistical models based on weather variables for pre-harvest forecasting of crop yield. Bhowmik (2009) developed logistic regression models for classification in agriculture using data pertaining to agricultural ergonomics. Agrawal *et al*. (2012) proposed forecasting models for wheat yield using discriminant function analysis by considering weekly data on weather variables. Wan *et al*. (2016) compared population means and variances under a semi parametric density ratio model using logistic regression technique. Vandita *et al*. (2016) applied logistic regression for forecasting of crop yield. Gurmeet (2019) studied the pre harvest forecast models of rice yield based on weather variables for eastern agro-climatic zone of Haryana using discriminant function analysis. Garde *et al*. (2020) studied pre-harvest forecast of *kharif* rice yield using weather parameters for strategic decision making in agriculture.

Keeping in mind that the early forecast of wheat crop yield would assist farmers to formulate crop trends, agricultural practices and their management, the present study was designed to develop different statistical models using weather variables to forecast wheat yield before harvesting.

MATERIALS AND METHODS

Time series data of wheat yield for north-central zone of Haryana have obtained from statistical abstract of Haryana. Daily weather data for the same period have been collected from the CSSRI, Karnal through Department of Agricultural Meteorology, CCS Haryana Agricultural University, Hisar, Haryana. The daily weather data have been converted into twelve



fortnights for wheat period (planting to harvesting) according to standard meteorological week (SMW). The data for five weather variables namely maximum temperature (° C), minimum temperature (° C), average relative humidity (%), sunshine hours (Hours) and accumulated rainfall (mm) have also been recorded from the same location.

At the outset, linear regression has been fitted between wheat yield (dependent variable) and year (independent variable) using data from 1980-81 to 2015-16. The equation hence obtained has been utilized for attainment of residuals. These obtained residuals have been used to categorize the crop yield into two groups: Residuals with negative value have been taken as zero and residuals with positive value have been taken as one. In case of three groups, residuals have been arranged in ascending order and crop yield is categorized into three group viz., adverse (0), normal (1) and congenial (2). OLR techniques have used for two/three groups to obtain probabilities using weather variables. These probabilities along with years were used as regressors for the development of crop yield forecast models using stepwise linear regression procedure. Performance and validation of fitted models have been checked on the basis of various performance measures like PRESS, RMSE, MAPE and MAD.

In the first step i.e. 22nd fortnight (14th Oct -28th Oct.), the weather variables along with years have been used to calculate probabilities using OLR for wheat yield. In the second step, the weather variables of 23rd fortnight (29th Oct – 12th Nov.), probabilities obtained in the first step and years have been used to calculate probabilities using OLR. These steps have been done again in 24th fortnight and so on up to final fortnight. Finally, the forecast models were fitted at various fortnights, starting from 5th fortnight (2nd March – 16th March) using stepwise linear regression by taking probabilities along with years as regressors.

(a) Model 1: Ordinal logistic regression

The linear relationship between the dependent variable Y and the independent variable X is given below

$$Y = \alpha + \beta X + \varepsilon$$

Where,

α is the intercept, β is the regression coefficient and $\varepsilon \sim N(0, \sigma 2)$

Let P_1 denotes probability that Y= 1 for given X, the model for the probability would be as follows:

$$P_1 = P(Y=1|x) = P(Y=1)$$

Thus, $P(Y=0) = 1-P_1$

Under the assumption that $E(\varepsilon) = 0$, it follows that

$$E(Y \mid x) = \hat{\alpha} + \hat{\beta} X = P_1$$

Since the error is dichotomous (discrete) so normality assumption is violated further variance of ε is not constant but depends on x through its effect on P^1 that is

$$V(\varepsilon) = P_1(1-P_1)$$

The logistic regression function is written as following:

$$P_1 = \frac{\exp(\alpha + \beta x)}{1+\exp(\alpha + \beta x)} = \frac{1}{1+\exp\{-(\alpha + \beta x)\}}$$

For this model odds of making response are following

$$\frac{P_1}{1-P_1} = \exp(\alpha + \beta x)$$

Taking log on both sides we get $g(x) = \log \frac{P_1}{1-P_1} = \alpha + \beta x$. This transformation is called as logit transformation.

The fitted model is given by

$$Yield = a + b_1 P_1 + b_2 T + \varepsilon$$

Where, a is the intercept, b_1 is the regression coefficient, P_1 is probability of Y=1,T is year and ε is error $\sim N(0, \sigma 2)$

Case II: Three groups

(b) Model 2: Ordinal logistic regression

If the response variable Y is having three categories

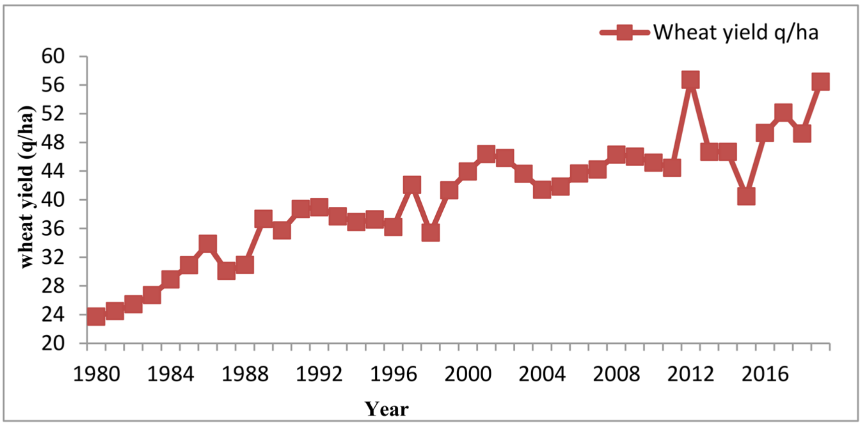


Fig. 1. Trend analysis of wheat yield.

(0, 1, 2) then the ordered multiple response models assume the relationship

$$\text{Logit}\,[P\,(Y \leq j\text{-}1|x)] = \alpha_1+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p, j=1, 2$$

The ordinal logistic regression is given below:

$$P_0 = \frac{\exp\,(\alpha_1+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p)}{1+\exp\,(\alpha_1+\beta_1 X_1+\beta_2 X+\ldots+\beta_p X_p}$$

$$P_0 + P_1 = \frac{\exp\,(\alpha_2+\beta_1 X+\beta_2 X_2+\ldots+\beta_p X_p)}{1+\exp\,(\alpha_2+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}$$

and $P_0+ P_1+ P_2 = 1$

Where,

$P_0 = P(Y = 0)$, $P_1 = P\,(Y=1)$ and $P_2 = P(Y=2)$

The fitted model is given by

$$\text{Yield} = a+ b_1P_1+b_2P_2+b_3T+\varepsilon$$

Where P_1 and P_2 are probabilities of Y = 1 and Y = 2 respectively. Other, symbols are same as defined earlier.

RESULTS AND DISCUSSION

In accordance with the study, the analysis has been carried out to develop wheat yield forecast models based on fortnightly weather parameters of Karnal, Haryana using logistic regression.

Trend analysis for wheat crop

Average wheat yield data of Karnal district was used

Table 1. Wheat yield forecast models using OLR. Note: Figures in brackets represent standard errors. ** Significant at 1% level of significance. * Significant at 5% level of significance.

		Fortnight	Forecast regression equation	p- value
		5th	Yield = -1374.298+0.710**T+4.483**P_1 (0.026) (0.573)	< 0.01
Two groups	Model-1	6th	Yield = -1364.690+0.702**T+5.022**P_1 (0.027) (0.548)	< 0.01
		7th	Yield = -1335.041+0.684**T+5.167**P_1 (0.034) (0.601)	< 0.01
		5th	Yield = -1379.420+0.713**T +1.465**P_1 +5.933**P_2 (0.022) (0.768) (0.654)	< 0.01
Three groups	Model-6	6th	Yield = -1369.088+0.703**T +2.614**P_1 +6.724**P_2 (0.023) (0.635) (0.601)	< 0.01
		7th	Yield = -1413.566+0.726**T +2.855**P_1 +7.650**P_2 (0.032) (0.746) (0.698)	< 0.01

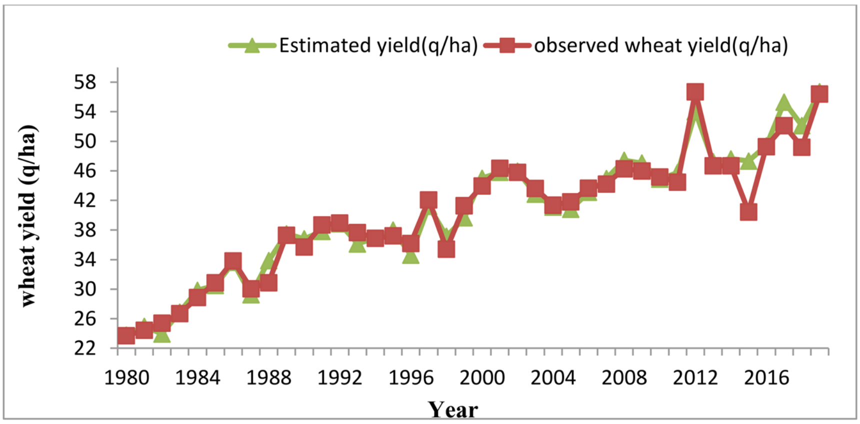


Fig. 2. Observed and estimated wheat yield.

Table 2. Performance measures for different fortnights.

Performance measures			5^{th}	6^{th}	7^{th}
PRESS	Two group	Model – 1	57.64	98.36	110.14
	Three group	Model – 2	61.00	70.07	119.73
RMSE	Two group	Model – 1	4.38	5.72	6.05
	Three group	Model – 2	2.60	2.79	3.64
MAPE	Two groups	Model – 1	7.78	10.28	10.79
	Three groups	Model – 2	7.33	8.50	12.53
MAD	Two groups	Model – 1	3.39	4.44	4.65
	Three groups	Model – 2	2.97	3.62	5.51

for trend analysis by considering year (time) as independent variable (Fig.1).

Forecast models

Regression models have been fitted for wheat yield data from 1980-81 to 2015-16 using stepwise linear regression method by taking wheat yield as the dependent variable and probabilities obtained by OLR along with year as regressors for different fortnights (Table 1).

Comparison of OLR (probabilities) technique

The results obtained for OLR (probabilities) have compared using various comparative measures such as PRESS, MAPE, RMSE and MAD for different fortnights (Table 2).

From Table 2: It was concludes that 5^{th} fortnight has lowest value of all comparative measures (PRESS, RMSE, MAPE and MAD) as compared to other fortnights. If we compare both the groups on probability based then OLR of three groups has given minimum value of all comparative measures. The comparison of observed and estimated wheat yield using selected model as following Fig. 2.

CONCLUSION

The present study was planned to develop pre-harvest forecast models for wheat yield based on weather parameters using stepwise multiple regression and ordinal logistic regression techniques. For this, different forecast models have been fitted based on fortnightly weather data by dividing the wheat yield into two and three groups. On basis of findings of the study, it has been observed that the values of all comparative measures were minimum in 5^{th} fortnight under both the categories (two/three groups). Further, it has also been observed that OLR based on probability in case of three categories has given optimum results as compared to other techniques. Therefore, on the basis of above results, the best time for forecasting the wheat yield is one month before the harvesting. Hence, it is concluded that the wheat yield forecast using OLR based on probability provide reliable forecast in 5th fortnight (2^{nd} March – 16^{th} March).

The findings of the present study will help various stakeholders to take decisions in advance for timely procurement, distribution, price-structure, storage and import-export. The estimation of wheat yield also influence prices in domestic and global markets and processors will take suitable decisions regarding quantum of wheat to be milled in different months. Central and state Governments jointly may also make arrangement for procurement to maintain stock and supply to privileged section of population through Targeted Public Distribution System (TPDS).

REFERENCES

Agrawal R, Chandrahas, Aditiya K (2012) Use of discriminant function analysis for forecasting crop yield. *MAUSAM* 63(3): 455-458.

Anonymous (2015) The increasing integrated rice productivity in South Sulawesi (Peningkatan Produktivitas Padi Terpadu di Sulawesi Selatan). *Department of Agriculture for Food Crops and Horticulture in South Sulawesi, Makassar*.

Bhowmik A (2009) A study on logistic regression modeling for classification in agriculture. *Ind Agric Stat Res Institute* (Thesis).

Gang C, Haiguang W, Zhanhong M (2006) Forecasting wheat stripe rust by discrimination analysis. Department of Plant Pathology, China Agricultural University.

Garde YA, Thorat VS, Pisal RR, Shinde VT (2020) Pre harvest forecasting of *kharif* rice yield using weather parameters for strategic decision making in agriculture. *Int J Environ Climate Change* 10(12): 162-170.

Gurmeet N (2019) Pre harvest forecast models of rice yield based on weather variables for eastern agro-climatic zone of Haryana using discriminant function analysis. *CCS Haryana Agricultural University, Haryana* (Thesis).

Vandita K, Agrawal R, Kumar A (2016) Use of ordinal logistic regression in crop yield forecasting. *MAUSAM* 67(4): 913-918.

Wan S, Xu B, Zhang B (2016) Using logistic regression for semi parametric comparison of population means and variances. *Commun Stat - Theory Methods* 45(9): 2485-2503.