Int.J.Curr.Microbiol.App.Sci (2021) *10*(03): 967-973

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 10 Number 02 (2021)**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2021.1003.123

Statistical Models for Wheat Yield Forecasting using Discriminant Function Analysis of Weather Parameters

Ravi Prakash Gupta*, V. N. Rai, Sarvesh Kumar, K. K Mourya and Annu

Department of Agricultural Statistics, Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, Uttar Pradesh, India

**Corresponding author*

A B S T R A C T

The study was based on, an application of discriminant function analysis of Agroclimatic zone for developing suitable statistical models to forecast wheat yield in Ayodhya district of Eastern Utter Pradesh has been demonstrated. The study utilized the crop yield data and corresponding weekly weather data of last 27 years (1990-91 to 2016-17). The model development was carried out at 44th and 6th SMW (Standard Meteorological Week) for getting forecast well in advance of actual harvesting of the field crop. The discriminant scores obtained from this have been used as regressor variables along with time trend in development of statistical models. In all six procedures using weekly weather data have been proposed. The models developed have been used to forecast the wheat yield for the year 2014-15, 2015-16 and 2016-17 which were not included in the development of the models. It has been found that most of the models provide reliable forecast of the wheat yield about two months before the harvest. However, the model-6 has been found to be the most suitable among all the models developed. It was observed high value of Adj. R2= 0.78 and low value of RMSE= 6.22. The model can be used in different crop for reliable and dependable forecast and these forecasts have significant value in agricultural planning and policy making.

Keywords

Pre harvest forecast, Wheat crop yield, weather indices, and discriminant function

Article Info

Accepted:
10 February 2021
Available Online:
10 March 2021

Introduction

Wheat (Triticum aestivum L.) is a third most important cereal crop after maize and rice, play significance role in food security. It is a type of grasses family cultivated for millions of peoples as a staple food. Wheat is major cereal crop in temperate region being utilized as a human food and livestock feed. In India wheat is the second most cultivated food crop after rice in production as well as in area. The crop has been under cultivation in about 30 million hectares (14% of global area) to produce the all-time highest output of 99.70 million tones of wheat (13.64% of world production with a record average productivity of 3371 kg/ha. It produced 101.20 million tone wheat during crop year 2018-19 which is 1.3% higher than previous year. In India mostly wheat grown in northern region during winter season. It covered 26% areas of total cereals production of country. Uttar Pradesh, Punjab, Rajasthan, Haryana, Madhya Pradesh, Bihar and Gujarat are the major wheat growing states in the country. Uttar Pradesh ranks first in area (36.58%) as well as in production (36.27%). It is the highest wheat growing state of the country and produces 28 million tonnes of wheat. Uttar Pradesh holds the first rank in wheat area (9.75 million ha) as well as production (31.88 million tonnes) during 2017-18 However productivity wise, Uttar Pradesh (3.27 tonnes per ha, hereafter 't/ha') is not the leading state which is even less than the national average (3.37 t/ha) in the same period (Ministry of Agriculture & Farmers' Welfare, 2019).Uttar Pradesh produces maximum wheat in India, accounting for about 33 percent of India's total wheat production (2017-18). The Punjab is the second largest producer of wheat in India, accounting about 18.50 percent of the nation's total wheat production (2017-18). The sowing of winter wheat begins about the first of October and runs through the end of December. Wheat usually begins to head in January, with harvest following in last week of March and April. Crop yield is affected by technological changes and weather variables. The technological changes include use of improved qualities of seed, increased fertilizer applications, better irrigation facilities, improved management practices and pest control, etc. Weather variability both within and between seasons is the second and only uncontrollable source of variability in yield. Weather variables affect the crop differently during different stages of development. Reliable and timely forecasts are essential for agriculture policy making and also for crop production, marketing, storage and transportation decision. This helps in managing risk associated with these activities (Bannayan and Crout, 1999 and Potgieter *et al.,* 2005). Rai and Chandrahas (2000) made use of discriminant function analysis of weather variables to develop statistical models for pre-harvest forecasting of rice-yield in Raipur district of Chhattisgarh. Chowdhury and Das (1993) made a multiple regression model for forecasting the kharif food production of India, using Indian SW monsoon rainfall as one of the parameters of the model. Crop yield in different years are affected due to technological change, system productivity and climatic variability. Individual effects of weather parameters on crop yields were studied by Jain *et al.,* (1980) and yield forecasting models based on weather factors were given by Agrawal *et al.,* (1986) and Munu *et al.,* (2013). Agrawal *et al.,* (2012) have recently developed forecast models for wheat yield in Kanpur district (U.P.) using discriminant functions analysis of weakly data on weather variables.

Since the discriminant function analysis discriminates best between sets of observations from two or more groups and classify the future observations into one of the previously defined groups, an attempt has been made in the present paper to develop suitable statistical models for forecasting of pre harvest wheat yield in Ayodhya district of Uttar Pradesh using discriminant functions analysis of weekly data on weather variables.

Development of statistical forecast models

In order to apply discriminant function analysis for modeling yield using weather variables, crop years under consideration have been divided into three groups, namely adverse, normal and congenial on the basis of crop yield adjusted for trend effect. Data on weather variables in these three groups were used to develop linear discriminant functions and the discriminant scores were obtained for each year. These discriminant scores were used along with year index (trend variable) as regressors and crop yield as regress and in developing the forecast models. In the present study the number of groups is three and number of weather variables is six, therefore only two discriminant functions can be obtained which are sufficient for discriminating a crop years into either of the three groups.

Three groups of crop years, viz. adverse, normal and congenial have been obtained as follows: Let $\bar{y}$ and s be the mean and standard deviation of the adjusted crop yields of n years. The adjusted crop yields less than or equal to $\bar{y}-s$ would form adverse group, the adjusted crop yields between $\bar{y}-s$ and $\bar{y}+s$ would form normal group and adjusted crop yields above or equal to $\bar{y}+s$ would form congenial group. The adjusted crop yields were assigned codes 1, 2 and 3 if they belong to adverse, normal and congenial groups, respectively.

It is also known that weather variables affect the crop differently during different phases of crop development. Its effect depends not only on its magnitude but also on its distribution pattern over the crop season. Therefore, using weekly weather data as such in developing the model process a problem as number of independent variables in the regression model would increase enormously. To solve this problem, following weather indices have been developed using the procedure of Agrawal *et al.*, (1983, 1986).

$$Z_{ij} = \sum_{w=1}^{n} r_{iw}^{j} X_{iw} \Big/ \sum_{w=1}^{n} r_{iw}^{j}$$

$$Z_{ii',j} = \sum_{w=1}^{n} r_{ii'w}^{j} X_{iw} X_{i'w} \Big/ \sum_{w=1}^{n} r_{ii'w}^{j},$$

j=0,1 and i=1,2,…,p. (3.3.2.5)

where Z_{ij} is un-weighted (for j=0) and weighted (for j=1) weather indices for ith weather variable and $Z_{ii',j}$ is the un-weighted (for j=0) and weighted (for j=1) weather indices for interaction between ith and i$'^{th}$ weather variables. X_{iw} is the value of the ith weather variable in wth week, $r_{iw}/r_{ii'w}$ is correlation coefficient of yield adjusted for trend effect with ith weather variable/product of ith and i$'^{th}$ weather variable in wth week, n is the number of weeks considered in developing the indices and p is number of weather variables.

Here, p=7 and n=15, i.e. 15 weeks data from 44th week to 52nd week of a year and 1st week to 6th week of the next year have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions. Here only the first 27 years data from 1990-91 to 2016-17 have been utilized for model fitting and remaining three years were left for the validation of the model.

The growth process of the crop has various phases and weeks within phases. In the development of pre-harvest model based on discriminant function analysis the entire 15 weeks data from 44th SMW to 52nd SMW of a year and 1st SMW to 6th SMW of the next year have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions.

In all 56 indices (28 weighted and 28 un-weighted) consisting of 7 weighted and 21 weighted interaction weather indices and 7 un-weighted and 21 un-weighted interaction weather indices have been constructed.

For quantitative forecasting, linear regression models are fitted by taking the discriminant scores and the trend variable as the regressors and crop yield as the regressand. The following models are considered.

Model-D_1

This model is the 2nd model of Agrawal *et al.*, (2012). This model utilizes the complete data over 15 weeks and also considers relative importance of weather variables in different weeks. Using seven weighted weather indices of seven weather variables as discriminating variables, discriminant function analysis has been carried out and two discriminant functions have been obtained. Two sets of discriminant scores for the years under consideration from these two discriminant functions were obtained. For developing forecast model, these two sets of discriminant scores along with the trend variable were utilized as the regressors and the yield as the regressand. The form of model considered is as follows:

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where α = intercept of the model, β_1 s(i= 1 2 3) the regression coefficients, $ds1$ $ds2$ are the two discriminant scores, T and ε are as defined in model-1.

This model utilizes complete data over all 15 weeks and also considers relative importance of weather variables in different weeks as against model-1 in which equal importance in different weeks was assigned.

Model-D_2

This model is 4th model of Agrawal *et al.*, (2012).Two discriminant functions and there from two sets of discriminant scores have been obtained using the first week data (44th SMW) on seven weather variables. Next, two sets of discriminant scores obtained from first week and the second week (45st SMW) data on seven weather variables data have been used as discriminating variables, so in all there were 9 discriminating variables, and based on these 9 discriminating variables the discriminant function analysis has been done and, therefore, two sets of discriminant scores have been obtained. This process was repeated up to the last week till the time of forecast (6th SMW or 15st week) and finally two sets of discriminant scores have been obtained. Based on these two sets of discriminant scores, the forecasting model taking yield as the regressand and the discriminant scores and the trend variable as the regressor variables has been fitted. The form of model considered is as follows:

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where symbols are as defined earlier.

Model-D_3

In this procedure, all 56 (weighted and un-weighted including interaction indices) have been used as discriminating variables in discriminant function analysis and two sets of discriminant scores from two discriminant functions have been obtained. Forecasting model has been fitted taking un-trended yield as the regressand variable and the two sets of discriminant scores and the trend variable (T) as the regressor variables. The form of the model fitted is as follows:

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where symbols are as defined earlier.

Model-D_4

In this procedure, 7 weighted and 7 un-weighted weather indices have been used as discriminating variables. Now, based on these 14indices, the discriminant function analysis has been done and two sets of scores have been obtained. On the basis of these two sets of scores, the regression model has been fitted taking the yield as the regressand and the two sets of scores and the trend variable (T) as the regressors. The fitted model here is

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where symbols are as defined earlier.

Model-D_5

In this procedure, discriminant function analysis has been carried out using the un-weighted and weighted average (weather indices) for the first weather variable (here, discriminating factors will be only two). Using the two sets of discriminant scores on the basis of first weather variable, and un-weighted and weighted average (weather indices) for the second weather variable, discriminant function analysis has been further carried out (here, the discriminating factors will be 4). This process is continued up to seven weather variables, and finally we get two sets of discriminant scores ds_1 and ds_2. Using crop-yield as regressand and discriminant scores ds_1& ds_2 and the time trend T as regressor variables, the following model is fitted for the development of forecast model.

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where symbols are as defined earlier.

Model-D_6

In this procedure, discriminant function analysis have been carried out using weekly data of the first weather variable spread over 15 weeks as discriminating variable. Using two sets of discriminant scores obtained from two estimated discriminant functions based on data of the first weather variable and 6 weeks data of the second variable, discriminant function analysis has been again performed and two sets of discriminant scores are obtained (here discriminating variable will now become 17). Using these two sets of discriminant scores and 15 weeks data of third weather variable have been again used to carry out discriminant function analysis and subsequently two sets of discriminant scores have been obtained. This process is continued up to seven weather variables, and ultimately we get two sets of discriminant scores ds_1 and ds_2. These two sets of scores and the trend variable (T) as the regressor variables and crop-yield as the regressand were utilized to develop fitting forecast model by the following model.

$$y = \beta o + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

where symbols are as defined earlier.

Measures for the comparison and validation of different models

Different models have been used in the present study for the comparison and the validation of the models developed. These models are given below:

R^2 (Coefficient of Determination)

It is in generally used for checking the adequacy of the model. R^2 is given by the following formula;

$$R^2 = 1 - \frac{ss_{res}}{ss_t}$$

wheress_{res} and ss_t are the residual sum of square and the total sum of square respectively.

R^2 never decreases when a regressor is added to the model, regardless of the value of the contribution of the variable in the model. Therefore, it is difficult to judge whether an increase in R^2 is really important. So, it is preferable to use Adjusted R^2 when models to be compared are based on different number of regressors. Adjusted R^2 is given by the following formula

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

wheren is the number of observation and p is the number of regressor variables. The total mean square is constant regardless of how many variables are in the model.

On adding a regressor in the model Adjusted R^2 increases only if the addition of the regressor reduces the residual mean square. It also penalizes for adding terms that are not helpful, so it is very important in evaluating and comparing the regression models.

Percent Deviation

This measures the deviation (in percentage) of forecast from the actual yield data. The formula for calculating the percent deviation of forecast is given below:

$$\text{Percentage deviation} = \frac{\text{(actual yield- forecasted yield)}}{\text{Actual Yield}} \times 100$$

Percent Standard Error of the Forecast (CV)

Let $\hat{y}_f$ be forecast value of crop yield and X_0 be the vector of selected values for regressor variables for the yield is forecasted.

The variance of $\hat{y}_f$ as given in (Draper and Smith, 1998) is obtained as

$$V(\hat{y}_f) = \hat{\sigma} X_0{}' (X'X)^{-1} X_0$$

Where $X^{'}X$ is the dispersion matrix of the sum of square and cross products of regressor variables used for the fitting the model and $\hat{\sigma}^2$ is the estimated residual variance.

The percent standard error (PSE) of forecast yield $\hat{y}_f$ is given by

$$\text{PSE} = \frac{\sqrt{V(\hat{y}_f)}}{\text{-Standard year}} \times 100$$

Infect, the PSE is the coefficient of variation (C.V.) of forecast yield.

Root Mean Square Error (RMSE)

It is also a measure of comparing two models. The formula of RMSE is given below

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

Where O_i and the E_i are the observed and forecasted value of the crop yield respectively and n is the number of years for which forecasting has been done.

Results and Discussion

The forecast model obtained under the fore strategies (as discussed under section 3) along with adjusted coefficient of determination (adj R^2) are presented in Table 1. In all the models, trend variable *T* was found significant. Apart from trend *T* other significant variables were found as discriminant scores *ds*1, *ds*2,in model-1, *to ds1* &*ds2* in model-4. Adjusted coefficient of determination (Adj R^2)varied between 0.55 to 0.78 in different modes, the maximum (0.78) being in model-6. RMSE was com on the basis of yield forecasts for the years 2014-15 to 2016-17. The results (Table 2) revealed that the percent deviation of forecast varied from 3.77 to 27.58 in model-1, 14.67 to 28.6 in model-2, 7.46 to 39.68 in model-3,1.3.57 to 27.95 in model-4,3.57 to 27.35 in model – v and 7.46 to 39.68 in model VI over the three years.

Table.1 Wheat forecast yield models

Model	Forecast regression equation	R^2	R^2adj
1	Yield=22.119+0.671 ds_1 +0.203 ds_2+0.232 T	61.5	55.8
2.	Yield=22.242- 0. 816ds_1 + 0.166 ds_2 + 0.219T	69.9	65.4
3.	Yield=22.654+0.680ds_1+0.022ds_2+0.1817T	81.0	78.2
4.	Yield=22.118- 0.671 ds_1-0.204 ds_2 + 0.232 T	61.6	55.8
5.	Yield=22.118- 0.671 ds_1-0.204 ds_2 + 0.232 T	61.6	55.8
6.	Yield=22.654+ 0.680ds_1+0.022 ds_2 + 0.181 T	81.0	78.2

Note :Figures is denoted Regression Equation and R.

Table.2 Actual and forecasts of wheat yield in (Q/ha)

Crop Year	Actual Yield	Forecast Using					
		Model-I	Model-II	Model-III	Model- IV	Model- V	Model- VI
2014-15	22.60	26.28 (27.58)	26.11 (26.79)	28.77 (39.68)	26.23 (27.35)	26.23 (27.35)	28.77 (39.68)
2015-16	27.40	26.43 (3.77)	28.08 (2.50)	29.44 (7.46)	28.38 (3.57)	28.38 (3.57)	29.44 (7.46)
2016-17	32.89	28.00 (14.86)	28.06 (14.67)	26.16 (20.43)	27.95 (15.0)	27.95 (15.0)	26.16 (20.43)
RMSE		4.36	4.25	8.68	4.36	4.36	6.22

Note- Figures in bracket denoted % deviation of Forecast.

The RMSE varied from a minimum of 4.25 in model-II to a maximum of 8.68 in model-III. Thus it is concluded that model-VI is the most suitable model among the models considered forecasting wheat yield for Ayodhya district of Uttar Pradesh. The models provides reliable forecast around two month before harvest.

References

Agrawal, R. and Mehta, S.C. (2007). Weather based forecasting of crop yields pest and diseases – IASRI model. *J. Ind. Soci. Agril. Stat.* 61(2). 255-263.

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

Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis. *John Wiley & Sons*, New York.

Choudhary, A. and Sarkar, M.B. (1981). Estimation of rice yield though weather factors in a dry sub- humid region. *Mausam*, 32(4).393-396.

Connor, W.P. (1999). Forecasting Survival and Passage of Migratory Juvenile Salmonids. *North American Journal of Fisheries Management* 2000, 20,651–60.

Huda, A.K.S., Ghildyal, R.A., Tomar, V.S. and Jain, R.C. (1975). Contribution of climatic variable in predicting rice yield. *Agric. Met.*, 15: 71-86..

How to cite this article:

Ravi Prakash Gupta, V. N. Rai, Sarvesh Kumar, K. K Mourya and Annu. 2021. Statistical Models for Wheat Yield Forecasting using Discriminant Function Analysis of Weather Parameters. *Int.J.Curr.Microbiol.App.Sci.* 10(03): 967-973.
doi: https://doi.org/10.20546/ijcmas.2021.1003.123