International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2019; 7(6): 2960-2962

Received: 04-09-2019
Accepted: 06-10-2019

Sarvesh Kumar
Department of Agriculture Statistics Narandra Deva University of Agriculture and Technology Kumarganj Ayodhya, Uttar Pradesh, India

VN Rai
Department of Agriculture Statistics Narandra Deva University of Agriculture and Technology Kumarganj Ayodhya, Uttar Pradesh, India

Mo Azfar
Department of Agriculture Statistics Narandra Deva University of Agriculture and Technology Kumarganj Ayodhya, Uttar Pradesh, India

Annu
Department of Agriculture Statistics Narandra Deva University of Agriculture and Technology Kumarganj Ayodhya, Uttar Pradesh, India

Ravi Prakash Gupta
Department of Agriculture Statistics Narandra Deva University of Agriculture and Technology Kumarganj Ayodhya, Uttar Pradesh, India

Corresponding Author:
Sarvesh Kumar
Department of Agriculture Statistics Narandra Deva University of Agriculture and Technology Kumarganj Ayodhya, Uttar Pradesh, India

Pre- harvest forecast model using linear regression model based on weather indices

Sarvesh Kumar, VN Rai, Mo Azfar, Annu and Ravi Prakash Gupta

Abstract
In the present paper, an application of regression analysis of weather variables (minimum & maximum temperature, relative humidity 7 hr & 14 hr, sun shine hour and wind velocity) for developing suitable statistical models to forecast Rapeseed and Mustard yield in Sultanpur district of Eastern Uttar Pradesh has been demonstrated. Time series data on Rapeseed and Mustard yield for 25 years (1990-91 to 2014-15) have been used in the regression mode. The forecast yield of Rapeseed and Mustard have been obtained from this model for the year 2012-13, 2013-14 and 2014-15, which were not included in the development of the model. This model has been found to be most appropriate on the basis of Adj R^2, percent deviation of forecast, percent root mean square error (%RMSE) and percent standard error (PSE) for the reliable forecast of Rapeseed and Mustard yield about one and half months before the crop harvest.

Keywords: Weather variables, rapeseed and mustard yield, regression analysis, forecast models

Introduction
In India Rapeseed and Mustard (*Brassica* spp.) is an important crop among oilseed crops in the world and second largest cultivator after China. According to area and production mustard stands in second place among oil seed crops of India.

Although there has been a significant increase in oilseed production since 1960s, the demand for oilseeds production in the future is likely to go up due to population increase and their income. Mustard is much sensitive to climatic variables and hence climate could have significant effect on its production. Uttar Pradesh is a leading mustard producing state of India 60% of total mustard production is from this state. Rapeseed & mustard is one of major oil seeds crops of Uttar Pradesh in general and particularly in Eastern Uttar Pradesh. There are very limited studies to assess the impact of climate change on oilseed crops as compared to cereals.

Forecasting opens menu window on to future. It plays a vital role in most of our activities and in all we do concern about future. Establishing a functional crop yield production forecasting system is an extremely important component of a national agricultural statistical system. Crop forecasts are used for a number of policy decisions such as those relating to procurement, trade, storage, transportation, distribution, import and export of food grains in general, and for implementing food security programmed in particular. Crop production forecasts are not only meant to serve the interests of governments but other stakeholders in the agricultural sector and also find use for crop forecast data in their day-to-day decision functions. The value of the various policy and business decisions could be enhanced if these are supported by a strong system of food crop production forecasting.

The crop yield depends on many types of variables which are discussed above viz. weather factors, plant characters during crop growth stages, pest and diseases attack, agricultural inputs etc. will provide better forecast if all these variables are considered while forecasting crop yield. But it may not reliable to develop a single model based on different type of the data. In such case separate models may be developed based on different group of variables and forecast obtained from these models may be combined to get a composite model (Agrawal *et al.* and Jain *et al.*).

Materials and Statistical Methodology

Sources and description of data

Yield data

Time series data on yield of rapeseed & mustard crop of Sultanpur district of Uttar Pradesh for 25 years (1990-91 to 2014-2015) have been collected from the Bulletins of Directorate of Agricultural Statistics and Crop Insurance, Govt. of Uttar Pradesh.

Weather data

Weekly weather data for the same period on the weather variables of Sultanpur district of Uttar Pradesh during the different growth phases of Rapeseed & Mustard crop have been obtained from Department of meteorological centre Amausi Airport Lucknow U.P. India. The data have been collected up to the first 28 weeks of the crop cultivation which includes 40th standard meteorological week (SMW) to 52nd SMW of a year and 1st SMW to 15th SMW of the next year. The data on six weather variables viz. Minimum Temperature, Maximum Temperature, Relative humidity at 7 hr, Relative humidity at 14 hr, Wind-velocity and Sun-shine (hr) have been used in the study.

Crop season

Rapeseed & Mustard is generally sown in the month of October and harvested in April. The entire crop growth phases are discussed below.

Statistical Methodology

$$Y = a + \sum_{i=1}^{p}\sum_{j=0}^{1} b_{ij} Z_{ij} + \sum_{i\neq i'=1}^{p}\sum_{j=0}^{1} b_{ii'j} Z_{ii'j} + cT + e$$

where $Z_{ij} = \sum_{w=1}^{m} r_{iw}^{j} X_{iw}$

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

Y is yield, p is number of weather variables used, m is number of weeks considered for the model building, r_{iw} / $r_{ii'w}$ are the correlation coefficients of yield and the i^{th} weather variable (X_i) the product of the two weather variables (X_i and $X_{i'}$) in w^{th} week $(i,i' = 1,2,.....6)$ T is time trend variable. a, b_{ij}, $b_{ii',j}$ and c are model parameters. e are error terms assumed to follow independently normal distribution with mean zero and variance σ^2. The model has been fitted for different values of m (m=19, 20,...., 23). The data after 23rd week have not been used as the idea was to forecast yield well in advance of harvest.

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;

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

where ss_{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

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

where n 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:

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

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

$$PSE = \frac{\sqrt{V(\hat{y}_f)}}{\text{Forecast yield}} \times 100$$

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

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

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.

Result and Discussion

Considering the actual Rapeseed & Mustard yield as regressand and 42 weather indices generated and time trend (T) as regressor variables, the model was fitted using step-wise regression analysis. The results are presented in Table 1.

Table 1: Estimate of regression coefficient of finally entered variables along with their standard error.

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2 (%)
1	Constant	9.589	2.266	96.5**
2	Z_{11}	-1.243**	0.160	
3	Z_{61}	0.015**	0.003	
4	Z_{71}	-0.006	0.002	
5	Z_{561}	0.194**	0.030	
6	Trend	0.119**	0.016	

* P< 0.05, ** P<0.1

Forecast model- A

$Y = 9.589 - 1.243\ Z_{11} + 0.015\ Z_{61} - 0.006\ Z_{71} + 0.194 Z_{561} + 0.199T$

where, Z_{10} = unweighted average of minimum temperature.

Z_{11} = weighted average of minimum temperature.

Z_{561} = weighted interaction between wind velocity and sunshine hour.

These three weather indices and time trend have been found significant variables for forecasting the pre-harvest Rapeseed & Mustard yield at 8th SMW of crop production.

The model is validated by forecasting the Rapeseed & Mustard yield for the years 2012-13 to 2014-15. The results of validation are given in Table 2. The values of percent deviation of forecast yield from actual yield, %RMSE and PSE were also computed and are presented in the Table 2.

It can be observed from the results of the tables 1 and 2 that finally entered regressor variables were significant. The value of adjusted R^2 is also quite high to the extent of 96%. The PSE and RMSE are also quite low indicating thereby that the model is best fitted and it has high power to pre-harvest forecast Rapeseed & Mustard yield at reproductive about one and half months before the harvest.

Table 2: Validation of the Model

Year	Actual yield	Predicted yield	R^2 % (Adj. R^2)	Percent deviation	PSE	RMSE
2012-13	7.32	10.13	97.3** (96.5**)	38.51	2,96	2.43
2013-14	4.99	8.02		60.90	2.16	
2014-15	4.45	5.21		17.20	5.40	

** P<0.1

References

1. Agrawal R, Jain RC, Mehta SC. Yield forecast based on weather variables and agricultural input on agro climatic zone basis. Ind. J Agric. Sci. 2001; 71(7):487-90.
2. Agrawal R, Mehta SC. Weather based forecasting of crop yields pest and diseases – IASRI model. J Ind. Soci. Of Agril. Stat. 2007; 61(2):255-263.
3. Agrawal R, Chandrahas, Kaustav A. Use of discriminant function analysis for forecasting crop yield, *Mausam*, 2012; 63(3):455-458.
4. Jain RC, Sridharan H, Agrawal, Ranjana. Principal component technique for forecasting of sorghum yield. Indian Journal of Agril. Sci. 1984; LI, 1:61-72.
5. Jain RC, Jha MP, Agrawal, Ranjana. Use of growth indices in yield forecast. Biometrical Journal. 1985; 27(4):435-439.
6. Yadav ST, Patil BP. Influences of agro-climate indices on fruit yield of cucumber. J Maha. Agric. Univ. 2008; 33(2):237-2.