Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
www.phytojournal.com
JPP 2021; Sp 10(2): 44-46
Received: 19-01-2021
Accepted: 24-02-2021

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

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

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

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

Corresponding Author:
Ravi Prakash Gupta
Department of Agricultural Statistics, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya, Uttar Pradesh, India

Statistical models for wheat yield using linear regression model based on meteorological parameters

Ravi Prakash Gupta, VN Rai, Sarvesh Kumar and Snehdeep

Abstract
In the present paper, an application of regression analysis of weather variables (minimum & maximum temperature, relative humidity 7 hr & 14 hr, rainfall, rainy day and wind velocity) for developing suitable statistical models to forecast Wheat yield in Ayodhya district of Eastern Uttar Pradesh has been demonstrated. Time series data on Wheat yield for 27 years (1990-91 to 2016-17) have been used in the regression model. The forecast yield of Wheat have been obtained from this model for the year 2014-15, 2015-16 and 2016-17, which were not included in the development of the model. This model has been found to be most appropriate on the basis of Adj R2, percent deviation of forecast, percent root mean square error (% RMSE) and percent standard error (PSE) for the reliable forecast of Wheat yield about two months before the crop harvest.

Keywords: Weather variables, wheat yield, regression analysis, forecast models

Introduction
Wheat (*Triticumaestivum* 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 utilised as a human food and livestock feed. It can be grow in tropical, sub -tropical and temperate region. Wheat shows highest yield potential in regions favoured with cool, moist weather and followed by dry condition.

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 programme in particular. Crop production forecasts are not only meant to serve the interests of governments but other stakeholders in the agricultural sector 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.

Weather variables affect the crop at different stages of crop development. Thus extent of weather influence on crop yield depends not only on magnitude of weather variable but also on the distribution pattern of weather over the crop season which, as such calls for the necessity of dividing the whole crop season in to finer intervals. However doing so will increase number of variable in the models and in turn a large number of observations may not be available for precise estimation of these parameters. Suitable dimension technique such as like transformation say forming indices or multivariate technique such as principle component, discriminate function etc. can be employed to tackle such problem.

Materials and statistical methodology
Sources and description of data
Yield data
Time series data on yield of wheat crop of Ayodhya district of Uttar Pradesh for 27 years (1990-91 to 2016-2017) 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 Ayodhya district of Uttar Pradesh during the different growth phases of Wheat crop have been obtained from Department of meteorology A.N.D.U.A. & T Kumarganj Ayodhya U.P. India. The data have been

collected up to the first 20 weeks of the crop cultivation which includes 44th standard meteorological week (SMW) to 52nd SMW of a year and 1st SMW to 11th SMW of the next year. The data on seven weather variables *viz.* Minimum Temperature, Maximum Temperature, Relative humidity at 7 hr, Relative humidity at 14 hr, Wind-velocity Rainfall and Rainy days have been used in the study.

Statistical methodology

$$Y = a + \sum_{i=1}^{p}\sum_{j=0}^{1} b_{ij} Z_{ij} + \sum_{i\neq i'=1}^{p}\sum_{j=0}^{1} b_{ii'j} Z_{ii'j} + cT + e \quad \text{... (3.3.1.5)}$$

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 ith weather variable (X_i) the product of the two weather variables (X_i and $X_{i'}$) in wth 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 = 13, 14…17). The data after 11th week have not been used as the idea was to forecast yield well in advance of harvest.

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

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

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:

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

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

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

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 Wheat as regress and and 56 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 finely entered variables along with their standard error

S. No.	Variables	Regression coefficient	Standard error	Adjusted R^2%
1.	Constant	22.947	3.669	90.8
2.	Z_{11}	-.988	.191	
3.	Z_{141}	.007	.003	
4.	Z_{231}	-.006	.001	
5.	Z_{241}	-.009	.001	
6.	Z_{371}	-.023	.008	
7.	T	.260	.023	

$$Y = 22.947 - .988Z_{11} + .007Z_{141} - .006Z_{231} - .009Z_{241} - .023Z_{371} + .260T$$

Where, Z_{10} = unweighted average of minimum temperature.
Z_{11} = weighted average of minimum temperature.
Z_{671} = weighted interaction between Rainy day and wind velocity.

These three weather indices and time trend have been found significant variables for forecasting the pre-harvest Wheat at 6th SMW of crop production.
The model is validated by forecasting the Wheat yield for the years 2014-15 to 2016-17. 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 90.8%. 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 Wheat yield at reproductive about Two months before the harvest.

Table 2: Validation of the model

Year	Actual Yield	Predicted Yield	R2% (AdjR2)	Percent deviation	PSE	RMSE
2014-15	20.60	26.21	92.9** (90.8)	27.27	14.32	6.51
2015-16	24.40	21.64		21.03	18.61	
2016-17	32.89	24.97		24.07	16.94	

References

1. Agrawal R, Jain RC. Composite model for forecasting rice yield. Ind. J Agri. Sci 1982;52(3):177-181.
2. Agrawal R, Mehta SC. Weather based forecasting of crop yield, pest and diseases – IASRI Models. J Ind. Soc. of Agril. Stat 1982;61(2):255-263.
3. Agrawal R, Jain RC, Jha MP. Joint effects of weather variables on rice yields. Mausam 1983;34(2):177-81.
4. Fisher RA. The influence of rainfall on the yield of wheat at Rothamsted. Roy. Soc. (London), Phil. Trans. Ser. B 1924;213:89-142.
5. Garde YA. Some statistical Models for crop yield forecasting, M.Sc. (Agril. Stat.) Thesis, G.B.P.U.A.T., Pantnagar, Uttarakhand 2009.
6. Garde YA, Shukla AK, Singh S. Pre-harvest forecasting of rice yield using weather indices in Pantnagar, Uttarakhand 2012.
7. Jain RC, Agrawal Ranjana, Jha MP. Effect of climatic variables on rice yield and its forecast. Mausam 1980;31(4):591-96.
8. Jain RC, Sridharan H, Agrawal Ranjana. Principal component technique for forecasting of sorghum yield. Indian Journal of Agril. Sci 1984;51(1):61-72.
9. Yadav RR, Sisodia BVS, Kumar S. Application of Principal Component Analysis in Developing Statistical Models to Forecast Crop Yield Using Weather Variables. Mausam 2015;65(3):357-360.