*Journal of Agrometeorology 23 (4) : 452-456 (December 2021) : https://doi.org/10.54386/jam.v23i4.162*

***Short communication***

# Development of yield prediction model of rice crop for hilly and plain terrains of Uttarakhand

**PARUL SETIYA* and A. S. NAIN**

*Department of Agrometeorology, G.B. Pant University of Agriculture and Technology, Pantnagar, India*
*Corresponding author email *: parul.setiya@gmail.com*

Indian economy is largely based on agriculture, the production capacity of which is predisposed to exceptional weather conditions observed over prolonged period, if other factors such as soil, management practices etc. are assumed to remain unchanged in different years. Favoring weather might give bumper production as against in a year with some exceptional weather hazards or incidence of any yield reducing pathogen and pests. The great geographical diversity of India owing to its range/variety of climate, soils and terrain types produce numerous farming situations, each unique in itself, favoring some crops while restraining others. Decision makers and planners need information on crop statistics related to crop acreage and expected production at local level, which is obtained by surveying and ground truthing, and then are produced at regional, state and country level. This information helps to have knowledge of current state of the food grain in the state and country, i.e., identify the crops in which a region is surplus and deficit, and in this way facilitates equitable and timely distribution and thus ensure food security to all. Therefore, an advance estimate of crop yield before harvest is very crucial task mainly for the regions that have high chances of climatic uncertainties. A very common approach for developing the prediction model is to measure the trend of historical yield data with weather parameters.

Several yield prediction models have been developed using number of weather variables and satellite based spectral indices to establish crop weather relationship for different crop grown in different regions in the country. These have been widely used to evaluate the association between weather parameters and crop yield of different regions (Banakara, *et al.,* 2018; Latwal, *et al*., 2017; Diwan *et al*., 2018 and Kumar *et al*., 2014). Bankara *et al*. (2018) compared the technique of Multiple Linear Regression (MLR) and Principal Component Analysis (PCA) to study the crop weather relationship and concluded that the MLR technique performs better than PCA for pre harvest forecasting of rice crop yield. Diwan *et al.* (2018) also developed crop yield forecast model by employing stepwise linear regression technique and found that temperature (maximum & minimum) and relative humidity were significant predictors in crop yield forecast. Similar studies were conducted by Kalubarme and Ahuja, (1996); Chauhan *et al.,* (2009) to develop agrometeorological data based rice yield prediction model for Karnal, central Punjab and Bulsar district of Gujarat respectively. Ghosh *et al.,* (2014) also developed rice forecast model for various district of West Bengal. Moreover, Latwal *et al.* (2017) used a combination of NDVI and weather variables to develop spectral-meteorological models for district-level crop yield prediction. Researchers observed that spectral-meteorological models could help to forecast the crop yield with highest accuracy. Kumar *et al.* (2014) developed regression models by employing modified Hendrick and Scholl technique to develop forecast model for paddy and sugarcane crop for six districts of south Gujarat and concluded that the combination of weather and yield data is appropriate and consistent option for yield forecasting.

In view of the above, the present study was carried out to examine the effect of weather variability on rice yield production at Udham Singh Nagar and Nainital district of Uttarakhand, India.

For the development of the crop yield model fortnightly data of weather parameter including rainfall, minimum temperature, maximum temperature, relative humidity and solar radiation spanning from 2001 to 2018 was obtained from the Agrometeorological observatory located at Norman E. Borlaug Crop Research Centre (NEBCRC), G. B. Pant University of Agriculture and Technology, Pantnagar, Udham Singh Nagar, Uttarakhand, whereas the historical yield data has been drawn from the Dacnet website. The yield data were detrend before performing the statistical analysis.

An attempt was made to develop a regression model that can be used in the prediction of rice crop yield based on the variability in rainfall (mm), maximum temperature ($^0$C), minimum temperature ($^0$C) and solar radiation over the area of study. The models were developed based on the yield and weather data of 15 years, i.e. 2001-2015 and the developed models were validated one the data of three years, i.e. 2016-2018. For the statistical analysis of dataset, linear regression with step wise technique was employed. The analysis was performed with the help of Statistical Package for Social Sciences (SPSS) software. The significance of the model was tested by considering the alpha level at 5%. Before conducting the regression analysis, the required assumptions (normality, homoscedasticity and multicollinearity) for conducting the regression analysis were evaluated.

Normality was evaluated using a Q-Q scatter plot (Bates *et.al.*, 2014; Field, 2013). The Q-Q scatter plot compares the distribution of the residuals with a normal distribution (a theoretical distribution). For the assumption of normality to be met, the quantiles of the residuals must not be deviate from the theoretical quantiles i.e. the points should form a relatively straight line. Strong deviations of points from straight line could indicate that the parameter estimates are unreliable. Moreover, the condition of homoscedasticity was evaluated for each model by plotting the model residuals against the predicted model values (Osborne & Walters, 2002). The assumption is met if the points appear randomly distributed and have no apparent curvature. The condition of multicollinearity was tested by using Durbin-Watson test. The value of Durbin-Watson test statistics close to 2 indicates the absence of multicollinearity.

Further, Adjusted $R^2$ was used to check the proportion of variance in the yield due to the weather parameters. Besides this, Percentage Error and Root Mean Square Error (RMSE) have been used as a statistical measure. Percentage Error (PE) measures the discrepancy between an observed and expected value while Root Mean Square Error (RMSE) measures the model performance. It shows the concentration of the data around the line of best fit.

***Multiple linear regression analysis***

Multiple linear regression models involved more than one independent variable and one dependent variable. It is of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

Where, $\beta_0$ is constant, $\beta_i$'s are coefficients of $X_i$'s, $X_i$'s are the independent variables also known as predictors and Y is the dependent variable and ε is the error.

***Percentage error (P.E.) and root mean square error (RMSE)***

The comparison of developed pre harvest forecasting model can be done by using the forecast percentage error and it is calculated by

$$P\ E = \frac{Observed\ \ Yield - Expected\ \ Yield}{Observed\ Yield} * 100$$

In addition to the Per cent Error (P.E.), Root Mean Square Error (RMSE) was also calculated to compare the developed models. The formula of RMSE is

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Observed\,Yield - Expected\,Yield)^2}$$

Here, n represents the number of years.

***Regression analysis for conceptual model***

Linear regression analysis was conducted to assess whether the weather parameters (Tmax, Tmin, rainfall, relative humidity and solar radiation) significantly predicted the yield of the rice crop of Udham Singh Nagar and Nainital district. The 'Stepwise' variable selection method was used to develop the linear regression model. In the stepwise technique, at each step predictor variables get added and the significance of the model was tested at each step.

**Table 1:** Error analysis of forecasting models

| Disrtrict | Year | Rice Yield | | | |
|---|---|---|---|---|---|
| | | Forecasted | Observed | Error (%) | RMSE |
| U.S.N. | 2016 | 3.57 | 3.28 | -8.94 | |
| | 2017 | 3.58 | 3.54 | -1.05 | 0.03 |
| | 2018 | 3.19 | 3.22 | 0.76 | |
| Nainital | 2016 | 2.69 | 2.83 | 4.93 | 0.05 |
| | 2017 | 2.81 | 3.43 | 17.90 | |
| | 2018 | 2.90 | 3.19 | 8.92 | |

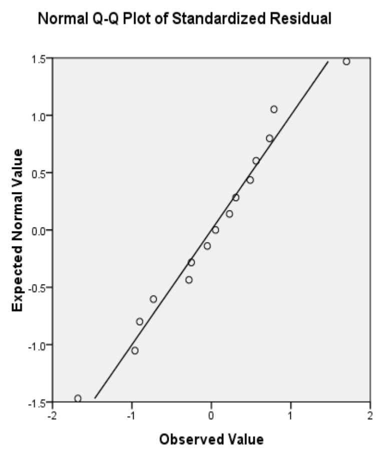


**Fig. 1:** Q-Q scatterplot testing normality (USN)

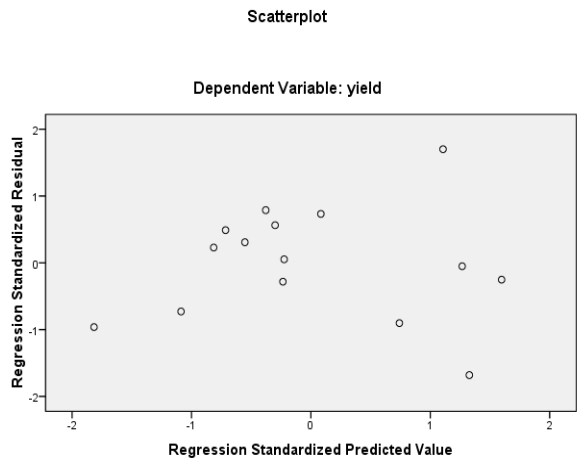


**Fig. 2**: Residuals scatterplot testing homoscedasticity (USN)

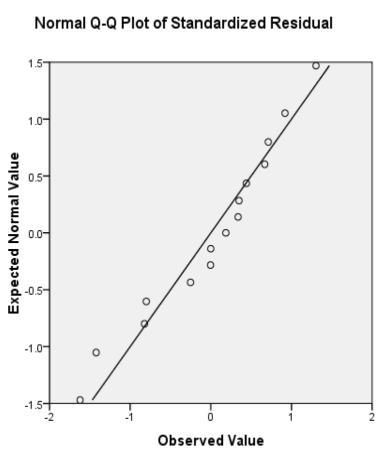


**Fig. 3:** Q-Q scatterplot testing normality (Nainital)

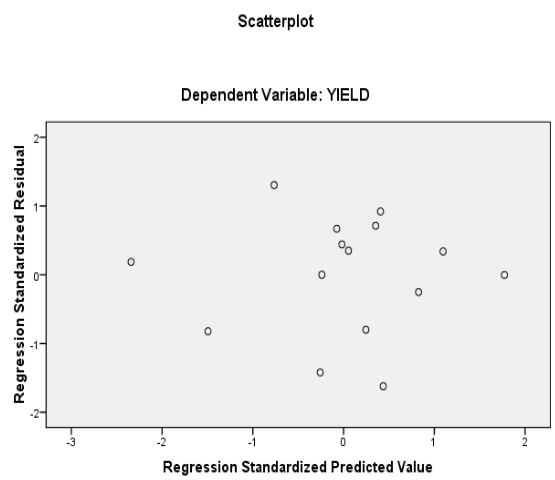


**Fig. 4**: Residuals scatterplot testing homoscedasticity (Nainital)

***Regression model for Udham Singh Nagar and Nainital district predicting rice yield***

Regression model was developed for the prediction of rice crop based on the weather parameter at Udham Singh Nagar and Nainital district of Uttarakhand. Prior to conducting the regression analysis, the assumptions of normality of residuals, homoscedasticity (equal variance) of residuals and multicollinearity were examined. Q-Q scatter plot was used to assess normality,

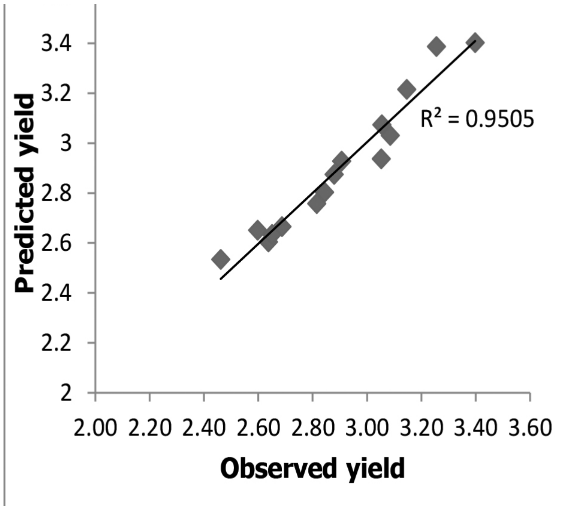


**Fig. 5:** Scatter plot of predicted and observed yield of Udham Singh Nagar district

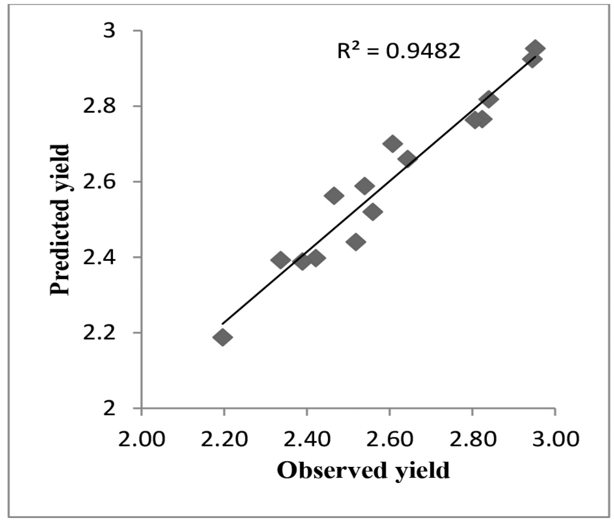


**Fig. 6:** Scatter plot of predicted and observed yield of Nainital district

homoscedasticity was assessed with a residuals scatter plot and Durbin-Watson test was calculated to check the multicollinearity. The Q-Q scatter plot for normality for the dataset of USN and Nainital district is presented in Fig. 1 and Fig. 3 while Fig. 2 and Fig. 4 shows the scatter plot of predicted values and model residuals for USN district and Nainital district, respectively. The condition of normality was met for the dataset of both the districts as the points in Q-Q scatter plot (Fig. 1 and Fig. 3) form a straight line. The condition of homoscedasticity was also met for the dataset of both districts as the points in Fig. 2 and Fig. 4 were randomly distributed around the mean.

Durbin-Watson test was used to access the multicollinearity, the value of Durbin-Watson test statistic for the dataset of USN and Nainital districts was 1.60 and 2.50, respectively, indicated the absence of multicollinearity. The regression analysis shows significant effect of predictors (weather parameters) on rice crop yield.

The statistical model to predict the rice crop for Udham Singh Nagar and Nainital district are shown in eq(1) and eq(2) respectively.

$Y= - 404.35+8.20^* (X_1)+0.02^* (X_2)+2.17^* (X_3)+2.31^* (X_4)$

Here, Y = rice crop yield (dependent variable); $X_{1=}$average maximum temperature of the second fortnight of the month of August; $X_{2=}$ Total rainfall of the first fortnight of the month of August; $X_{3=}$ average maximum temperature of the second fortnight of the month of September; $X_{4=}$ average maximum temperature of the second of the month of October;

$Y=-71.249+4.601^*(X_1) +0.244^*(X_2) -0.038^*(X_3) -2.929^*(X_4)$

Here, Y = rice crop yield (dependent variable); $X_{1=}$average maximum temperature of the second fortnight of the month of July; $X_{2=}$average relative humidity of the first fortnight of the month of October; $X_{3=}$ Total rainfall of the second fortnight of the month of September; $X_{4=}$ average maximum temperature of the second fortnight of the month of September

Linear regression models were significant for U.S.N. ($p< .001$, $R^2 = 0.90$) as well as for Nainital ($p< .001$, $R^2 = 0.89$) district, indicating that approximately 90 per cent of the variation in rice yield of USN district was explainable by weather parameters with RMSE = 0.03, whereas 89 per cent variation in rice yield of Nainital district was explainable by weather parameters with RMSE = 0.05.

These developed models have also been validated for three year i.e. 2016, 2017 and 2018. Value of per cent error indicates little under estimation of rice yield for USN district, whereas the model showed over-estimation for Nainital district (Table 1). The scatter plots of predicted and observed yield of Udham Singh Nagar and Nainital district has been shown in Fig. 5 and Fig. 6, respectively.

The results indicated that that maximum temperature of the second fortnight of August and rainfall of the first fortnight of August played important role in the development of rice crop at USN district, while relative humidity of the first fortnight of October, maximum temperature of the second fortnight of July rainfall of the second fortnight of September were significant predictor in the development of rice crop at Nainital district. The developed multiple regression models were found to be efficient in forecasting yields for Udham Singh Nagar as well as for Nainital district. Therefore these models can be used by various stakeholders to predict the rice yield for Udham Singh Nagar and Nainital districts of Uttarakhand.

***Conflict of Interest Statement***: The author(s) declare(s) that there is no conflict of interest.

***Disclaimer***: The contents, opinions, and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

***Publisher's Note:*** The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

Banakara, K. B. and Popat R. C., Amaresh and Pandya H. R. (2018). Pre-harvest forecast of Kharif Rice using weather parameters in Bharuch district of Gujarat state. *Int. J. Chem. Stud.*, *6*(6), 1111-1116.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Chauhan, V.S., Shekh, A.M., Dixit, S.K., Mishra, A.P. and Kumar,S. (2009). Yield prediction model of rice in Bulsar district of Gujarat. *J. Agrometeorol.,* 11 (2): 162-168.

Diwan, U. K., Puranik, H. V., Das, G. K. and Chaudhary, J. L. (2018). Yield prediction of wheat at pre-harvest stage using regression based statistical model for 8 district of Chhattisgarh, India. *Int. J. Curr.Microbiol. Appl. Sci*.,7(1):2180-2183,

Field, A. (2013). Discovering statistics using *SPSS* (4th ed.). Thousand Oaks, CA: Sage.

Ghosh, K., Balasubramanian, R., Bandopadhyay, S., Chattopadhyay, N., Singh, K. K. and Rathore, L. S. (2014). Development of crop yield forecast models under FASAL-a case study of kharif rice in West Bengal. *J. Agrometeorol*, 16(1) 1:8.

Kalubarme, M.H., and Ahuja, S.C. (1996). Agrometeorological rice yield modelling in Karnal. *J. Indian Soc. Remote Sens.*, 24:125–132.

Kumar N., Pisal R., Shukla, S. P. and Pandey, K. K. (2014). Crop yield forecasting of paddy, sugarcane and wheat through linear regression technique for south Gujarat. *Mausam*. 65: 361-364.

Latwal, A., Nain, A. S., Jha, A. and Ranjan, R. (2017). Wheat yield prediction using satellite data and meteorological parameters for Udham Singh Nagar district of Uttarakhand. *Int. J. Basic Appl. Agril. Res*., 15(3): 130-132.

Osborne, J. and Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Pract. Assess. Res. Evaluation*, 8(2): 1-9.

*Received : October 2020 : Accepted : August 2021*