

## Prediction of Wheat Yield in Uttar Pradesh using Multiple Linear Regression Approach

Nitesh Kumar Yadav<sup>1\*</sup>, R. Vasanthi<sup>2</sup>, R. Pangayar Selvi<sup>2</sup> and Tmt. G. Vanitha<sup>3</sup>

<sup>1</sup>PG Scholar, Department of Physical Sciences & Information Technology,  
AEC & RI, TNAU, Coimbatore-641003 (Tamil Nadu), India.

<sup>2</sup>Associate Professor (Mathematics), Department of Physical Sciences & Information Technology,  
AEC&RI, TNAU, Coimbatore-641003 (Tamil Nadu), India.

<sup>3</sup>Assistant Professor (CS), Office of Dean SPGS, TNAU, Coimbatore – 641003 (Tamil Nadu), India.

(Corresponding author: Nitesh Kumar Yadav\*)

(Received 22 April 2022, Accepted 23 June, 2022)

(Published by Research Trend, Website: [www.researchtrend.net](http://www.researchtrend.net))

**ABSTRACT:** Development of effective agricultural and food policy at the regional and global levels, accurate crop production estimates are essential. Multiple regression approach has been used to prediction of the crop yield production widely. Thirty years (1991-2020) of weather data and yield data of wheat for thirteen districts of Uttar Pradesh are taken from Directorate of economics and Statistics and NASA metrological data, which are main producer of wheat was used to develop yield prediction equations. MLR model were validated by predicting yield of three years (2018,2019 and 2020) data. The results indicated that the district Muzaffarnagar has less Mean Absolute Percentage Error (0.07376), Mean Absolute Error (0.25552), Standard Error (0.3622) comparing to remaining districts. Using Backward regression in SPSS the variable Minimum temperature ( $T_{\min}$ ) is found to be significant and influences Wheat yield in Muzaffarnagar district. For Aligarh district Relative humidity is found to be significant for Wheat Yield prediction. By analysing the results MLR is found to be the best fit for the district Muzaffarnagar and Aligarh since it has less MAPE value comparing to other districts. The model can be used to some extent to forecast the yield in various districts of Uttar Pradesh.

However, One of the main producing region for wheat in both India and the rest of the world is the Uttar Pradesh Wheat yields in this area will be impacted by climate change due to variations in temperature and precipitation as well as decreasing water availability for irrigation, posing serious questions about the security of the food supply on a national and international level.

**Keywords:** Multiple linear regression, Mean Absolute Error, Mean Absolute Percentage Error, Standard Error, Weather Indices.

### INTRODUCTION

Agriculture is the backbone of the Indian economy, the expansion of agriculture and related industries remains a vital role in the Indian economy's overall performance. For effective planning and policymaking in the nation's agriculture sector, two components—crop acreage estimation and crop yield forecasting—are essential (Lal *et al.*, 1998). The estimation of agricultural output at the regional level serves as the foundation for planning crop production forecasts at the national level. Model based on weather variables can provide accurate predictions of crop output before harvest and early warning of pest and disease attacks, allowing for the timely implementation of relevant plant protection measures to protect the crops (Agrawal

and Mehta 2007). An important project called FASAL, which is being carried out by the Ministry of Agriculture of the Government of India in conjunction with the Space Application Centre (SAC), the Institute of Economic Growth (IEG), and the India Meteorological Department, forecasts agricultural output using space, agrometeorology, and land-based observations (IMD). IMD develops intra-seasonal operational yield forecasts at district and state levels for 13 important crops of India throughout the kharif and rabi seasons using a statistical model as part of the FASAL project in partnership with 46 Agromet Field Units (AMFU) situated in various regions of the nation (Ghosh *et al.*, 2014). Wheat is commonly grown for its seed, a cereal grain that is a staple diet for all across the world. The genus *Triticum* has numerous wheat species

and the most frequently grown wheat is *Triticum aestivum*. Wheat is farmed on the most land on the comparison of any food crop. India's wheat production is 106.41 million tonnes in the 2021-22 crop season, which ends in June. Wheat and Rice is the main producing crop in India and these two crop play important role in the economic growth of our country. For the purpose of predicting the growth and yield of four cultivars of rice (*Oryza sativa* L.) under three transplanting dates, three agroclimatic models based on growing degree-days, helio-thermal units, and photo-thermal units were developed at Srinagar in 2004 and 2005. These cultivars were "Jhelum," "K 39," "Shalimar rice 1," and "China 1007." (25 May, 10 June and 25 June) (Singh *et al.*, 2010). Wheat production reached 761 million tonnes in 2020, making it the second most widely grown cereal after maize, with a yearly production volume of 131,696,392 tonnes China is the world's greatest wheat producer and India is in second with a yearly production of 93,500,000 tonnes. Wheat is grown in three agroclimatic zones: western Uttar Pradesh (3.29 million hectares), eastern Uttar Pradesh (5.24 million ha), and central Uttar Pradesh (5.24 million ha) (0.68 million ha). The total area is 9.2 million hectares, with a total production of 24.5 million tonnes and a yield of 2.7 tonnes per hectare. The state of Uttar Pradesh produces the most wheat among all the states in India. It is situated in the Ganga's extremely fertile river basin. The state produces 300,00 tonnes of wheat. Wheat is produced on 96 lakh hectares in the state, on the whole 37% in India.

## METHODOLOGY

**Data Collection:** Data were collected for 13 district in UP from 1991-2020 (30 Year) from NASA research centre for Solar and meteorological data and Directorate of economics and statistics. Variables considered for this study are namely Yield, Maximum temperature ( $T_{max}$ ), Minimum temperature ( $T_{min}$ ), Relative humidity, Rainfall (mm) and Ph respectively. In this study 13 district have selected because of there higher production in Uttar Pradesh.

**Statistical Method:** The stepwise regression approach was used to pick the optimum regression equation from a large number of independent variables. The data was analysed using Rstudio software, with a probability level of 0.05 for entering and 0.1 for removing variables. To predict the yield of wheat for the subsequent years, a regression model was fitted using the entered variables derived via individual stepwise regression analysis. The multiple linear regression analysis was carried out for examining the Standard Error (SE) of estimated values resulting from various weather parameters. The statistical measures namely Mean absolute percentage error (MAPE), Mean absolute error (MAE) and Root Mean Square Error (RMSE) are used to fit the measured data to the degree of precision of each investigated correlation. Yield

forecast model have been created for all thirteen wheat-producing districts.

**Model Performance Metrics:** The effectiveness of the developed statistical models is examined using the mean absolute error (MAE), Mean absolute percentage error (MAPE) and Root Mean Square Error (RMSE). The following formula was used to calculate them:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \right]^{1/2}$$

$Y_i$  = Actual value

$\hat{Y}_i$  = Model output

The generated models perform better when the RMSE approaches 0, and the model fits the data better when the MAE and MAPE values are lower.

**MAE:** Mean Absolute Error is a model evaluation metric used with regression models. The mean absolute error of a model with respect to a test set is the mean of the absolute values of the individual prediction errors on over all instances in the test set. Each prediction error is the difference between the true value and the predicted value for the instance.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

**MAPE:** The mean absolute percentage error (MAPE) it is also called the mean absolute percentage deviation (MAPD) which measures accuracy of a forecast system. It measures this accuracy as a percentage, and can be calculated as the

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

The mean absolute percentage error (MAPE) is the most common measure used to forecast error, probably because the variable's units are scaled to percentage units, which makes it easier to understand. It works best if there are no extremes to the data (and no zeros). It is often used as a loss function in regression analysis and model evaluation.

**Root Mean Square Error (RMSE)** is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit. Root mean square error is commonly used in climatology, forecasting, and regression analysis to verify experimental results.

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \right]^{1/2}$$

## RESULTS AND DISCUSSION

Table 1 shows the yield fluctuations explained by the model for the wheat crop, along with the standard error. In all thirteen districts of Uttar Pradesh, The SE was ranged between 0.3622 (Muzaffarnagar) and 0.6023(Kanpur). The MAPE was ranged between 0.07367 (Muzaffarnagar) and 0.15322 (Gorakhpur). The MAE was ranged between 0.25552 (Muzaffarnagar) and 0.42938 (Gorakhpur). It is found to be that the MLR model is best fit for the district Muzaffarnagar since it has less MAPE value (0.07367). Among remaining 12 district, MLR suites for the district Aligarh (0.09032) since it has less MAPE next to the district Muzaffarnagar. This is comparable with the result of Singh *et al.* (2011).

For the validation of predicted model of wheat according to Singh *et al.* (2011) & Timbadia *et al.* (2021) for different district in the year 2018, 2019 and

2020 are shown in Table 2. Predicted yield was closed to observed yield; hence it can be utilised for yield forecasting and planning purpose. The results showed that, in relation to the main wheat-growing districts, the agro-meteorological yield model adequately predicted the yield variability caused by differences in minimum and maximum temperatures, rainfall as well as relative humidity. According to claim Singh *et al.* (2022) and Kumar *et al.* (2018), the per-hectare yield of wheat in India has decreased over the past few years as a result of the temperatures progressively rising in January, February, and March (This three month period are most crucial for the wheat crop). According to Lal *et al.* (1998) and Saxena *et al.* (2016)., in India, a rise of 0.5 °C in winter temperature is predicted to lower wheat production by 0.45 t ha<sup>-1</sup> because maximum and lowest temperatures are particularly sensitive weather parameters for the wheat crop.

**Table 1: Yield forecast models of wheat for different districts of Eastern Uttar Pradesh.**

District	Regression equation	SE	RMSE	MAPE	MAE
Agra	$7.222 - 0.113*X_1 - 0.093*X_2 + 0.044*X_3 + 0.015*X_4$	0.4939	0.44173	0.10766	0.35067
Aligarh	$3.962 - 0.42*X_1 - 0.049*X_2 + 0.040*X_3 + 0.01*X_4$	0.4721	0.42228	0.09032	0.29707
Bulandshahr	$9.051 - 0.118*X_1 - 0.128*X_2 + 0.037*X_3 - 0.089*X_4$	0.4523	0.40457	0.08941	0.32316
Etawah	$2.447 - 0.057*X_1 - 0.069*X_2 + 0.013*X_3 + 0.338*X_4$	0.5096	0.45579	0.12502	0.35946
Farrukhabad	$4.28 - 0.07*X_1 - 0.07*X_2 + 0.04*X_3 + 0.19*X_4$	0.5461	0.49	0.11049	0.35825
Fatehpur	$-4.399 + 0.102*X_1 - 0.246*X_2 + 0.016*X_3 + 0.278*X_4$	0.4219	0.37735	0.11051	0.30881
Gorakhpur	$-4.741 + 0.075*X_1 - 0.138*X_2 - 0.005*X_3 + 0.589*X_4$	0.5996	0.53628	0.15322	0.42938
Kanpur	$8.004 - 0.129*X_1 - 0.175*X_2 + 0.026*X_3 + 0.090*X_4$	0.6023	0.53875	0.15247	0.4193
Meerut	$-0.842 + 0.042*X_1 - 0.047*X_2 + 0.034*X_3 + 0.102*X_4$	0.5377	0.48091	0.10768	0.37072
Moradabad	$4.633 - 0.054*X_1 - 0.067*X_2 + 0.112*X_4$	0.4512	0.40354	0.10252	0.27720
Muzaffarnagar	$4.544 - 0.031*X_1 - 0.107*X_2 + 0.03*X_4$	0.3622	0.32394	0.07367	0.25552
Saharanpur	$-1.734 + 0.042*X_1 - 0.08*X_2 + 0.062*X_3 + 0.147*X_4$	0.4913	0.43947	0.12061	0.32819
Varanasi	$5.614 - 0.056*X_1 - 0.129*X_2 + 0.02*X_3 - 0.121*X_4$	0.4291	0.38380	0.12023	0.30289

Where;  $X_1 = T_{max}$ ,  $X_2 = T_{min}$ ,  $X_3 = RH$ ,  $X_4 = RF$ ,  $X_5 = Ph$

**Table 2: Validation of predicted model of wheat yield for different district of UP.**

District	Wheat					
	2018		2019		2020	
	Forecasted	Predicted	Forecasted	Predicted	Forecasted	Predicted
Agra	4.37	3.843	4.04	3.964	4.12	3.787
Aligarh	4.27	3.894	4.12	3.460	4.21	3.786
Bulandshahr	4.53	4.552	4.38	4.485	4.62	4.484
Etawah	4.37	2.687	3.11	2.525	4.21	2.754
Farrukhabad	4.28	3.968	4.21	4.305	4.55	4.103
Fatehpur	3.99	2.480	3.96	2.683	4.04	2.768
Gorakhpur	4.02	3.848	3.70	2.909	4.65	3.150
Kanpur	4.17	3.293	4.07	3.911	4.42	3.504
Meerut	4.66	3.469	4.56	3.570	4.53	3.549
Moradabad	4.00	2.669	3.77	2.819	3.02	2.571
Muzaffarnagar	4.25	4.478	3.98	4.663	4.21	4.082
Saharanpur	4.22	4.368	3.89	4.712	3.21	3.07
Varanasi	3.5	2.587	3.3	2.953	3.24	3.0

**Table 3: MLR Result for the district Muzaffarnagar.**

Sr. no.	Model	P-value
1.	Constant	0.129
	Maximum Temperature ( $T_{max}$ )	0.586
	Minimum Temperature ( $T_{min}$ )	0.123
	Relative Humidity (RH)	0.158
	Rainfall (RF)	0.111
	Ph	0.915
2.	Constant	0.114
	Maximum Temperature ( $T_{max}$ )	0.580
	Minimum Temperature ( $T_{min}$ )	0.111
	Relative Humidity (RH)	0.148
	Rainfall (RF)	0.105
3.	(Constant)	0.001
	Minimum Temperature ( $T_{min}$ )	0.100
	Relative Humidity (RH)	0.148
	Rainfall (RF)	0.093
4.	(Constant)	0.000
	Minimum Temperature ( $T_{min}$ )	0.012
	Rainfall (RF)	0.361
5.	(Constant)	0.000
	Minimum Temperature ( $T_{min}$ )	0.016*

**Table 4: MLR Result for the district Aligarh.**

Sr. no.	Model	P-value
1.	(Constant)	0.375
	Maximum Temperature ( $T_{max}$ )	0.637
	Minimum Temperature ( $T_{min}$ )	0.559
	Relative Humidity (RH)	0.186
	Rainfall (RF)	0.518
	Ph	0.937
2.	(Constant)	0.359
	Maximum Temperature ( $T_{max}$ )	0.635
	Minimum Temperature ( $T_{min}$ )	0.554
	Relative Humidity (RH)	0.165
	Rainfall (RF)	0.105
3.	(Constant)	0.074
	Minimum Temperature ( $T_{min}$ )	0.591
	Relative Humidity (RH)	0.153
	Rainfall (RF)	0.093
4.	(Constant)	0.052
	Relative Humidity (RH)	0.056
	Rainfall (RF)	0.424
5.	(Constant)	0.020
	Relative Humidity (RH)	0.050*

The maximum and minimum temperatures that prevailed during the cropping season have a significant impact on this region's wheat growing belts. The result revealed that agrometeorological yield model explained the yield variability due to variations in temperatures, rainfall and relative humidity during the different stages (tillering, panicle initiation, booting and physiological maturity). Maximum and minimum temperatures were found common agrometeorological indices for most of the districts of this region.

However, rainfall with relative humidity is also proved important agrometeorological indices for some of the districts of Uttar Pradesh (Naushad Khan *et al.*, 2020).

Table 3 shows the result of MLR by Backward Method in SPSS. It is found to be that the Minimum temperature is significant at 5% level and hence the

temperature plays a significant role in the yield of wheat in Muzaffarnagar district.

Table 4 shows the result of MLR by Backward Method in SPSS. It is found to be that the Relative Humidity is significant at 5% level and hence the temperature plays a significant role in the yield of wheat in Aligarh district.

### CONCLUSION

For thirteen districts in Uttar Pradesh, yield forecasts have been made for the wheat crops and it is found to be that the MLR model is best fit to the district Muzaffarnagar because it has less MAPE value (0.16), RMSE value (0.55) & MAE value (0.45) and also by seeing the result of Backward method in SPSS, Minimum temperature and Relative humidity influences Wheat yield in the district Muzaffarnagar

and Aligarh respectively. The MLR model are also validated because the predicted yield was closed to observed yield. As a result, it could be used to predict wheat yield for all thirteen districts of Uttar Pradesh and found that Multiple Linear Regression is found to be the best fitted model.

### FUTURE SCOPE

For quick decision-making, it is crucial for decision-makers at the national and regional levels to predict crop yields. Farmers may choose what to produce and when to grow it with the help of an accurate crop production prediction model.

**Conflict of Interest.** The authors affirm that they do not have any conflicts of interest.

**Acknowledgments.** We acknowledge the Directorate of economics and statistics and NASA research center for Solar and meteorological data.

### REFERENCES

Agrawal, R., & Mehta, S. C. (2007). Weather based forecasting of crop yields, pests and diseases-IASRI models. *J. Ind. Soc. Agril. Statist*, 61(2), 255-263.

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

Kumar, A., Kumar, N., & Vats, V. (2018). Efficient crop yield prediction using machine learning algorithms. *International Research Journal of Engineering and Technology*, 5(06), 3151-3159.

Lal, M., Singh, K. K., Rathore, L. S., Srinivasan, G., & Saseednran, S. A. (1998). Vulnerability of rice and wheat yields in NW India to future changes in climate. *Agricultural and forest meteorology*, 89(2), 101-114.

Naushad Khan, Ajay Kumar, Vijay Dubey, C.B. Singh, Sanjeev Kumar and Shubham Singh (2020). Estimated Yield Forecasting of Rice and Wheat for Central Uttar Pradesh using Statistical Modal. *Int. J. Curr. Microbial. App. Sci.*, 9(8): 2490-2495.

Singh, H., Singh, K. N., Kasan, B., & Khan, A. A. (2010). Agroclimatic models for prediction of growth and yield of rice (*Oryza sativa*) under temperate Kashmir conditions. *Indian Journal of Agricultural Sciences*, 80(3), 254-257.

Saxena, A., Singh, S. P., Rani, M., Kumar, M., Parmar, V., Shekhar, S., & Gogna, A. A Study of Crop Yield pattern with Climate Change based on Physical Parameters: Temperature and Rainfall in Western Uttar Pradesh to make future predictions for better Crop Management and Yield.

Singh, K., Sharma, S. N., & Sharma, Y. (2011). Effect of high temperature on yield attributing traits in bread wheat. *Bangladesh Journal of Agricultural Research*, 36(3), 415-426.

Timbadia, D. H., Sudhanvan, S., Shah, P. J., & Agrawal, S. (2021, April). Crop Yield Prediction for India Using Regression Algorithms. In *International Conference on Advances in Computing and Data Sciences* (pp. 241-251). Springer, Cham.

**How to cite this article:** Nitesh Kumar Yadav\*, R. Vasanthi, R. Pangayar Selvi and Tmt. G. Vanitha (2022). Prediction of Wheat Yield in Uttar Pradesh using Multiple Linear Regression Approach. *Biological Forum – An International Journal*, 14(2a): 266-270.