Annals of Agri-Bio Research **28** (1): 136-139, 2023

Impact Score : 0.28 (Scopus)

Forecasting of Cotton Yield Using Weather Variables in Haryana

ALISHA MITTAL*, URMIL VERMA AND MANOJ KUMAR

Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar-125 004 (Haryana), India
**(e-mail: alishamittal602@hau.ac.in; Mobile : 89501 38848)*

(Received: June 20, 2022; Accepted: July 21, 2022)

ABSTRACT

Multiple linear regression models incorporating weather variables each spanning successive fortnights within the growth period of the cotton crop and crop condition term were developed to predict the cotton yields in four cotton growing districts, namely, Hisar, Fatehabad, Sirsa and Bhiwani of the Haryana state. Although, the weather variables were found statistically significant as predictors and gave predictions with reasonably high coefficients of determination (R^2) but the predictions had too high per cent deviations to be acceptable and hence were deemed unsuitable for routine crop yield forecasting. To improve the predictive accuracy, a dummy variable in the form of Crop Condition Term (CCT) was added to the weather models. Addition of CCT to the weather models significantly improved the accuracies of the cotton yield predictions.

Key words: Crop condition term, Durbin-Watson d-test, per cent deviation, weather variables

INTRODUCTION

Forecasting of crop yield is a formidable challenge. Crop yield models are abstract presentation of interaction of the crop with its environment and can range from simple correlation of yield with a finite number of variables to the complex statistical models with predictive end. India has a well established system for collecting agricultural statistics. The primary responsibility for collection of data regarding the area under different crops and production of crops is that of the State Governments. The official forecasts (advance estimates) of major cereal and commercial crops are issued by the Directorate of Economics and Statistics, Ministry of Agriculture, New Delhi. However, the final estimates are given a few months after the actual harvest of the crop. There is always a considerable scope of improvement in the conventional system. In many previous studies, yield forecasting models have incorporated a series of weather predictors (Gurung *et al.*, 2017; Panwar *et al.*, 2018; Singh *et al.*, 2019). Models developed by Shastry *et al.* (2017), Kamath *et al.* (2021) and Mittal and Mishra (2021) were successfully used for forecasting yields of various crops at district as well as agro-climatic zone level in different States of India. Verma *et al.* (2016), Goyal (2018), Das *et al.* (2018), Abrougui *et al.* (2019), Mamata *et al.* (2020) and Nihar *et al.* (2022) have also developed models for crop yield prediction.

MATERIALS AND METHODS

The Haryana state comprising 22 districts is situated between 74°25′ E to 77°38′ E longitude and 27°40′ N to 30°55′ N latitude. The districts Hisar, Fatehabad, Sirsa and Bhiwani comprising the western zone of the state were considered for the model building. The Department of Agriculture (DOA) cotton yield estimates for the period 1980-81 to 2017-18 in respect of the four districts were collected from Statistical Abstracts of Haryana. The daily weather data on maximum temperature (TMAX), minimum temperature (TMIN), rainfall (RF), sunshine hours (SSH) and relative humidity (RH) were collected for the same period. Weather data starting from 1st fortnight of May to 1 month before harvest (i.e. 11 fortnights) were utilized for the model building (crop growth period: May to October/ November). The time-series yield/weather data from 1980-81 to 2010-11 for Hisar, Sirsa and Bhiwani districts and 1997-98 to 2010-11 of Fatehabad district were used for the training set and the remaining data i.e. 2011-12 to 2017-18 were used for the post-sample validity checking of the models.

The multiple linear regression model was used to relate crop yield to the average maximum temperature, average minimum temperature, average relative humidity, average sunshine hours and accumulated rainfall calculated for 11 fortnights covering the period May to October/ November and CCT was included as an additional regressor (dummy) variable in extending the weather models. The standard linear regression model considered may be written in the form below:

$$Y = Xb+\varepsilon$$

Where, Y is an (n × l) vector of observations (DOA yields), X is an (n × p) matrix of known form (weather variables and CCT), b is (p × l) vector of parameters, ε is an (n × l) vector of errors and E (ε) = 0, V (ε) = $I\sigma^2$ so the elements of ε are uncorrelated.

The best subsets of weather variables were selected using a stepwise regression method. The predictive performance of the cotton yield equations was compared on the basis of adj-R^2, per cent deviations of yield estimates from the real-time yields and root mean square errors (RMSEs).

For detecting serial correlation, the Durbin-Watson *d* statistic was used, which is defined as:

$$d = \frac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n} \sum \hat{u}_t^2}$$

It is simply the ratio of the sum of squared differences in successive residuals to the RSS. Note that in the numerator of the d statistic the number of observations is *n-1* because one observation is lost in taking successive differences.

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

Since $\sum \hat{u}_t^2$ and $\sum \hat{u}_{t-1}^2$ differ in only one observation, they are approximately equal. Therefore, setting.

$$\sum \hat{u}_{t-1}^2 \approx \sum \hat{u}_t^2$$

$$d \approx 2\left(1 - \frac{\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2}\right)$$

Now taking $\hat{\rho} = \frac{\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2}$

As the sample first-order coefficient of autocorrelation, an estimator of $\hat{\rho}$.

Thus, $d \approx 2(1 - \hat{\rho})$

But since $-1 \le \rho \le 1$, therefore, $0 \le d \le 4$.

If $\hat{\rho}$. = *0, d = 2*; that is, if there is no serial correlation (of the first- order), *d* is expected to be about 2.

RESULTS AND DISCUSSION

Department of Agriculture yield data of cotton for Hisar, Fatehabad, Sirsa and Bhiwani districts were regressed against the cropping year considered as an independent variable along with the weather variables and crop condition term as dependent variables. The multiple linear regression models with alternative combinations of weather variables were tried for fitting the accurate models. Following the steps required in SPSS; Model 1, Model 2 and Model 3 for Hisar, Fatehabad, Sirsa and Bhiwani districts with fortnightly weather variables (viz., RH3, RH6, RH10, RF1, RF10, TMAX1, TMAX6, TMIN6, SSH4 and SSH10) and CCT as independent variables were tried for the same and Durbin-Watson d-statistic (i.e. DW=1.5 to 2.5) indicated that there was no serial auto correlation, either positive or

Table 1. Multiple linear regression models for cotton yield (kg/ha) of Hisar, Fatehabad, Sirsa and Bhiwani districts

Districts	Model(s)	Yield forecast models	R^2	Standard error	Durbin-Watson statistic
Hisar	1	Y=41.67+190.73 × CCT+0.43 × RF10	.82	51.45	1.96
	2	Y=390.81+204.79 × CCT+0.51 × RF10-9.56 × TMAX1	.86	46.29	2.17
	3	Y=321.91+206.54 × CCT+0.54 × RF10-10.54 × TMAX1+2.34 × RH3	.87	42.54	2.08
Fatehabad	1	Y= -503.67+220.99 × CCT-0.24 × RF1+14.58 × TMAX6	.86	57.96	1.80
	2	Y= -494.81+222.89 × CCT+14.67 × TMAX6	.87	56..15	1.79
	3	Y= -718.55+219.14 × CCT+28.02 × TMIN6	.86	57.03	1.41
Sirsa	1	Y= 84.70+199.72 × CCT-5.55 × SSH4	.85	57.64	1.53
	2	Y= -88.73+198.09 × CCT+2.12 × RH10	.86	56.54	1.64
	3	Y= 155.00+200.76 × CCT-12.88 × SSH10	.85	57.23	1.68
Bhiwani	1	Y=335.50+148.32 × CCT-25.05 × SSH10	.75	44.42	2.15
	2	Y=117.88+145.44 × CCT+3.71 × RH10	.79	41.91	2.29
	3	Y=196.36+149.43 × CCT-24.36 × SSH10+1.86 × RH6	.77	42.84	2.32

negative i.e. residuals were not auto correlated (Table 1). Cotton yield forecasts for the years 2011-12, 2012-13, 2013-14, 2014-15, 2015-16, 2016-17 and 2017-18 were obtained on the basis of Model 1, Model 2 and Model 3 for all four districts. The per cent deviations of forecast yield(s) from observed yield(s) based on the three regression models are shown in Tables 2, 3, 4 and 5. The performances of the contending models were compared on the basis of different statistics i.e. adjusted R^2 and RMSE as shown in Table 6. The results indicated the preference of using Model 3 for Hisar, Model 2 for Fatehabad, Sirsa and Bhiwani districts as multiple linear regression models for cotton yield forecasting.

CONCLUSION

This study had demonstrated the utility of understanding and quantifying the relationships between cotton yield and weather variables. The relationships can be employed in studies that explore the impact of climate change on probable future crop yields at regional scales. In particular, adding the crop condition term to the weather yield models substantially improved the predictive accuracy of the model and produced what was considered

Table 2. Comparative view in terms of per cent deviations of forecast yield(s) from observed yield(s) based on regression models for Hisar district

Year	Observed yield (kg/ha)	Model 1		Model 2		Model 3	
		Estimated yield (kg/ha)	RD (%)	Estimated yield (kg/ha)	RD (%)	Estimated yield (kg/ha)	RD (%)
2011	734.63	638.76	13.05	645.63	12.11	660.12	10.14
2012	608.28	617.92	-1.60	647.85	-6.52	636.33	-4.63
2013	501.39	482.93	3.67	496.35	0.98	528.69	-5.46
2014	368.23	423.20	-14.94	444.23	-20.65	442.40	-20.15
2015	276.02	240.86	12.73	227.08	17.72	246.33	10.75
2016	623.13	615.16	1.26	617.56	0.87	633.36	-1.66
2017	489.02	423.20	13.46	407.42	16.69	421.61	13.78

Table 3. Comparative view in terms of per cent deviations of forecast yield(s) from observed yield(s) based on regression models for Fatehabad district

Year	Observed yield (kg/ha)	Model 1		Model 2		Model 3	
		Estimated yield (kg/ha)	RD (%)	Estimated yield (kg/ha)	RD (%)	Estimated yield (kg/ha)	RD (%)
2011	818.86	682.21	16.68	686.95	16.10	681.46	16.77
2012	715.13	697.65	2.44	703.39	1.64	719.57	-0.62
2013	766.09	682.69	10.88	682.28	10.93	688.46	10.12
2014	653.51	665.83	-1.89	678.17	-3.78	701.91	-7.41
2015	209.04	226.71	-9.52	222.91	-7.68	239.81	-15.85
2016	686.03	661.11	3.63	661.31	3.60	651.75	4.99
2017	454.11	439.64	3.19	438.28	3.49	467.92	-3.04

Table 4. Comparative view in terms of per cent deviations of forecast yield(s) from observed yield(s) based on regression models for Sirsa district

Year	Observed yield (kg/ha)	Model 1		Model 2		Model 3	
		Estimated yield (kg/ha)	RD (%)	Estimated yield (kg/ha)	RD (%)	Estimated yield (kg/ha)	RD (%)
2011	769.83	659.45	14.34	664.20	13.72	640.26	16.83
2012	767.61	641.47	16.43	648.35	15.54	648.75	15.48
2013	697.16	638.82	8.37	666.60	4.38	667.29	4.28
2014	619.21	640.41	-3.42	627.98	-1.42	629.31	-1.63
2015	295.23	235.33	20.29	249.27	15.57	265.26	10.15
2016	694.14	639.94	7.81	652.38	6.02	649.53	6.43
2017	321.13	245.03	23.70	237.53	26.03	269.51	16.08

Table 5. Comparative view in terms of per cent deviations of forecast yield(s) from observed yield(s) based on regression models for Bhiwani district

Year	Observed yield (kg/ha)	Model 1		Model 2		Model 3	
		Estimated yield (kg/ha)	RD (%)	Estimated yield (kg/ha)	RD (%)	Estimated yield (kg/ha)	RD (%)
2011	613.88	552.80	9.93	596.37	2.84	554.46	9.67
2012	625.93	569.33	9.03	568.61	9.15	559.66	10.58
2013	523.92	457.07	12.76	455.13	13.12	462.30	11.76
2014	441.18	383.19	13.13	387.48	12.16	390.21	11.54
2015	334.24	307.75	7.86	272.63	18.37	313.65	6.10
2016	518.16	422.60	18.43	430.23	16.94	443.46	14.39
2017	402.34	464.34	-15.41	397.52	1.20	485.60	-20.69

Table 6. Comparative view in terms of adjusted R^2 and RMSEs of cotton yield forecasts in relation to real time yield(s) based on multiple linear regression models

District		Adj. R^2	RMSEs
Hisar	Model 1	.81	51.10
	Model 2	.84	59.01
	Model 3	.87	50.87
Fatehabad	Model 1	.85	62.37
	Model 2	.86	61.21
	Model 3	.85	65.01
Sirsa	Model 1	.84	79.52
	Model 2	.85	72.94
	Model 3	.84	73.16
Bhiwani	Model 1	.73	63.74
	Model 2	.76	57.02
	Model 3	.75	62.40

to be satisfactory district-level yield(s) estimation. Hence, based on this empirical study, use of CCT as an indicator variable was recommended in addition to the weather variables to enhance the predictive accuracy of the zonal weather models. The predictive performance of the models was assessed using the adjusted R^2, the per cent deviations of the forecast yields from the DOA yield estimates and the RMSEs.

ACKNOWLEDGEMENT

The weather data received from Department of Agrometeorology, CCSHAU, Hisar, India are gratefully acknowledged.

REFERENCES

Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R. and Chehaibi, S. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil Tillage Res.* **190**: 202-208.

Das, B., Nair, B., Reddy, V. K. and Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *Int. J. Biometeorol.* **62**: 1809-1822.

Goyal, M. (2018). Pre-harvest wheat yield forecast through agro-meteorological indices for northern region of Haryana. *IJAMSS* **7**: 29-40.

Gurung, B., Panwar, S., Singh, K. N., Banerjee, R., Gurung, S. R. and Rathore, A. (2017). Wheat yield forecast using detrended yield over a sub-humid climatic environment in five districts of Uttar Pradesh, India. *Indian J. agric. Sci.* **87**: 87-91.

Kamath, P., Patil, P., Shrilatha S., Sushma and Sowmya, S. (2021). Crop yield forecasting using data mining. *Global Transitions Proc.* **2**: 402-407.

Mamata, Mittal, A. and Bharti, D. (2020). Forecasting of barley production in India using ARIMA model. *J. Pharmacogn. Phytochem.* **9**: 1193-1196.

Mittal, A. and Mishra B. (2021). Cotton yield forecasting in districts of Haryana using fortnightly weather variables: A time series approach. *J. Cotton Res. Dev.* **35**: 280-285.

Nihar, A., Patel, N. R. and Danodia (2022). A. machine-learning-based regional yield forecasting for sugarcane crop in Uttar Pradesh, India. *J. Indian Soc. Remote Sens.* **50**: 1519-1530.

Panwar, S., Kumar, A., Singh, K. N., Paul, R. K., Gurung, B., Ranjan, R., Alam, N. M. and Rathore, A. (2018). Forecasting of crop yield using weather parameters–Two-step non-linear regression model approach. *Indian J. agric. Sci.* **88**: 117-119.

Shastry, A., Sanjay, H. A. and Bhanusree, E. (2017). Prediction of crop yield using regression techniques. *Int. J. Soft Comput.* **12**: 096-102.

Singh, K. N., Singh, K. K., Kumar, S., Panwar, S. and Gurung, B. (2019). Forecasting crop yield through weather indices through LASSO. *Indian J. agric. Sci.* **89**: 540-544.

Verma, U., Piepho, H. P., Goyal, A., Ogutu, J. O. and Kalubarme, M. H. (2016). Role of climatic variables and crop condition term for mustard yield prediction in Haryana (India). *Int. J. Agric. Stat. Sci.* **12**: 45-51.