Plant Archives Vol. 20, Special Issue (AIAAS-2020), 2020 pp. 510-515 ***ISSN 0972- 5210***

ROLE OF CROP CONDITION TERM AS AN INDICATOR VARIABLE ALONGWITH WEATHER PARAMETERS TO FORECAST COTTON YIELD IN HARYANA

Ajay Kumar[1,*], Aditi[2], Urmil Verma[2] and Salinder Malik[3]

[1]Department of Mathematics, Guru Jambheshwar University of Science and Technology, Hisar

[2]Department of Mathematics and Statistics, CCS Haryana Agricultural University, Hisar

[3]Department of Agriculture and Farmers Welfare, Haryana

Abstract

The purpose of this study is to forecast cotton yield based on weather-yield models incorporating crop condition term as an indicator variable using multiple linear regression analysis. The models developed for Hisar, Sirsa, Fatehabad and Bhiwani districts of Haryana are based on weather parameters *viz*., maximum temperature, minimum temperature, sunshine hours, relative humidity and rainfall. The predictive performance of the fitted models have been assessed on the basis of per cent relative deviations of the predicted yield(s) from the Department of Agriculture yield estimates and root mean square error(s). Inclusion of crop condition term as an indicator variable in the weather-yield models significantly improved the predictive accuracy of cotton yield in Hisar, Sirsa, Fatehabad and Bhiwani districts of Haryana. The percent relative deviations for post-sample (2018-19 & 2019-20) validity checking of district-level cotton yield prediction in most of the time-regimes were observed between 5-12% except Bhiwani district in 2018-19.

***Keywords*:** Crop Condition Term (CCT), Weather Variables, Stepwise Regression, Adjusted R^2, RMSE and Cotton Yield Prediction.

Introduction

Regression analysis is used for the estimating the relationships between a dependent variable and one or more independent variables. It measures the strength of relationship between variables and helps in finding the future values of the response variable. The process of performing a regression analysis allows to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other.

Cotton is one of the finest natural fibers available to mankind for clothing from time immemorial. Cotton is an important commercial crop of the country and contributing nearly 85% of the total domestic fiber consumption. India's textile sector is one of the oldest industries in Indian economy dating back to several centuries. The Indian textile industries consume a diverse range of fibers and yarns. It is estimated that cotton requirement in India by 2025 will be around 140 lakh bales of lint and the present production is around 123 lakh bales. It has also the distinction of having largest area under cotton cultivation in the world constituting about 36 percent of world area under cotton cultivation. The lint productivity of cotton is 524 kg/ha, which is the lowest and far below that of the world average of 765 kg/ha. Haryana is the fifth largest producer of cotton in India. The cotton crop is grown during *kharif* season and generally sown during May-June and is harvested during early winter (Oct.-Nov.). Cotton is the crop of tropical and sub-tropical warm humid climate. Annual temperature requirement is 20^0C to 28^0C. Equitable temperature distribution and bright sunshine is desirable. Loamy soil mixed with lime and potash is good for cotton. It grows well in black soil which is sticky in nature and has water retentive capacity. The modest requirement of water can be met by an average rainfall of 50-100 cm. However, it is successfully grown in the area of lesser rainfall with the help of irrigation.

Sharma *et al.* (1992) tried to estimate cotton acreage and condition assessment using digital data from IRS-1A. LISS-1 and the stratified random selection strategy during *kharif* 1990-91 in Hisar and Sirsa districts. Subsequently, the efforts were made to improve the accuracy and timeliness of this process by modifying the stratification and sampling procedure by Yadav *et al.* (1994). To mention a few more work specifically on cotton crop; Ray *et al.* (1994, 1999), Dubey *et al.* (1995, 1997) and Kalubarme and Dubey (1999) etc. had attempted good efforts in this direction. Viator *et al.* (2005) have worked to observe the effect of climatic factors on cotton boll formulation. Nair *et al.* (2012) estimated the impact of adoption of variable rate technology on cotton yield in Texas and used dummy variable that assumes the value 1 if a farmer adopts VRT for application of any input and 0 otherwise. Verma *et al*. (2014) developed models to predict cotton yields in five cotton growing districts; Hisar, Sirsa, Bhiwani, Rohtak and Jind covering more than 90% of cotton production of Haryana State. Incorporating Crop Condition Term in weather models improved the accuracy of yield prediction in the State.

*Corresponding author Email: ajaystatistics@gmail.com

Koukouli *et al.* (2018) described an assessment of simulated potential cotton yield using simulation model in Northern Greece. The results indicated that alteration of crop management practices such as changing the planting date could be used as potential adaptation measures to address the impact of climate change on cotton production.

The Haryana state comprises of 22 districts and is having total geographical area of 44,212 sq. km. The major cotton growing districts in the state *viz*., Hisar, Sirsa, Bhiwani and Fatehabad covering more than 85% of the crop, have been considered for the model building. The state Department of Agriculture (DOA) cotton yield estimates for the period 1981-82 to 2019-20 collected from Department of Agriculture and Farmers Welfare, Haryana have been used. The daily weather data on maximum temperature, minimum temperature, rainfall, sunshine hours and relative humidity of Hisar district were collected for the same period. Weather data starting from 1st fortnight of May to 1 month before harvest (*i.e.*10 fortnights) were utilized for the model building (crop growth period: May to October/November). The time-series yield and weather data from 1981-82 to 2017-18 have been used for the training set and the remaining data *i.e.*, 2018-19 and 2019-20 are used for the post-sample validity checking of the developed models.

The fortnightly weather data were prepared from daily weather data as given below:

Average Maximum Temperature (Tmx) = $\frac{\sum_{i=1}^{15} Tmx_i}{15}$

Average Minimum Temperature (Tmn) = $\frac{\sum_{j=1}^{15} Tmn_j}{15}$

Accumulated Rainfall (Arf) = $\sum_{k=1}^{15} Arf_k$

Average Relative Humidity (Rh) = $\frac{\sum_{l=1}^{15} Rh_l}{15}$

Average Sunshine Hours (Ssh) = $\frac{\sum_{m=1}^{15} Ssh_m}{15}$

Where,

Tmx_i = i^{th} day maximum temperature

Tmn_j = j^{th} day minimum temperature

Arf_k = k^{th} day rainfall

Rh_l = l^{th} day relative humidity

Ssh_m = m^{th} day sunshine hours

i,j,k,l,m = 1,2,..........,15 (daily weather data)

Crop productivity is affected by technological change and weather variability. It can be assumed that the technological factors will increase crop yield smoothly through time and therefore, year or some other parameter of time can be used to study the overall effect of technology on yield. The linear time-trend forecast model(s) fitted for all the districts may be expressed as T_r = a+bt, where T_r = Yield(kg/ha), a = Intercept, b = Slope and t= Year. Predictions T_r based on this model yielded a predictor variable that has been denoted as 'trend yield'. Trend yield may be considered as an indication of technological advancement, qualitative/ quantitative changes in fertilizer/ insecticide/ pesticide/ weedicide use and increased use of high yielding varieties over time.

Weather variability both within and between seasons is major uncontrollable source of variability in yield. Weather variables affect the crop differentially during different stages of development. This increases the number of variables in the model and in turn, a large number of parameters are to be evaluated from the time series data for precise estimation. Thus, a technique based on relatively smaller number of manageable parameters and at the same time, taking care of entire weather distribution is always preferred to solve the problem.

For quantitative forecasting, the regression models have been fitted by taking weather variables and trend yield as regressors and crop yield as the regressand. The multiple linear regression model considered may be expressed as follows:

$$Y = a + cT_r + \sum_{i=1}^{10} b_i\, Tmx_i + \sum_{j=1}^{10} b_j\, Tmn_j + \sum_{k=1}^{10} b_k\, Arf_k + \sum_{l=1}^{10} b_l\, Rh_l + \sum_{m=1}^{10} b_m\, Ssh_m + e$$

Where,

Y - Cotton yield (kg/ha)

T_r - Trend yield (kg/ha)

a- Overall mean effect

c - Regression coefficient of trend yield

b_i, b_j, b_k, b_l, b_m - Regression coefficients of weather variables shown above

(i, j, k, l, m - fortnights1,2,3…10 over crop growth period)

e - Error termwith assumption NID $(0, \sigma^2)$

The multiple linear regression analysis was carried out for the development of weather-yield models. The best subsets of weather variables have been selected using the stepwise regression method (Draper and Smith, 2003) in which all the variables are first included in the model and eliminated one at a time with decisions at any particular step conditioned by the results of previous step. The best supported weather variables in the model are retained if they had the highest adjusted R^2 and the lowest standard error of estimate at a given stage. Once a regression model has been constructed, it may be important to confirm the goodness of fit of the model and the statistical significance of estimated parameters. Commonly used checks of goodness of fit include R^2, analysis of the pattern of residuals and hypothesis testing. Statistical significance is checked by an F-test of the overall fit, followed by t-test of individual parameters. The performance of fitted weather-yield models were compared on the basis of per cent relative deviations and root mean square errors to obtain pre-harvest mustard yield forecasts.

To further enhance the predictive performance, the weather-yield models were again fitted by taking crop condition term (CCT) as categorical/dummy variable along with weather variables as regressors and DOA mustard yield as regressand. The CCT being an indicator variable is generated by splitting the trend yield data into different non-overlapping classes.

Results and Discussion

Time-trend analysis often reflects an underlying pattern/ behaviour in a time series which would otherwise be partly or nearly completely hidden by noise. The DOA cotton yield data were used by considering time (year) as an independent variable and regressed against yield to get the trend equations of the form as have been shown in Figure-1. The following time versus yield graphs show the overall increasing trend for cotton crop in all the districts.

Next, the weather-yield models were developed by stepwise regression method using trend-based yield and weather parameters (rainfall, minimum temperature, maximum temperature, sunshine hours and relative humidity) computed over different fortnights of crop growth period, as regressors. The suitable weather-yield models bearing highest adjusted R^2 and lowest standard error of estimate are shown in Table-1. The yield forecasts based on weather-yield models-1 and 1* had higher percent relative deviations from real-time cotton yield(s) in most of the districts, sometimes even too high than to be acceptable. Thus, the above weather-yield models were found unsuitable for operational yield forecasting purpose *i.e.* adding linear time-trend to the model, along with selected weather variables couldn't significantly improve the forecast accuracy of cotton yield.

Further, an attempt was made to improve the predictive accuracy of weather-yield models by identifying and adding additional covariate CCT/indicator variable to the model. The CCT being a categorical covariate has been obtained by dividing the trend based yield series into three non-overlapping classes corresponding to adverse, normal and high yield(s). Again, following the stepwise regression analysis, the models-2 and 2* were selected on the basis of highest Adj. R^2 and lowest standard error of estimate as have been shown in the Table-1.

The forecast performance(s) of the above weather-yield models have been observed in terms of per cent relative deviations of yield forecasts from the real time yield(s). Model predicted yield(s) of all the districts along with observed yield(s) and percent relative deviations are given in Table-2.

Percent Relative Deviation = 100 (observed yield – estimated yield) / observed yield); measures the deviation (in percentage) of forecast yield from the actual yield

Thus, incorporating CCT as an indicator variable along with weather parameters enhanced the predictive accuracy of weather-yield models inspite of showing lower adj. R^2 and higher standard error of estimate.

Regression diagnostics of the fitted models

Residual Diagnostics are concerned with testing the goodness of fit of a model and suggesting appropriate modifications if required. Thus, residual histogram and normal-probability plots for the fitted model were prepared for examining normality assumptions of the residuals. Histograms show approximate behaviour with slight deviation from normality. The P-P plots also infer the same. Standardized residual plots appear fine. On the whole, the following plots (Figure-2) do not exhibit serious violations of the model assumptions. Thus, on the basis of this empirical work, using CCT as an indicator variable is recommended in addition to the weather parameters to enhance the predictive accuracy of the weather–yield models for district-level cotton yield forecast in Haryana.

Acknowledgement

The yield data and weather data received from Department of Agriculture and Farmers Welfare, Haryana and Department of Agro-meteorology, CCS HAU, Hisar, India are gratefully acknowledged.

References

Dubey RC, Chaudhary A and Kale JD (1995). The estimation of cotton yield based on weather parameters in Maharashtra. *Mausam,* **46(3):** 275-278.

Dubey RP and Sehgal VK (1997). Development of weather based zonal models for cotton crop in India, *Paper presented at INTROMET (International Symposium on Advances in Tropical Meteorology) under theme, "Monsoon and Management of Agricultural activities, IIT, New Delhi,* 2-5.

Kalubarme MH and Dubey RP (1999). State-level agro-meteorological models for national cotton production forecast. *Scientific Note: RSAM/SAC/FASAL-TD/SN/05/99, Space Applications Centre, Ahmedabad.*

Koukouli P and Georgiou P (2018). Evaluation of climate change impacts on cotton yield using crop system and regression models. *Journal of Advances in Agriculture*, **8**: 1433-1450.

Nair S, Wang C, Segarra E, Johnson J and Rejesus R (2012). Variable rate technology and cotton yield response in Texas. *Journal of Agricultural Science and Technology*, **3**: 1034-1043.

Ray SS, Pokharna SS and Ajai (1994). Cotton production estimation using IRS-1B and meteorological data. *International Journal Remote Sensing*, **15(5):** 1085-1090.

Ray SS, Pokharna SS and Ajai (1999). Cotton yield estimation using agro-meteorological model and satellite-derived profile. *International Journal Remote Sensing*, **20(14):** 2693-2702.

Sharma SA, Ajai, Hooda RS, Mothikumar KE, Yadav M and Manchanda ML (1992). Cotton acreage and condition assessment for Hisar and Sirsa districts of Haryana (1990-91). *Scientific Note*, RSAM/SAC/CACA/SN.

Verma U, Piepho HP, Ogutu JO, Kalubarme MH and Goyal M (2014). Development of agromet models for district-level cotton yield forecasts in Haryana State. *International Journal of Agricultural and Statistical Sciences*, **10(1):** 59-65.

Viator RP, Nuti RC, Keith L, Edmisten and Wells R (2005). Predicting cotton boll maturation period using degree days and other climatic factors. *Agronomy Journal*, **97**: 494-499.

Yadav M, Hooda RS, Mothikumar KE, Ruhal DS, Khera AP, Singh CP, Hooda IS, Verma U, Dutta S and Kalubarme MH (1994). Cotton acreage in Hisar and Sirsa districts of Haryana using remote sensing techniques. *Technical Report*, HARSAC/TR/03.

Table 1: Regression based weather-yield forecast models for Hisar, Sirsa, Fatehabad and Bhiwani districts

Types	Fitted Models	Adj. R^2	SE
Model 1	Y_{est} = -405.38 +1.26 T_r + 29.90Ssh_7 + 4.44Rh_8– 27.26 Ssh_4 + 0.65 Arf_7 - 4.72 Rh_3 - 3.51 Rh_6 + 6.97 Rh_2 + 5.62 Rh_{10} – 4.45 Rh_4	0.652	94.61
Model 1*	Y_{est} = -452.85 +1.19 T_r – 13.48 Tmx_2+ 11.45 Tmx_6 + 40.81Ssh_7 + 4.56Rh_8– 17.15 Ssh_4 + 0.53 Arf_7 - 4.82 Rh_3 + 3.48 Rh_{10}	0.601	101.29
Model 2	Y_{est} = -274.16 + 135.37 CCT + 3.91Rh_9–16.74 Ssh_4+ 32.38Ssh_7 + 0.62Arf_7-14.72 Tmx_8 + 15.63 Tmx_6	0.513	111.99
Model 2*	Y_{est} = -454.99 + 137.48 CCT + 5.04Rh_9-20.57Ssh_4+ 33.69Ssh_7 + 0.83Arf_7 + 14.22 Tmx_6 -15.32 Tmx_1 + 11.11 Tmn_1	0.511	112.15

Where,
Y_{est}- Model predicted yield (kg/ha)
T_r- Trend yield (kg/ha)
CCT- Crop condition term
R^2- Coefficient of Determination
SE-Standard error of Estimate
Model 1, 1* - Weather parameters and trend yield as regressors
Model 2, 2* - Weather parameters and CCT as regressors

Table 2: Post-sample cotton yield forecasts based on finally selected models for all the districts

District/ Forecast Years Hisar	Observed Yield (kg/ha)	Model 1		Model 2	
		Fitted Yield (kg/ha)	RD (%)	Fitted Yield (kg/ha)	RD (%)
2018-19	501	573.07	-14.39	528.47	-5.48
2019-20	545	567.62	-4.15	515.72	5.37
Sirsa					
2018-19	559	697.31	-24.74	528.47	5.46
2019-20	714	696.29	2.48	651.09	8.81
Fatehabad					
2018-19	591	787.86	-33.31	663.84	-12.33
2019-20	653	794.55	-21.68	651.09	0.29
Bhiwani					
2018-19	321	497.30	-54.92	393.10	-22.46*
2019-20	447	492.86	-10.26	380.34	14.91

*Due to deficient rainfall and whitefly attack

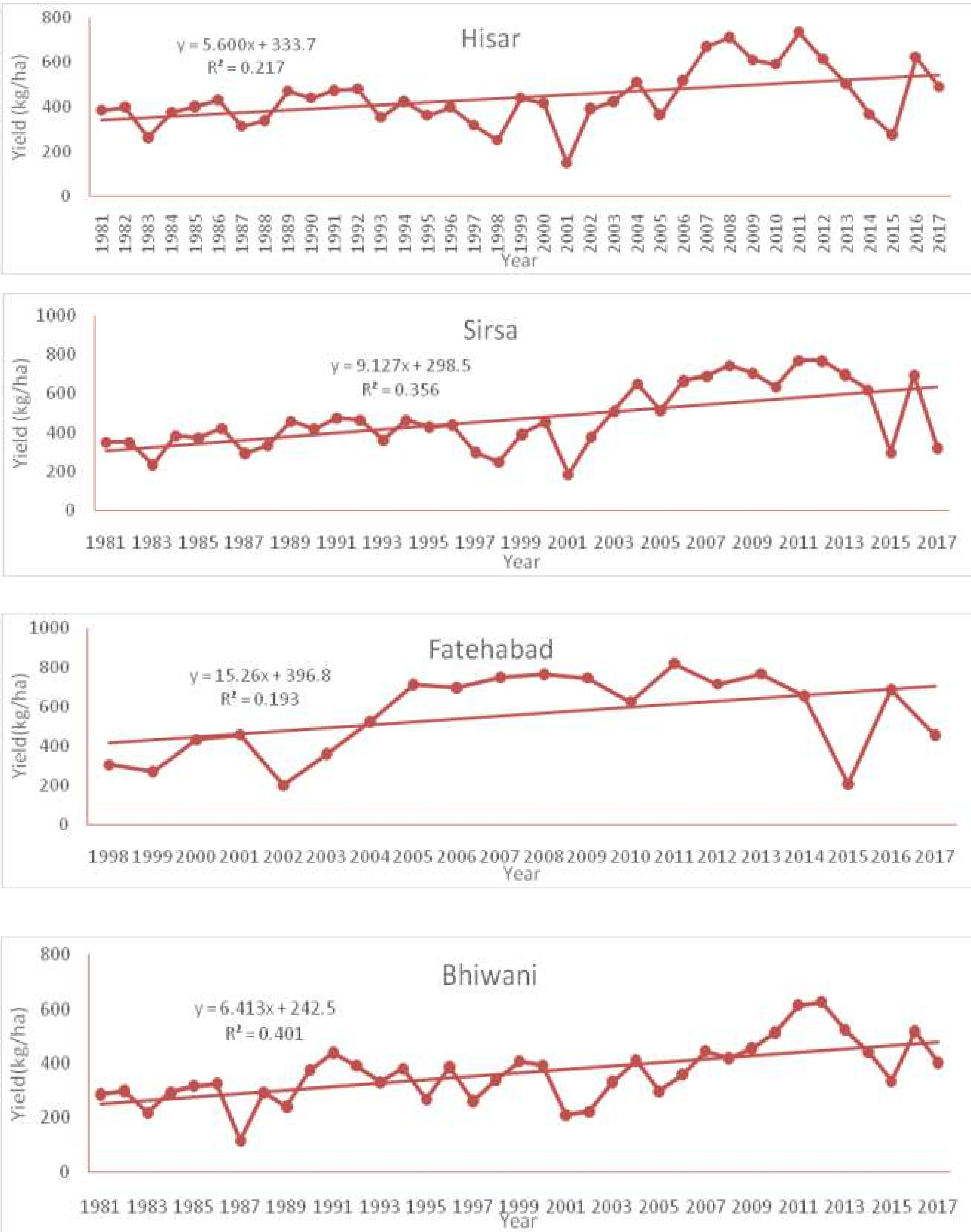

Figure 1: Time versus Yield graph(s) of cotton crop for all the districts

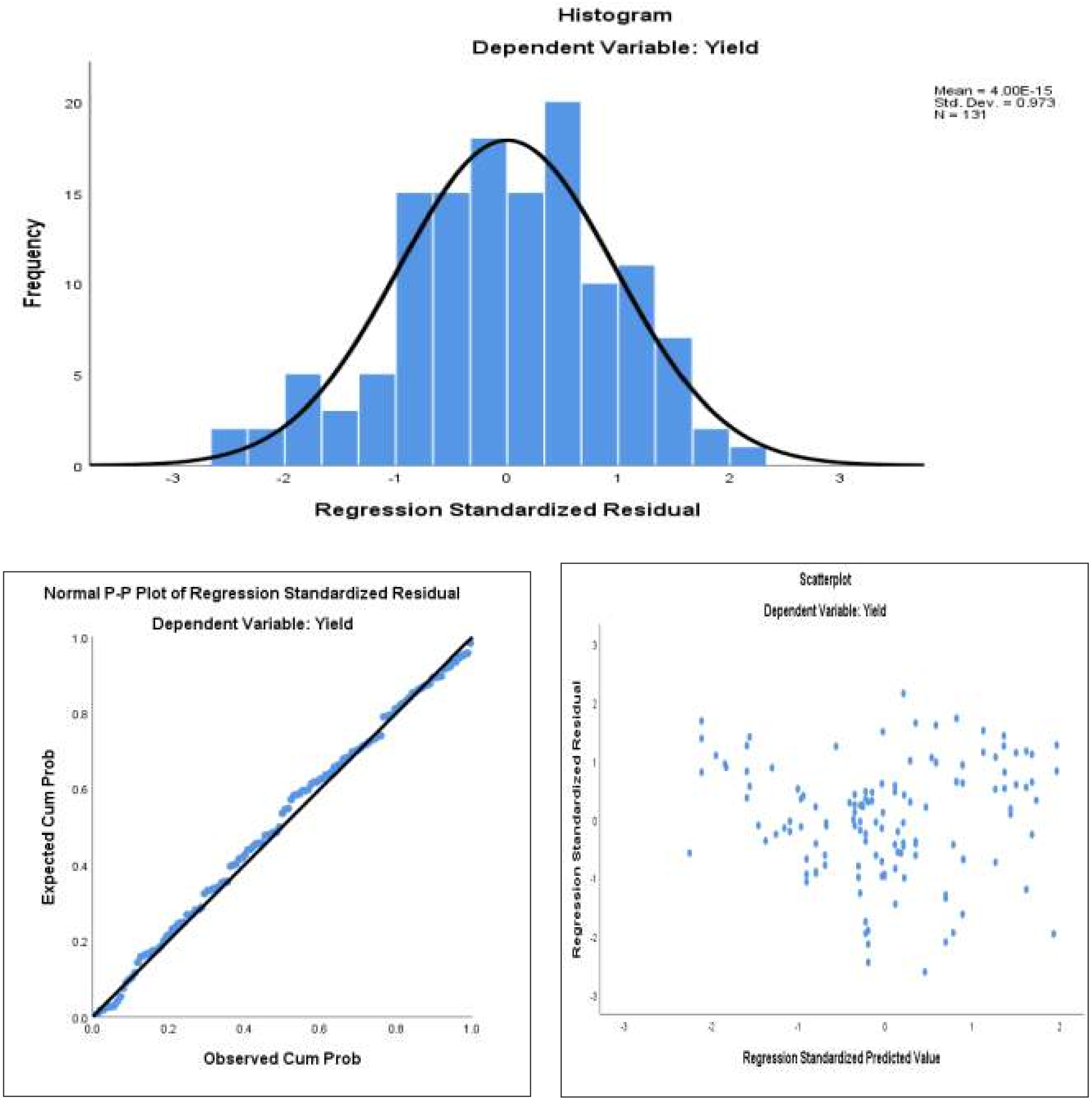

Figure 2: Regression diagnostics of the fitted model (CCT + weather variables)