Vol. 15, No. 2, pp. 157-162 (2015)
Journal of Agricultural Physics
ISSN 0973-032X
http://www.agrophysics.in

Research Article

Wheat Yield Prediction using Weather based Statistical Model in Central Punjab

K.K. GILL, S.S. SANDHU, PRABHJYOT KAUR, RITU BABUTA AND KAVITA BHATT

School of Climate Change and Agricultural Meteorology, Punjab Agricultural University, Ludhiana-141004, Punjab

ABSTRACT

A study was undertaken to investigate the impact of climatic and technological factors on crop productivity of wheat. Four models, Basic, Modified model I (Technology Trend), Modified model II (Composite Index) and SPSS were employed to forecast wheat yields over central Punjab (Ludhiana district) by using weather and yield data of 41 years. The sensitive period of statistical and phenological significance were selected for regression analysis. The regression model explained 63% variation in yield, while the technological trend model improved the prediction considerably (R^2 = 0.89). The regression models were compared to analyze the impact of climate and technology together on wheat productivity, and SPSS model emerged as the best model, which explained 90% variation in wheat yield due to weather. Among other models, exponential function was better than the power function (R^2= 0.87). However, the multiple linear and exponential functions explained similar variation (88-89%) in wheat yield, respectively. The model was validated for five years and performed well for wheat yield prediction in the district.

Key words: Wheat, Weather parameters, Technology, Statistical models and Climate change

Introduction

Agriculture and allied sectors constitute the single largest component of India's economy, but is highly dependent on weather. Average yield of wheat in India is 2.9 t ha^{-1} and in many place, the yield has reached to a plateau. Again, changing global climate is threatening the sustainability on agriculture, and the developing countries like India are more vulnerable.

Weather affects crop growth at different phenological phases and is therefore, responsible for variation in yields from year-to-year and place-to-place. A number of statistical techniques such as multiple regressions, principal component analysis (Jain *et al*., 1984), Markov chain analysis (Ramasubramanian & Jain, 1999) and agro-meteorological models (Walker, 1989) have been used to quantify the response of crops to weather. Multiple regression models were used to forecast crop yields in India (e.g., Appa Rao, 1983; Mahi *et al*., 1991). Time series analysis is also used to analyze yield trends and to predict yields under different scenarios. Kumar (2000) compared the performance of linear and quadratic regression models with those of exponential smoothing and moving averages.

Crop-weather models for forecasting crop parameters in wheat are used in various agro-meteorological regimes (Bal *et al*. 2004). Studies used multiple regression approach (e.g., Kandiannan *et al*., 2002; Sharma *et al.*, 2004) to capture the effect of climatic variability on crop yield. Multiple regression models are not only easy to use, but are also more accurate than the simulation model approach. Predicting grain quality before wheat harvest would be useful for grain buyers, and for farmers to help optimize agronomic activity (Smith and Gooding, 1999). Similarly, crop yield estimates are necessary to meet farmers' needs, land appraisers and government agencies all over the world.

The current study compared the performance of four statistical models, including technology for analyzing yield time series and predicting yields. We chose wheat as the test crop, because wheat is an important cereal and its yield time series show a great diversity of trends (increasing, plateauing or decreasing). We collected weather parameters to offer several multiple regression models for prediction of wheat grain yields. By coupling technology trend with weather variables, models were developed for corn and soybean (Da Mota, 1983) and wheat (Esfandiary *et al*., 2009; Nain *et al*., 2002) crops. If wheat yield and grain quality response to weather could be predicted early, the information would have immense utility. Keeping this in mind, an attempt has been made to develop agrometeorological wheat yield-forecasting model for the Ludhiana district of Punjab. The study was conducted to examine the wheat yields in Ludhiana district over the last 41 years (1970-2010).

Materials and Methods

Ludhiana is located in central region of Punjab (30°54¢ N latitude, 75°48¢ E longitude and an altitude of 247 above mean sea level. The climate is sub-tropical, where winter is mild to cold and summer is hot dry to humid. Average temperature ranges as 22.5-23.9°C and average annual rainfall is 755 mm, where 75% is received during monsoon and 25% during winter season.

Wheat Yield Data: Yearly production (q) and area (ha) under wheat crop in Ludhiana district for the period 1970-2013 were collected from the Statistical Abstract of Punjab (Stats, 2013-14). For each year, the total production of the district was divided by the total acreage to calculate the wheat productivity.

***Weather Data*:** Weekly data of maximum and minimum temperatures, relative humidity, rainfall, sunshine hours and number of rainy days for the period 1970-2013 were collected from the agrometeorological observatory located in the School of Climate Change & Agril. Meteorology, PAU, Ludhiana for the crop season except the harvesting month, since the forecast was made before the harvesting. The solar radiation was converted to sunshine hours through DSSAT model (version 4.5).

***Technology trend*:** Since 1950, crop yield in majority of meteorological sub-divisions registered upward trend due to advance in agri-technology, which has contributed to significant rise in crop yield. In this model, technology trend is used through suitable time-scale dummy variable to account for the effect of technology on wheat yield (Appa Rao, 1983; Sarkar, 2000).

Technology data: To analyze the impact of changing climatic parameters on the productivity of wheat, data on technological variables viz., area under high yielding varieties, fertilizer consumption, irrigated area and farm mechanization were also collected from the statistical abstracts of Punjab and was used to calculate the composite index. It was calculated by taking the value of each technological variable with respect to total cropped area and then averaged out the value as composite index to estimate cotton yield (Gill and Bhatt, 2015).

SPSS Model: SPSS (Statistical Product and Service Solutions) was used to compute Pearson's correlations between observed yield and weather parameters, and with combinations of weather parameters. Sum of weather parameter and sum product of different weather parameter and correlation coefficient has been derived. Multiple regressions between dependent variable (yield) and independent variables (time, sum and sum products for different weather parameters) were carried out. Regression equation was written using the regression formula.

***Method of analysis*:** Functional analysis was used to investigate the impact of various climatic as well as technological factors on crop

Table 1. Effect of meteorological parameters on sensitive period of wheat crop and its yield

Meteorological parameters	Sensitive period (SMWs)	Stage of the crop	Effect on yield
Maximum temperature	1st - 3rd	Vegetative	- ve
Rainfall	7th -10th	Reproductive	- ve
Number of rainy days	7th - 9th	Late vegetative	- ve
Sunshine hours	52nd - 3rd	Vegetative	- ve
Solar radiation	4th - 5th	Early reproductive	+ ve

productivities. Linear, exponential and power functions were employed for the purpose. R-square and significance values were considered to select the best-fit model and the traditional multiple regression technique was employed to develop yield-forecasting models. To avoid the multicollinearity among various technological variables, composite technology index was formed. Firstly, a basic model was developed by using weather parameters (Table 1) without including technological trend from a data series of 41 years (1970-2010). Secondly, a modified model was obtained by introducing an assumed technological trend in the basic model keeping other independent variables constant. Thirdly, a more advanced model was framed by considering the composite index of different technologies. The development of modified models was intended to improve the accuracy of forecast of wheat yield, by superimposing the impact of agricultural technology in the form of linear time scale dummy variable as well as technology trend. Lastly, SPSS model was used to predict the wheat yield. The district average crop yield taken from statistical abstract was taken as dependent variable with weather parameters and technology as independent variables. The general form of the model is:

$$Y_e = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} a_j x_j + \sum_{i=1}^{n} a_j x_k$$

where, Y_e = Estimated yield, kg ha^{-1}, a_0 = Regression constant, a_i = Regression coefficients for meteorological predictor variables, x_i (i = 1, 2, ……n), a_j = Regression coefficients for technological trend variables; x_j = jth technological trend variable and x_k = kth composite index variable.

To analyze the impact of climatic and technological factors on productivities of wheat, weekly data on each climatic variable pertaining to specific crop growing period were correlated with respective crop yield. Using correlation techniques, attempt was made to find the degree and nature of relationship between yield and weekly weather over the crop growing periods i.e., 43rd to 13th standard meteorological week (SMW) to identify the critical periods when weather parameters exert significant influence on yield. These were used to calculate multiple correlations with yield of all combinations by dropping one or more variables, which were found less significant. Those parameters (maximum and minimum temperatures, rainfall, sunshine hour, number of rainy days, and maximum relative humidity), which are statistically significant at the mandatory levels, were used in the final equation. The model was finally verified with independent data for 5 years outside their sampling series. The performance of the model was examined by computing percentage deviations of estimates and forecast yield figures.

Sensitive periods and parameters: The sensitive periods of statistical and phenological significance were selected in terms of standard meteorological weeks (SMWs) for regression analysis.

Results and Discussion

Basic model: The linear model successfully accounted for 63% of total variation in yield with a multiple correlation coefficient (F Ratio =7.58). Both the exponential and power function models accounted for 64% variations in yields with F Ratio of 7.81 and 7.85, respectively.

Multiple linear regression equation (MLRE) for the basic model is

$Y = 1241.21-(47.35*X_1)+(201.46*X_2)+(15.28*X_3)-(65.79*X_4)-(11.26*X_5)+(99.6*X_6)+(83.85*X_7)$

Modified model I: Improved agricultural technology necessitated the need to modify the basic model by introducing technological trend as an independent linear time scale dummy variable. The modified models gave multiple correlation coefficients of 0.73-0.85. Models accounted for 83% (Linear), 85 % (Exponential), and 73 % (Power) of total variation in yields.

The MLRE for Model I is

$Y = 4152.3-(13.79*X_1)-(14.73*X_2) (11.85*X_3) - (105.72*X_4)+(3.43*X_5)-(142.45*X_6)+(28.37*X_7) + (104.17*X_8)$

Modified model II: To analyze the impact of technologies as a composite technological index on wheat productivity, it was found that high yielding varieties, fertilizer usage, mechanization etc. had positive and significant impact on wheat productivity. The model accounted for 88, 89 and 87% yield variability through linear exponential and power functions.

The MLRE for Model II is

$Y = 3520.93-(35.72*X_1)-(32.28*X_2)-(8.95*X_3)-(48.27*X_4)-(0.26*X_5)-(137.76*X_6)+(38.46*X_7) +(8.07*X_8)$

Where, X_1, X_2, X_3, X_4, X_5, X_6, X_7 and X_8 are T_{max} (1-3), T_{min} (9-12), RH_e (1-3), SSH (52-3), Rain (7-10), NoRD (7-9), SR (4-5) and the trend in model I & composite index in model II, respectively

SPSS Model III: The multi-regression analysis using SPSS has been employed for the estimation of wheat yield. The regression for SPSS Model is

$Y = 1232.648+(63.49029*Time)+ (0.383764*Z170)$

Here, Z170 is the sum product of maximum temperature * sunshine hours.

The regression equation showed that maximum temperature and sunshine hours plays the most important role on wheat yield with errors ranged between -10.2 % in year 1972 to 17.7 % in year 2009. The forecasted yield and error based on above regression equation is given in Table 3. The validation results show that the basic linear model predicts yield better in comparison to the exponential and power function models (Table 2). All the statistical parameters show good performance of basic linear model for yield prediction. However, the –ve value of MBE indicates that the linear model under estimate the yield. For modified model-1, linear model predicts better than exponential and power function models. However RMSE 237.7, NRMSE 4.96 and MAE 189.4 represent good yield prediction under exponential model as compared to linear model. The NRMSE value of <10 shows excellent yield prediction under linear and power function models. In modified model II, The CRM values of 0 under linear model represent perfect fit, while the –ve values in exponential and power function models show over-estimation. The NRMSE of 6.92 represent excellent fit under linear model compared to exponential and power function model. In SPSS model, negative CRM represents agreement between actual and predicted yields, while NRMSE of 8.58% represent excellent fit. However the MBE value of 154.4 shows model over estimate the yield.

Table 2. Statistical significance of the basic, modified and SPSS models

Statistical parameters	Basic			Modified 1			Modified 2			SPSS
	Linear	Exponen-tial	Power	Linear	Exponen-tial	Power	Linear	Exponen-tial	Power	
CRM	0.04	0.11	-0.10	0.01	0.07	0.04	0	-0.10	-0.12	-0.03
RMSE	678.51	780.49	828.18	315.83	503.64	237.70	331.33	769.18	1671.95	415.61
NRMSE	14.17	16.30	17.29	6.59	10.52	4.96	6.92	16.06	34.91	8.58
MAE	590.20	683.60	696.60	245.80	476.40	189.40	304.40	704.80	1597.80	344.8
MBE	-171.40	-534.40	496.20	-59.00	-340.40	-168.60	-12.80	489.60	578.60	154.4

Table 3. Validation of the models over independent data set

Year	Yield (kg ha^{-1})				Deviation (%)		
	Predicted			Actual Yield	Linear	Exponential	Power
	Linear	Exponential	Power				
A. Basic model							
2009	4311	4292	4709	4390	-1.8	-2.2	7.3
2010	5411	4737	5941	4364	16.8	2.2	28.2
2011	4590	4019	5505	4964	-7.5	-19.0	10.9
2012	4591	4380	4874	5375	-14.6	-18.5	-9.3
2013	4186	3846	5398	4853	-13.7	-20.7	11.2
B. Modified model I							
2009	4453	3878	4442	4390	1.4	-11.7	1.2
2010	4682	3714	4350	4364	1.0	-19.9	-6.1
2011	4802	4725	4677	4964	-3.3	-4.8	-5.8
2012	4775	4734	4976	5375	-11.2	-11.9	-7.4
2013	4939	5193	4658	4853	1.8	7.0	-4.0
C. Modified model II							
2009	4690	4987	5467	4390	6.8	13.6	24.5
2010	4793	4645	6690	4364	3.4	0.2	44.4
2011	4717	6094	3630	4964	-5.0	22.8	-26.9
2012	4922	4837	7413	5375	-8.4	-10.0	37.9
2013	4760	5831	3639	4853	-1.9	20.1	-25.0
D. SPSS							
2009		5165		4390		17.7	
2010		5008		4634		8.1	
2011		5063		4964		2.0	
2012		5143		5375		-4.3	
2013		4609		4853		-5.0	

The models were validated independently for 5 years (2009-2013) (Table 3). Results revealed that for all the five years, modified model II could predict the yield more satisfactorily over basic models. The yield forecast by multiple linear regression model II lies mostly within -2 to 7%. The satisfactory performance of the composite technological index model signifies that this model has positive and significant bearing on wheat productivity in Ludhiana district and it can be used to forecast wheat yield in the Ludhiana district which represents the central region of Punjab.

Conclusions

Models included county-fixed and crop-year random effects. Yield level is strongly influenced by weather variables as indicated by high R^2 values. For yield, maximum temperature and precipitation have a positive relation with yield, while minimum temperature, relative humidity (evening), sunshine hours and solar radiation were negatively related to yield. In the forecast evaluation, the forecasting ability of yield models was enhanced by adding the composite technological trend factor. Developed models are more accurate than using a basic model to predict yield.

References

Appa Rao, G.J. 1983. Estimation of wheat yields over Punjab using district and state models. *Mausam* **34** (3): 275-80.

Bal, S.K., Mukherjee, J., Mallick, K. and Hundal, S.S. 2004. Wheat yield forecasting models for Ludhiana district of Punjab state. *J. Agrometeorology*. **6**: 161-65.

Da Mota, F.S. 1983. Weather- technology models for corn and soybeans in south of Brazil. *Agric. Meteorol.* **28**: 49-64.

Esfandiary, F., Aghaie G. and Ali D.M. 2009. Wheat yield prediction through Agrometeorological Indices for Arelebil district. *Int. J. Biol. Life Sci.* **5**: 2.

Gill, K.K. and Bhatt, K. 2015. Development of Agrometeorological models for estimation of cotton yield. *J. Krishi Vigyan* **3** (2): 27-34.

Jain, R.C., Sridharan, H and Agarwal, R. 1984. Principal component technique for forecasting of sorghum yield. *Indian J. Agric. Sci.* **54** (6): 467-70.

Kandiannan, K., Karithikeyan, R., Krishnan, R., Kailasam, C. and Balasubramanian, T.N. 2002. A Crop-Weather Model for Prediction of Rice Yield Using an Empirical- Statistical Tehnique. *J. Agron. Crop Sci.* **188**:59-62.

Mahi, G.S., Mavi, H.S., Chaurasia, R., Singh, G. and Jhorar, O.P. 1991. Climate based wheat yield model. In: National Symposium on statistical Methodology for Dryland Agriculture. 28-30 Jan. CRIDA. Hyderabad. 187-94.

Nain, A.S., Dadhwal, V.K. and Singh, T.P. 2002. Real time wheat yield assessment using technology trend and crop simulation model with minimal data set. *Curr. Sci.* **82** (10): 1255-58.

Ramasubramanian, V and Jain, R.C. 1999. Use of growth indices in Markov Chain model for crop yield forecasting. *Biomet. J.* **41** (1): 99-109.

Sarkar, J. 2000. Wheat yield forecasting over Gujarat using agro meteorological model. *J. Maharashtra Agric. Univ.* **25** (3): 294-97.

Sharma, A. Sood, R. K and Kalubarme, M. H. 2004. Agrometeorological wheat yield forecast in Himachal Pradesh. *Journal of Agrometeorology*. 6: 153-60.

Smith, G. P., and Gooding, M.J. 1999. Models of Wheat Grain Considering Climate Cultivar and Nitrogen Effects. *Agric. For. Meteor.* **94**:159-170.

Walker, G. K. 1989. Model for operational forecasting of western Canada wheat yield. *Agric. For. Meteor.* **44**: 339-51.

Received: April 18, 2014; Accepted: June 19, 2015