Crop yield forecasting of paddy, sugarcane and wheat through linear regression technique for south Gujarat

NEERAJ KUMAR, R. R. PISAL, S. P. SHUKLA and K. K. PANDEY*

N. M. College of Agriculture, Navsari Agricultural University, Navsari- 396 450 (Gujarat), India

*V. C. S. G. College of Horticulture, Uttarakhand University of Hortculture and Forestry, Uttarakhand, India

(Received 17 October 2012, Modified 27 May 2013)

e mail : neeraj34012@gmail.com

सार – धान, गन्ना तथा गेहूँ की उपज के पूर्वानुमान के लिए गुजरात के दो जिलों (नवसारी तथा भरूच) के लिए बहु समाश्रयण निदर्शों का विकास किया गया। इसमें नवसारी के 31 वर्षों (1980-2010) तथा भरूच के 27 वर्षों (1984-2010) के मौसम तथा उपज के आँकड़ों का उपयोग किया गया। दोनों जिलों के लिए मॉडल तैयार करने के लिए नवसारी जिले के 27 वर्षों तथा भरूच जिले के 23 वर्षों के डिट्रेंड उपज के आँकड़ें तथा विकसित किए गए मौसम के आँकड़ों का उपयोग किया गया। मॉडल के विकास के लिए महत्वपूर्ण मौसम आँकड़ों का चयन अधिकतम R² तथा P के आधार पर किया गया तथा परीक्षण और त्रुटि पद्धति के आधार पर बहु समाश्रयण विश्लेषण किया गया। मॉडल का मान्यकरण चार सालों (2007-2010) के स्वतंत्र आँकड़ा समुच्चयों के आधार पर किया गया । मॉडल के मान्यकरण के दौरान नवसारी के लिए धान, गेहूँ तथा गन्ना में विचलन क्रमश: -7.30 से 3.41% , 1.68 से 2.05% तथा -8.27 से 11.51% पाया गया। इसी प्रकार भरूच जिले के लिए धान, गेहूँ तथा गन्ना में विचलन क्रमश: 5.35 से 11.76%, -12.65 से 7.18% तथा -12.07 से 6.86% पाया गया।

ABSTRACT. Multiple regression models were developed for yield forecasting of paddy, sugarcane and wheat for two districts of Gujarat (Navsari and Bharuch). The historical weather and crop yield data of 31 years of Navsari (1980-2010) and 27 years of Bharuch (1984-2010) were used. The data of de-trend yield and generated weather variables for 27 years of Navsari (1980-2006) and 23 years of Bharuch (1984-2006) were used for generation of the model for both districts. Significant weather variables are obtained on the basis of highest R^2 and significant P-value. The multiple regression analysis was executed by trial and error method. The models were validated with 4 years independent data set (2007 to 2010) of these two districts. During the validation period, Navsari district model deviations for paddy, sugarcane and wheat were between -7.30 to 3.41%, 1.68 to 2.05% and -8.27 to 11.51% respectively. Similarly Bharuch model deviations for paddy, sugarcane and wheat were between 5.35 to 11.76%, -12.65 to 7.18% and -12.07 to 6.86% respectively.

Key words – Regression models, R^2, Validation, Yield forecasting.

1. Introduction

Agriculture is an economic sector which depends highly on climatic conditions. Crop models are frequently used to evaluate the ability of climate forecasts in guiding crop management practices. In statistical model approach, one or several variables (representing weather or climate) are related to crop responses such as yield and yield contributing characters. Crop yield is the integrated effect of a number of physical and physiological processes that occur during the crop-growing period. These processes are influenced by the characteristics of the crop, weather, soil and time management factors (Rafi and Rehan, 2005). Statistical models provide simple alternative to process-based models. The main advantages of statistical models are their limited reliance on field calibration data and their transparent assessment of model uncertainties (Lobell and Burke, 2010). Shankar and Gupta (1988) related the meteorological parameters and yield data using the phenological periods, in which weather elements illustrate significant correlation with crop yield, by using multiple linear regressions. The work of Wickham (1973) also clearly brought out that the yield variation in paddy crop production, due to weather, management and biotic factors can be addressed through a statistical modelling approach. Regression models are often used because of their simple and straight forward nature, relating yield with one or more meteorological factors. As such these are used extensively in the studies on soils, agronomic practices and insect-disease; a combination of such factors is still difficult with some of the dynamic simulation models. Many earlier researchers (Norman 1979), Fisher (1924) and Hendrick and Scholl (1943) have suggested models which require small number of parameters to be used

while taking care of distribution pattern of weather over the crop season. Models based on weather parameters can provide reliable forecast of crop yield in advance of harvest (Agrawal and Mehta, 2007). Parthasarathy *et al*. (1988) and Ramakrishna *et al*. (2003) have developed forecast equations based on regression model, for total Indian food grain production using monsoon rainfall and soil index. Regression equations have also been developed for forecasting paddy yield (Shankar and Gupta, 1987), for estimation of sugarcane yield (Singh and Bapat, 1988) and for wheat yield (Agarwal *et al*., 2012). Keeping above facts in view, the present study was undertaken to predict yields of rice, wheat and sugarcane.

Rice is cultivated during Kharif season (June to October), wheat is sown during Rabi season (November to March) and Sugarcane which is one year crop, grown during Nov to Nov in both the districts of Gujarat.

2. Materials and methodology

Weekly weather data of crop growing season of 31 years (1980-2010) of Navsari district (20°57' latitude, 72°54' longitude and 11.89 m above mean sea level) and 27 years (1984-2010) of Bharuch district (21°42' latitude, 72°58' longitude and 15.0 m above mean sea level) were collected and used in the present study. The yield data was procured from the official site (Directorate of Agriculture, Gujarat) for Navsari and Bharuch for the same time period. Weekly data of maximum temperature (Tmax) °C, minimum temperature (Tmin) °C, morning relative humidity (RH-I)%, afternoon relative humidity (RH-II)%, bright sunshine hours (BSS) hours/day and rainfall (mm) were employed according to growing period of each crop. Correlations were worked out between forecasters (weather parameters) with respective year yield of crop. The yield data was detrended.

The step-wise forward model (1) can be used for obtaining the most significant weather variable, by which we can fit the multiple regression model (2) and these most significant variables can be used for correlation coefficient.

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + b_n X_n$$

$$n = 1, 2, \ldots 5 \tag{1}$$

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + b_n X_n$$

$$n = 1, 2, 3 \tag{2}$$

The summary measures describe the quality of simulation while the difference measures try to locate and quantify errors. The latter include the Mean Absolute Error (MAE), the Mean Bias Error (MBE) and the Root Mean Square Error (RMSE) (Varshneya *et al*., 2010).

$$MAE = \sum_{i=1}^{n} (1F_i - O_i 1)/n$$

$$MBE = \sum_{i=1}^{n} (F_i - O_i)/n$$

$$RMSE = \left[\sum_{i=1}^{n} (F_i - O_i)^2)/n \right]^{\frac{1}{2}}$$

MAE and RMSE indicate the magnitude of the average error, but provide no information on the relative size of the average difference between (F) and (O). The statistic MBE describes the direction of the error bias. The value of MBE is related to the magnitude of the values under investigation. A negative MBE indicates that the forecasting is smaller in values than those of the corresponding observations. PE is defined as ratio of RMSE to mean observed value expressed as percentage (Varshneya *et al*., 2010).

$$PE = \frac{RMSE}{\bar{O}} \times 100$$

3. Results and discussion

3.1. *Validation of models*

The observed and forecasted yields for period (2007-2010) have been presented in Table 2 and various error analysis of independent and all data set in Table 3. The regressions models were validated with the four years (2007-2010) of independent data set. The data exposed that the Navsari district paddy model deviated -7.30 to 3.41% from the observed yield, the error analysis revealed that the MAE, MBE, RMSE, PE and average error of model were 94.25 kg ha[-1], -94.25 kg ha[-1], RMSE ± 96.42 kg ha[-1], 2.87 % and 4.9% respectively. Bharuch regression model for paddy deviated between 5.35 to 11.76% and MAE, MBE, RMSE, PE and average error of model were 89.21 kg ha[-1], -89.21 kg ha[-1], ± 102.87 kg ha[-1], 2.68% and 3.6% respectively.

For sugarcane crop the Navsari district model diverged between 1.68 to 2.05% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 526.22 kg ha[-1], -526.22 kg ha[-1], ± 466.46 kg ha[-1], 5.66 % and 6.66% respectively. Similarly for Bharuch district the model deviated between -12.65 to 7.18% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 479.89 kg ha[-1], -479.89 kg ha[-1], ±378.59 kg ha[-1], 14.68% and 7.9%, respectively.

TABLE 1

Statistical models for paddy, sugarcane and wheat forecast at Navsari and Bharuch districts

Model	Crops	Equation	R^2	Model Error
Navsari	Paddy	$Y = 20.791 + 0.019*Rain + 0.721*BSS + 0.362*EP$	0.69	4.9%
	Sugarcane	$Y = 875.667 - 83.72*Tmin + 0.846*BSS$	0.63	6.6%
	Wheat	$Y = 64.942 - 1.687*Tmin$	0.57	7.2%
Bharuch	Paddy	$Y = 17.458 + 0.328*Rain - 0.871*BSS + 0.687*RH\text{-}II$	0.78	3.6%
	Sugarcane	$Y = 677.254 - 7.017*Tmax + 0.595*Tmin$	0.58	7.9%
	Wheat	$Y = 7.992 - 3.367*Tmax + 3.768*Tmin$	0.53	8.7%

TABLE 2

Validation of model for Navsari and Bharuch districts

Crops	Years	Navsari			Bharuch		
		Observed	Forecast	Error (%)	Observed	Forecast	Error (%)
Paddy (kg ha^{-1})	2007	2580	2492	3.41	1646	1558	5.35
	2008	2469	2567	-3.97	1671	1573	5.86
	2009	2487	2419	2.73	1590	1482	6.79
	2010	2577	2765	-7.30	1684	1486	11.76
Sugarcane (kg ha^{-1})	2007	72000	70789	1.68	74640	76491	-2.48
	2008	68700	67289	2.05	71010	73121	-2.97
	2009	70300	68989	1.86	72600	67389	7.18
	2010	72200	70989	1.68	72810	82021	-12.65
Wheat (kg ha^{-1})	2007	3179	3333	-4.84	1930	2163	-12.07
	2008	2503	2215	11.51	1374	1306	4.95
	2009	2571	2749	-6.92	1559	1452	6.86
	2010	3121	3379	-8.27	1600	1788	-11.75

TABLE 3

Error analysis of forecasting models for Navsari and Bharuch districts

Crops	Error analysis	Navsari		Bharuch	
		All data set	Independent data set (2006-2010)	All data set	Independent data set (2006-2010)
Paddy	MAE (kg ha^{-1})	87.75	94.25	78.88	89.21
	MBE (kg ha^{-1})	-16.46	-94.25	-13.46	-89.21
	RMSE (kg ha^{-1})	22.67	96.42	46.28	102.87
	PE (%)	3.64	2.87	3.2	2.68
	Average error (%)	2.2	4.9	8.7	3.6
Sugarcane	MAE (kg ha^{-1})	917.79	526.22	1012.38	479.89
	MBE (kg ha^{-1})	-64.33	-526.22	-83.26	-479.89
	RMSE (kg ha^{-1})	523.06	466.46	697.57	378.59
	PE (%)	3.33	5.66	7.99	14.68
	Average error (%)	0.88	6.6	2.28	7.9
Wheat	MAE (kg ha^{-1})	68.53	67.42	76.87	39.6
	MBE (kg ha^{-1})	-17.64	-67.42	-11.25	-39.6
	RMSE (kg ha^{-1})	89.82	77.28	86.42	62.11
	PE (%)	8.76	2.49	9.52	6.42
	Average error (%)	0.97	7.2	1.49	8.7

In Navsari district wheat model oscillated between -8.27 to 11.51% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 67.42 kg ha^{-1}, -67.42 kg ha^{-1}, ±77.28 kg ha^{-1}, 2.49% and 7.2% respectively, while at Bharuch district model variations were between -12.07 to 6.86% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 39.6 kg ha^{-1}, -39.6 kg ha^{-1}, ± 62.11 kg ha^{-1}, 6.42% and 8.7%, respectively. All the models performed very well under diverse conditions.

3.2. *Regression model evaluation for paddy (kg ha^{-1})*

The step wise regression analysis was indicating that rainfall, bright sunshine hours and evaporation at Navsari, while at Bharuch rainfall, bright sunshine hours and maximum relative humidity showed maximum and significant effect on paddy yield. The R^2 and model error percent values for paddy crop for Navsari were 0.69 and 4.9% respectively and for Bharuch, value of R^2 was 0.78 and model error percent was 3.6% (Table 1).

3.3. *Regression model evaluation for sugarcane (kg ha^{-1})*

The step wise regression analysis was indicating that minimum temperature and bright sunshine hours at Navsari, while at Bharuch maximum temperature and minimum temperature showed maximum and significant effect on sugarcane yield. Navsari district R^2 value and model error percent were 0.63 and 6.6% respectively and for Bharuch values of R^2 along with model error percent were 0.58 and 7.9% respectively (Table 1). These results are comparable with the sugarcane forecasting by Singh and Bapat (1988) and Mehta *et al*. (2000).

3.4. *Regression model evaluation for wheat (kg ha^{-1})*

The step wise regression analysis was indicating that only one parameter like minimum temperature at Navsari while at Bharuch maximum temperature and minimum temperature showed maximum and significant effect on sugarcane yield. Navsari district R^2 value and model error percent were 0.57 and 7.2% respectively and for Bharuch district values of R^2 and model error percent were 0.53 and 8.7% respectively (Table 1). Similar results have also been reported by Singh *et al*. (2008).

4. **Conclusion**

In this paper an attempt has been made to obtain yield forecast for paddy, wheat and sugarcane by using weather data. Analysis is based on time series weather and yield data. The performance of weather variables indicate that rainfall is important for rice crop, while minimum temperature for sugarcane and wheat are more important in both the districts. The multiple regression models were found to be efficient in forecasting yields satisfactorily in respect of the three crops.

References

Agrawal, Ranjana, C. and Aditya, K., 2012, "Use of discriminant function analysis for forecasting crop yield", *Mausam*, **63**, 3, 455-458.

Agrawal, Ranjana and Mehta, S. C., 2007, "Weather based forecasting of crop yields, pests and diseases - IASRI Models", *J. Ind. Soc. Agril. Stati. ft.*, **61**, 2, 255-263.

Fisher, R. A., 1924, "The influence of rainfall on the yield of London wheat at Rothamsted", *Phil. Trans. Roy. Soc.*, **213**, 89-142.

Hendrick, W. A. and Scholl, J. E., 1943, "Technique in measuring joint relationship the joint effects of temperature and precipitation on crop yield", *North Carolina Agric. Exp. Sta. Tech. Bull.*, **74**.

Lobell, D. B. and Burke, M. B., 2010, "On the use of statistical models to predict crop yield responses to climate change", *Agric. Forest Meteorol*, 10.1016/j. agrformet. 2010.07.008

Mehta, S. C., Agrawal R. and Singh, V. P. N., 2000, "Strategies for composite forecast", *J. Ind. Soc. Agril. Sci*., **53**, 3, 262-272.

Norman, J. M., 1979, "Modelling the complete crop canopy. In Modification of the aerial environment of the crops", *ASAE*, St. Josephs, Michigan, USA, p538.

Parthasarathy, B., Mount, A. and Kothawala, D., 1988, "Regression model for estimation of Indian food grain production for summer monsoon rainfall" *Agricultural and Forest Meteorology*, **42**, 167-182.

Rafi, Z. and Rehan, A., 2005, "Wheat crop model based on water balance for agro-meteorological crop monitoring", *Pakistan Journal of Meteorology*, **2**, 3, 23-33.

Ramakrishna, Y. S., Singh, H. P. and Nageswara Rao, G., 2003, "Weather based indices for forecasting national food grain production", *J. Agrometeorology*, **5**, 1, 1-11.

Shankar, U. and Gupta, B. R. D., 1987,"Forecasting the yield of paddy at Chisurah in west Bengal using multiple regression technique". *Mausam*, **38**, 415-418.

Shankar, U. and Gupta, B. R. D., 1988, "Forecasting paddy yield in Bihar and Orissa states in India based on weather parameters and multiple regression technique", *Tropical Agriculture*, UK, **65**, 3, 265-267.

Singh, B. H. and Bapat, S. R., 1988, "Pre-harvest forecast model for prediction of sugarcane yield", *Ind. J. Agric. Sci*., **58**, 6, 465-469.

Singh, H., Hundal, S. S. and Kaur, P., 2008, "Effect of temperature and rainfall on wheat yield in south western region of Punjab", *J. Agrometeorology*, **10**, 1, 70-74.

Varshneya, M. C., Chinchorkar, S. S., Vaidya, V. B., and Pandey, V., 2010, "Forecasting model for seasonal rainfall for different regions of Gujarat", *J. Agrometeorology*, **12**, 2, 202-207.

Wickham, T. H., 1973, "Predicting yield in low land Paddy through a water balance model in Philippine irrigation system: Research's and operations", *Los Banos*, Philippine, 155-181.