Crop yield forecasting of paddy and sugarcane through modified Hendrick and Scholl technique for south Gujarat

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

Agricultural Meteorological Cell, Department of Agriculutral Engineering, N. M. College of Agriculture,

Navsari Agricultural University, Navsari – 396 450 (Gujarat), India

*Department of Agricultural statistics, V. C. S. G. College of Horticulture, Uttatrakhand University of

Horticulture and Forestry, Bharsar, Pauri – Garhwal, Uttatrakhand, India

e mail : neeraj34012@gmail.com

सार – दक्षिणी गुजरात के छह जिलों की धान और गन्ने की फसलों के लिए परिष्कृत हैंड्रिक और शोल तकनीक से समाश्रयण निदर्श विकसित किए गए। 27 वर्षों की फसलों और मौसम के प्राचलों के आँकड़ों का विश्लेषण किया गया है। मॉडल को विकसित करने के लिए 25 वर्षों के आँकड़े लिए गए हैं। 2010 और 2011 के डाटासेटों का उपयोग करते हुए मॉडल को प्रमाणित किया गया है। क्रमवार समाश्रयण विश्लेषण ट्रायल एरर पद्धति से किया गया जिससे 5% स्तर पर विशेष प्रागुक्तों का श्रेष्ठ सम्मिश्रण प्राप्त हुए। मॉडल की फिटिंग के लिए बहुसमाश्रयण तकनीकों का इस्तेमाल किया गया और R² अधिकतम और निम्नतम प्रतिशत त्रुटि से सर्वश्रेष्ठ पाया गया। फसल उत्पादों के सभी पूर्वानुमान मॉडलों से अच्छे परिणाम प्राप्त हुए हैं और उत्पन्न त्रुटि प्रतिशत स्वीकार्य श्रेणी में है। इस विश्लेषण से यह पता चलता है कि धान और गन्ने के फसल उत्पादन जिलों में मॉडल एरर प्रतिशत क्रमश: -10.0 से 8.1% और -12.2 से 1.5% था। वर्ष 2012 के लिए फसल उत्पादन पूर्वानुमान प्रमाणित मॉडल के आधार पर है जो नवसारी, सूरत, भड़च, वालसाड, नर्मदा और तापी जिलों के लिए तैयार किए गए है।

ABSTRACT. Regression models by modified Hendrick and Scholl technique were developed on paddy and sugarcane for six districts of south Gujarat. The data on the yield and weather parameters were analyzed for 27 years. The 25 year data was used for development of the model. The validation of model was done using data set of 2010 and 2011. The stepwise regression analysis was executed by trial and error method to obtain the finest combination of predictors, significant at 5% level. The multiple regression techniques was used for fitting of the model and decided best by highest R^2 and lowest percent error. All crop yield forecasting models gave good estimates and produced error percent within acceptable range. Analysis revealed that the model error percent of paddy and sugarcane for respective crop growing districts were -10.0 to 8.1% and -12.2 to 1.5% respectively. Crop yield forecasting for year 2012 based on validated model was made for the districts of Navsari, Surat, Bharuch, Valsad, Narmada and Tapi.

Key words – Percent error, RMSE, Stepwise regression and yield forecasting etc.

1. Introduction

Analysis of data provides near real time information about the crop state in quality and quantity with the possibility of early warning, so that timely interventions can be planned and undertaken. Reliable and timely forecasts are essential for agriculture policy making and also for crop production, marketing, storage and transportation decision. This helps in managing risk associated with these activities (Bannayan and Crout, 1999; Potgieter *et al*., 2005 and Tripathi *et al*., 2012). The growth and fluctuations of annual food production (consisting of rice, wheat, coarse grains, and also grams and pulses) have been examined by Gadgil *et al*. (1999). They also observed a correlation coefficient of 0.61 between the rice production anomaly of India and the Indian SW monsoon rainfall. Chowdhury and Das (1993) made a multiple regression model for forecasting the *kharif* food production of India, using Indian SW monsoon rainfall as one of the parameters of the model. Yield forecasting utilize crop and weather data over long period of time pertaining to locations under consideration.

Crop yield in different years are affected due to technological change, system productivity and climatic variability. Individual effects of weather parameters on crop yields were studied by Jain *et al.* (1980) and yield forecasting models based on weather factors were constructed by Agrawal *et al.* (1986); Rao *et al.* (2012) and Munu *et al.* (2013). Agrometeorological crop yield forecasting (ACYF) methods provide a quantitative estimate of the expected crop yield over a given area, in advance of the harvest and in a way that constitutes an improvement over trends, provided no extreme conditions occur (Agrawal and Mehta, 2007). They are based on the common-sense assumption that weather conditions are the main factor behind the inter-annual (short-term) variations of detrended crop yield series (Petr, 1991). 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. 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 (Agrawal and Aditya, 2012). Keeping above facts in view, the present study was undertaken.

2. Materials and methods

Total six districts were selected for forecasting of paddy and sugarcane yields. District wise crop yield data for the period of last 27 years (1985 to 2011) were procured from Directorate of Agriculture, Gujarat state. Weather data were analyzed for each districts of similar period. Out of 27 year data base, the 25 year data were used for development of the model and rest two years yield data (2010 and 2011) were used for validation of the model. Weekly mean 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 (Rain) mm were considered according to growing period of each crop. District wise weekly weather data of growing season of each crop, *i.e.*, 22th Standard Meteorological Weeks (SMW) to 41th (SMW) for paddy and full year data starting from 44th (SMW) to next year 43th (SMW) were used for sugarcane.

2.1. *Individual generated weather variables*

In order to study the forecast of district wise yield using weather variables, two new variables from each weather variable were generated.

Let X_{iw} be the value of ith (i = 1, 2,…p) weather variable at wth weeks (w = 1, 2,…..n) in this study n is 20 for paddy and 52 for sugarcane for the development of the model.

$$Z_{ij} = \frac{\sum_{w=1}^{n} r_{iw}^{j} X_{iw}}{\sum_{w=1}^{n} r_{iw}^{j}}$$

j = 0 for unweighted generated variable and j = 1 for weighted generated variable and r_{iw} be the correlation coefficient between weather variable X_i at wth week used as a weight and detrended yield over the period of K years.

2.2. *Joint generated weather variables*

Weather variables on crop-yield, the joint effect also been calculated by including interaction terms in the model as follows:

$$Q_{ii',j} = \sum_{w=1}^{n} r_{ii',w}^{j} X_{iw} X_{i'w} \bigg/ \sum_{w=1}^{n} r_{ii',w}^{j} \qquad j = 0, 1$$

where, $r_{ii',w}$ is the correlation coefficient between crop yield (detrended) Y and product of weather variables X_{iw} and $X_{i'w}$. Clearly, we have two generated variables (interaction term), including these two interaction terms in the model.

The stepwise regression technique was used to find out the most significant variable. The multiple regression model was used on the basis of yield as regressor and significant generated variable as regress and for the forecasting of yield. The weather indices based modified Hendricks and Scholl model proposed by IASRI, New Delhi by Agarwal and Mehta (2007) was used.

The model is given below:

$$\text{Model } Y = a + \sum_{i=1}^{p} \sum_{j=0}^{2} b_{ij} z_{ij}' + \sum_{i \neq i'=1'}^{p} \sum_{j=0}^{2} b_{ii'j} Q_{ii'j} + CT + \varepsilon$$

where,

$$Z_{ij} = \sum_{w=1}^{m} r_{iw}^{j} X_{iw} \bigg/ \sum_{w=1}^{m} r_{iw}^{j}$$

and

$$Q_{ii',j}' = \sum_{w=1}^{m} r_{ii'w}^{j} X_{iw} X_{i'w} \bigg/ \sum_{w=1}^{m} r_{ii',w}^{j}$$

TABLE 1

Crop yield forecasting models for paddy and sugarcane for all six districts

Crops	District	Equations	R^2	Forested Yield for 2012
Paddy	Navsari	Y = 1358.80 + Z31 * 1.255 + 44.954 * time	0.71	2709
	Surat	Y = 1530.816 + Q231 * 0.058 + 26.376 * time	0.75	2289
	Bharuch	Y = 2346.909 + Q121 * 0.745 - Q150 * 0.022 + Q351 * 0.016 + 51.754 * time	0.89	1725
	Valsad	Y = 1403 + Z21 * 1.204 + 28.832 * time	0.60	2230
	Narmada	Y = 2489.618 + Z41 *14.76	0.76	920
	Tapi	Y = 1580.086 + Z31 * 1.174 + 20.309 * time	0.71	2040
Sugarcane	Navsari	Yield = 86803.520 + Q351 * 0.303 + 733.385 * time	0.75	66721
	Surat	Yield = 83397.060 + Z31 * 23.994 + Z50 * 87.451 + Z51* 237.223+ 720.913 * time	0.89	65635
	Bharuch	Yield = 1317.722 + Z41 * 233.562 + Q131 * 1.084 + 537.836 * time	0.60	68881
	Valsad	Yield = 73299.17 + Q231 * 1.765 + 802.240 * time	0.89	60862
	Narmada	Yield = 133450.000 - Q140 * 0.903 + Q141 * 8.420 + Q251 * 6.690 + 474.913 * time	0.76	70171
	Tapi	Yield = 82683.139 + Z31 * 24.215 + Z50 * 89.137 + Z51 * 241.589 + 707.808 * time	0.71	65211

where,

a, b and C	constant
Z_{ij}	generated Variable (individual)
$Q_{ii'j}$	generated Variable (interaction term)
r_{iw}	is correlation coefficient of yield with i-th weather variable in w-th period
$r_{ii'w}$	is correlation coefficient (adjusted for trend effect) of yield with product of i-th and i'-th weather variables in w-th period
m	is period of forecast
p	is number of weather variables used
T	time trend
ε	is random error distributed as $N(0,\sigma^2)$

In this approach for each weather variable, two type of indices were developed, one as simple total values of weather variable in different periods and other as weighted total, weights being correlation coefficients between yield and weather variable in respective periods. On similar lines for studying joint effects, un-weighted and weighted indices for interaction were computed with products of weather variables. The weighted and un-weighted weather variables were developed with their interaction with each other by taking two at a time Tripathi *et al.* (2012). Stepwise regression technique was used to select important weather indices. The models were validated with independent data set of years 2010 and 2011.

The summary measures describe the quality of simulation while the difference measures try to locate and quantify errors. The latter include the Root Mean Square Error (RMSE).

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

RMSE indicate the magnitude of the average error, but provide no information on the relative size of the average difference between forecasted and observed. PE is defined as ratio of RMSE to mean observed value expressed as percentage (Varshneya *et al.*, 2010).

<div align="center">

TABLE 2

Validation of models

</div>

Districts	Years	Paddy				Sugarcane			
		For	Obs	RMSE	Error (%)	For	Obs	RMSE	Error (%)
Navsari	2010	2551	2493	48.27	1.86	69831	70300	460.72	0.64
	2011	2740	2704			71863	72430		
Surat	2010	2506	2477	38.45	1.62	68988	69100	159.62	0.22
	2011	2314	2268			75181	74931		
Bharuch	2010	1697	1790	86.74	4.73	71654	72600	357.23	0.49
	2011	1791	1871			72956	72810		
Valsad	2010	1990	1949	39.52	1.92	70316	71550	554.70	0.79
	2011	2121	2159			67472	68010		
Narmada	2010	894	847	55.01	6.38	69646	70450	665.07	0.90
	2011	937	875			75979	76800		
Tapi	2010	2240	2278	40.04	1.82	70260	70760	385.05	0.54
	2011	2142	2100			62130	61050		

For : Forecast, Obs : Observed

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

3. Results and discussion

3.1. *Crop yield forecast models*

3.1.1. *Paddy*

As stated earlier, weekly weather data from 22th to 41th week were utilized to develop paddy yield forecast models for Navsari, Surat, Bharuch, Valsad, Narmada and Tapi districts by trial and error method for obtaining highest R^2 and lowest model error. Regression analysis was conducted to evaluate the cumulative effect of selective meteorological parameters on paddy yield. All the district paddy models showed significant correlations (R^2). The highest correlation was found in case of Bharuch and Dang (R^2 0.89) followed by Narmada, Surat, Navsari, Tapi and Valsad with values 0.76, 0.75, 0.71, 0.71 and 0.60, respectively (Table 1). Paddy yield forecast for Navsari showed their dependency on weighted maximum temperature (Z_{31}) and also affected by time trend. Surat model gave dependency on product weighted variable (Z) of maximum (Z_{11}) and minimum temperature

(Z_{21}). Bharuch showed more number of variables *viz.,* weighted product of Tmax and Tmin (Q_{121}), Rain and RH-II (Q_{351}), un-weighted product of Tmax and RH-II (Q_{150}). Valsad, Narmada, Tapi and Dangs models are depends on weighted weather variables of Tmin (Z_{11}), RH-I (Z_{41}), Rain (Z_{31}) and product of Tmin with Rain (Q_{231}), respectively (Table 1). The Navsari model forecasted the highest paddy yield (2709 kg ha[-1]), while Narmada district model forecasted lowest paddy yield (920 kg ha[-1]), for the remaining districts *viz.,* Surat, Bharuch, Valsad, and Tapi the model forecast were, 2289, 1725, 2230 and 2040 kg ha[-1], respectively.

3.1.2. *Sugarcane*

As stated earlier, one year weekly weather data starting from 44th was utilized to develop sugarcane yield forecast models for Navsari, Surat, Bharuch, Valsad, Narmada and Tapi districts. Regression analysis was conducted to evaluate the cumulative effect of selective meteorological parameters on sugarcane yield. All the district sugarcane models showed significant correlations (R^2). The highest correlation was found in case of Surat and Valsad (R^2 0.89) followed by Narmada, Navsari, Tapi, and Bharuch (R^2 0.76, 0.75, 0.71 and 0.60,

respectively) (Table 1). Sugarcane yield forecast for Navsari and Valsad showed their dependency on weighted product of Rain *vs* RH-II (Q_{351}) and *T*min *vs* Rain (Q_{231}), respectively. Number of weather variables were more for districts Surat, Bharuch, Narmada and Tapi. Similarly, for Surat it depends on weighted variables of Rain (Z_{31}), weighted and un-weighted variables of RH-II (Z_{51} and (Z_{50}). Bharuch model rely on weighted variables of RH-II (Z_{51}) and weighted product of Tmax with Rain (Q_{131}). Narmada yield prediction model depends on weighted products of *T*max with RH-I (Q_{141}), *T*min with RH-II (Q_{251}) and un-weighted product of *T*max with RH-II (Q_{150}). Tapi district model is rely on weighted variables of Rain (Z_{31}), RH-II (Z_{51}) and un-weighted variables of RH-II (Z_{50}) (Table 1). Among all sugarcane producing districts Narmada model forecasted the highest sugarcane yield (70171 kg ha^{-1}), while Valsad district model forecasted lowest sugarcane yield (60862 kg ha^{-1}), the remaining districts *viz.,* Navsari, Surat, Bharuch and Tapi of about 66721, 65635, 68881 and 65211 kg ha^{-1}, respectively. Different districts showed its variation in predictor is due to their different agro-climatic condition and having variations in response.

3.1.3. *Validation*

The observed and forecasted yields for period (2010-2011) and various error analysis of independent data have been presented in Table 2. The regressions models were validated with two years (2010 and 2011) of independent data set. The data exposed that the paddy yield forecasting models for all districts showed their reliability by producing error below 10% (Table 2). The error structures for districts Navsari, Surat, Bharuch, Valsad, Narmada and Tapi were *viz.,* RMSE $\pm$ 48.27, 38.45, 86.74, 39.52, 55.01 and 40.04 kg ha^{-1}, PE 1.86, 1.62, 4.73, 1.92, 6.38 and 1.82% (Table 2). Similar type of results was found with that of Rankja *et al.* (2009) and Tripathy *et al.* (2012).

Sugarcane yield forecasting models for all districts showed their reliability by producing error below 5%. The error structure for districts Navsari, Surat, Bharuch, Valsad, Narmada and Tapi were *viz.,* RMSE $\pm$ 460.72, 159.62, 357.23, 554.70, 665.07 and 358.05 kg ha^{-1}, PE 0.64, 0.22, 0.49, 0.79, 0.90 and 0.54% (Table 2).

4.   Conclusion

Forecasting models were developed based on modified Hendrick and Scholl technique for paddy and sugarcane crops by using past years yield and weather data. In this techniques time trend, various weather variables and their weighted and un-weighted indices were utilized. The combined effect of weather variables *viz.,*

rainfall and minimum temperature for paddy, similarly maximum, minimum temperature and rainfall for sugarcane played crucial role in yield determination. The models are relying on various weather parameters during their growing period. All models gave the good estimates for yield forecast by giving higher regression coefficient and lower error per cent during validation period. Hence combination of weather and yield data is appropriate and consistent option for yield forecasting.

References

Agrawal, Ranjana, Jain, R. C. and Jha, M. P., 1986, "Models for studying rice crop-weather relationship", *Mausam*, **37**, 1, 67-70.

Agrawal, Ranjana and Mehta, S. C., 2007, "Weather based forecasting of crop yields, pests and diseases - IASRI models", *J. Ind. Soc. Agril. Stat.*, **61**, 2, 255-463.

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

Bannayan, M. and Crout, N. M. J., 1999, "Stochastic modeling approach for real time forecasting of winter wheat yield", *Field Crop Research*, **62**, 85-95.

Chowdhary, A. and Das, H. P., 1993, "Southwest monsoon rains and *kharif* production", *Mausam*, **44**, 4, 381-394.

Gadgil, S., Abrol, Y. P. and Seshagiri Rao, P. R., 1999, "On growth and fluctuation of Indian food grain production", *Curr*ent *Science*, **76**, 548-556.

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., p74.

Jain, R. C., Agrawal, R. and Jha, M. P., 1980, "Effect of climatic variables on rice yield and its forecast", *Mausam*, **31**, 4, 591-596.

Munu, D., Hussain, R., Sarma, A., Sarmah, K. and Deka, R. L., 2013, "Yield prediction of winter rice employing meteorological variables in central and upper Brahmaputra valley zone of Assam", *J. Agrometeorology*, **15** (Special Issue 1), 162-166.

Petr, J., 1991, "Weather and Yield", *Dev. Crop Sci. No. 20*, Elsevier, Amsterdam, p288.

Potgieter, A. B., Hammer, G. L. Dohetry, A. and de Voil, P., 2005, "A simple regional-scale model for forecasting sorghum yield across North-Eastern Australia", *Agril. Forest Met.*, **132**, 143-153.

Rankja, N. J., Upadhyay, S. L., Varmora, S. L., Parmar, B. A. and Sahu, D. D., 2009, "Estimation of cotton yield based on weather parameters of Junagadh district in Gujarat state", *J. Agrometeorology*, **11**, 2, 152-155.

Rao, V. U. M., Nair, Linitha, Rao, B. B. and Kumar, P. V., 2012, "Climatic sensitivity of mustard crop in Northern India: An assesment of yields in future climates from real time data", *J. Agrometeorology,* **14** (Special Issue), 10-18.

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.

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.

Tripathi, M. K., Mehra, B., Chattopadhyay, N. and Singh, K. K., 2012, "Yield prediction of sugarcane and paddy for districts of Uttar Pradesh", *Journal of Agrometeorology*, **14**, 2, 173-175.

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.