GAU Res. J. 1999. 24 (2) : 79-83

Pre-Harvest Forecast Models for Mango (Var. *Kesar*) Yield Using Weather Variables

R. C. Bharati*, T. J. Khatri and R. K. Parikh

Department of Agricultural Statistics, N. M. College of Agriculture, Gujarat Agricultural University, Navsari

ABSTRACT

Data on mango (var. Kesar) tree yield along with age of the tree and weather variables such as minimum of minimum temperature during October, December, February and rainfall during whole year were collected from Regional Fruit Research Station and Observatory situated at the College Farm of Gujarat Agricultural University, Navsari for the past 13 years. Regression models for all possible subsets were fitted and Mallow's Cp criterion R^2, Adj R^2 & mean square error were utilised to select the best fit. Two models, one at flowering stage and the other at fruit setting stage were developed with $R^2 = 0.808$ and $R^2 = 0.928$ respectively . The last variable i.e. age of the tree helps to provide age-wise prediction of a tree which is necessary for a perennial tree crop.

(Received on 24th March, 1999)

INTRODUCTION

In forecast modelling, the knowledge of yield contributing factors is essential. In India weather variables play an important role in deciding the yield of a crop. Regression type prediction models have been developed by many workers to predict the yield of a crop using weather variables. Gangopadhyaya and Sarkar (1964), Thompson (1969) and Brown and Venderlip (1969) while studying the effect of weather factors on yield of different crops, agreed that the weather data when included for specific growth periods are better than monthly data to explain variation in the crop yield. LeDuc *etal.* (1980) described climatic crop relationship yield model in an operational system and recognised location-wise regression type yield prediction model. Huda *et al.* (1985) used step-wise regression technique and Mallow's Cp criterion to develop a model for predicting sorghum yield using environmental factors with 73% yield variation. But Hamaker and Oosterhaff in the book `Fitting Equations to Data' written by Cuthburt Daniel and Fred S. Wood (1980) showed that the step-wise regression can lead to confusing results, when the independent variables are highly correlated. In the present paper, an attempt has been made to develop a regression type mango (var. Kesar) yield prediction model based on weather variables and age of the tree collected from past records of Regional Fruit Reseach Station and

*Present address : Assistant Professor (Statistics), Dept., of Stat., Math. and Comp. Appl., F.B.S.H., R.A.U., Psua, Samastipur-848 125.

observatory situated at the college farm of Gujarat Agricultural University Campus Navsari by employing a most comprehensive approach of variable selection.

MATERIALS AND METHODS

Data for this study were taken from two sources. The yield of 10 mango (var. *Kesar*) trees of the same age, randomly selected with same set of trees over years and its age (calculated from the data of planting) for the past 13 years for the period from 1981 to 1993 were obtained from Regional Fruit Research Station situated at the college farm of Gujarat Agricultural University Campus, Navsari whereas, the weather data such as month-wise minimum of minimum temperature for October, December & February and the annual rainfall were obtained from the observatory situated at the college farm of Gujarat Agricultural University Campus, Navsari. The average yield based on 10 randomly selected trees is considered as dependent variable in the model. The data for the period 1981-1991 (11 years) were used to develop the model and the data for the period 1992 & 1993 were used to forecast/validate the yield. The crucial stages identified were panicle initiation stage in the month of October, flowering stage in the month of December and fruit setting and development stage in the month of February. As low temperature enhances the flowering and fruit setting of mango, the minimum of the minimum temperature can be regrarded as the instrumental variable in forecasting the yield. Thus variables included, were minimum of minimum temperature during the month of October, December & February, the annual rainfall and the age of the tree.

RESULTS AND DISCUSSION

Proposed Model-I

The all possible multiple regression analysis was carried out with above 5 variables and the 11 observations. The results obtained with $R^2 = 0.928$, Adjusted $R^2 = 0.86$, $Cp = 6$ and mean square error = 1379.64 and the significance of all variables except minimum of minimum temperature during fruit setting stage. However, the non-significant variables (X_3) was found to be very close to significant value.

$$Y = — \underset{}{1579.34} + \underset{(6.45)}{30.3^{*} X_1} + \underset{(6.71)}{35.9^{*} X_2} + \underset{(12.46)}{35.9 X_3} + \underset{(0.035)}{0.2^{*} X_4} + \underset{(2.95)}{16.9^{*} X_5}$$

Where, Y = Yield (in kg.) of mango (var. Kesar) tree.

X_1 = Minimum of Minimum temperature (°C) in the month of October.

X_2 = Minimum of minimum temperature (°C) in the month of December.

X_3 = Minimum of minimum temperature (°C) in the month of February.

X_4 = Annual rainfall (in mm)

X_5 = Age of the tree (in years)

Figures in the brackets denotes the standard error of the estimates.

Asterisk denotes the significance of regression coefficient at 5%.

Proposed Model-II

A subset above variables was obtained using Mallow's Cp criterion, R^2 adjusted R^2 and error mean square so as to get the forecast model for earlier prediction i.e., in the month of December. The Cp criterion is defined as $Cp = [RSS/s^2-(N-2p)]$; where RSS stands for residual sum of squares for the best subset being tested; `p' stands for the number of variables in the subset (including the intercept); $s^2$ stands for residual mean square based on regression using all independent variables and `N' stands for the number of observations. Netter and Wasserman (1974) described that the model which has the smallest Cp value, should be selected as best fit. Removing the main set from the set of all possible subset, we get the smallest value of Cp = 12.32 for the subset (X_1, X_2, X_4, X_5) as shown in the table 1 in bold printing. The subsets for the single variable were removed from the analysis as they appeared to be unwanted and the Cp values were also higher. Also the model has Maximum R^2, Maximum Adjusted R^2, minimum mean square error and minimum Cp value after removing the main set in Table 1. Thus the model given below was selected with $R^2 = 0.808$.

$$Y = -823.115 + \underset{(6.68)}{16.9^*}\, X_1 + \underset{(8.19)}{24.8^*}\, X_2 + \underset{(0.0142)}{0.142^*}\, X_3 + \underset{(4.28)}{15.03^*}\, X_4$$

Where,

Y = Yield (in kg.) of mango (var. Kesar) tree.

X_1 = Minimum of minimum temperature (°C) in the month of October.

X_2 = Minimum of minimum temperature (°C) in the month of December.

X_3 = Annual rainfall (in mm).

X_4 = Age of the tree (in years).

Figures in the brackets denotes the standard error of the estimates.

Asterisk denotes the significance of regression coefficient at 5%.

Neter and Wasserman (1974) showed that the model having little bias component tend to fall near the line Cp = p. This can be verified in our present study with the help of Table 1.

The model I was selected as it has Cp = p for the complete set of subsets of variables selected. It has Maximum R^2, Maximum Adjusted R^2, Minimum root mean square error and Minimum Cp value and there is no other subset of this set, where Cp = p can be attained. The model II is one of the subset with second Maximum R^2 & Adjusted R^2 and second Minimum root mean square error & Cp value. Although for model II Cp `is not equal' to p for the subset selected, it may be seen that Cp is more close to p as shown in the table 1 with bold printing. This explains that model II is the second-best-fit with the advantage of predicring mango (var. *Kesar*) yield, two months earlier than model I.

On the basis of proposed model I & II the yieid was forecasted and compared with the actual average yield obtained during the year 1992 and 1993, the years which were not included in the model formulation, as shown in Table 2.

TABLE 1

Subset showing the values of R^2, Adjusted R^2, p, Cp, RMSE.

Serial No.	*Subset Combination*	*R^2*	*Adjusted R^2*	*p*	*Cp*	*RMSE*
1	1,2	0.132	0.085	3	55.32	10402.78
2	1,3	0.123	0.096	3	55.79	10514.10
3	1,4	0.282	0.103	3	44.94	8611.17
4	1,5	0.287	0.109	3	44.58	8550.75
5	2,3	0.128	-0.090	3	55.62	10453.47
6	2,4	0.308	0.140	3	43.11	8296.05
7	2,5	0.261	0.070	3	46.36	88574.00
8	3,4	0.250	0.060	3	47.14	8991.90
9	3,5	0.317	0.150	3	42.45	8182.70
10	4,5	0.419	0.270	3	35.36	6959.97
11	1,2,3	0.153	-0.210	4	55.86	11599.97
12	1,2,4	0.451	0.210	4	37.68	8017.19
13	1,2,5	0.445	0.210	4	35.59	7606.12
14	1,3,4	0.283	-0.020	4	46.85	9825.53
15	1,3,5	0.339	0.050	4	42.92	9050.41
16	1,4,5	0.516	0.310	4	30.62	6625.79
17	2,3,4	0.316	0.020	4	44.54	9370.44
18	2,3,5	0.371	0.100	4	40.69	8611.74
19	2,4,5	0.603	0.430	4	24.57	5433.65
20	3,4,5	0.460	0.230	4	34.51	7392.76
21	1, 2, 3, 4	0.453	0.080	5	37.04	8747.99
22	1, 2, 3, 5	0.451	0.080	5	37.28	8769.14
23	1, 2, 4, 5	0.808	0.680	5	12.32	3063.40
24	1, 3, 4, 5	0.571	0.280	5	32.56	7718.59
25	2, 3, 4, 5	0.611	0.350	5	26.07	6224.57
26	1, 2, 3, 4, 5	0.928	0.860	6	6.00	1379.64

TABLE 2

Comparision of forecasted model based on proposed model I & II with observed yield based on average of 10 randomly selected trees of same set.

		Year 1992	*Year 1993*
Observed yield (kg/tree)		162	143
Forecasted yield	Model I	158.8	128.8
	Model II	127.6	105.5
Difference for	Model I	3.15	14.14
	Model II	34.4	37.46
Per cent different from observed yield	Model I	1.9	9.8
	Model II	21.2	26.2

The per cent different of forecasted yield based on model I from observed average yield of mango per tree is 1.9% and 9.8% for the year 1992 and 1993 respectively, indicated the satisfactory fit of the data. Again per cent difference of forecasted yield based on model II for the year 1992 and 1993 were 21.2% and 26.2%, indicated more deviation in comparision to model I (Table 2). Since model II includes temperature parameter only up to the month of December, it was regarded as earlier forecast model in comparision to model I.

The conlusions emerging from the study are : (1) Out of the two models developed in this study the model I ($Y = -1579.34 + 30.3 X_1 + 35.9 X_2 + 35.9 X_3 + 0.2 X_4 + 16.9 X_5$) that requires minimum of minimum temperature in the month of October (X_1), December (X_2), & February (X_3), the annual rainfall (X_4) and the age of the tree (X_5), is suitable for prediction at fruit setting stage and model II ($Y = -823.155 + 16.9 X_1 + 24.8 X_2 + 0.142 X_3 + 15.03 X_4$ that requires minimum of minimum temperature in the month of October (X_1), & December (X_2), the annual rainfall (X_3) and the age of the tree (X_4), is suitable for yield prediction flowering stage i.e., in the month of December. Thus an early forecast of yield by two month can be given using model II. (2) Since age of the tree is one of the significant variable in both the models, the age wise prediction will be more useful in planning the future needs in perennial plantation crops.

ACKNOWLEDGEMENT

The authors convey their deep scence of gratitude to Dr. B M. Patel, Professor of Horticulture, Regional Fruit Research Station and Observatory, College Farm, Gujarat Agricultural University, Navsari for providing yield and weather data respectively.

REFERENCES

Brown, J. T. and Vanderlip, R. L. (1969). Relations of Palmer drought variables to yield of wheat, *Agronomy Journal*, **61** : 95-98.

Cuthbert, Daniel and Fred, S. Wood (1980). Fitting Equations to Data (2nd Ed.) John Wiley and Sons, New York.

Draper, N. R. and Smith, H. (1966). Applied Regression Analysis, John Wiley and Sons, New York.

Gangopadhyaya, M. and Sarkar, R. P. (1964). Curviliner study on effect of weather on growth of sugarcane, *Indian Journal of Mereorology and Geophysics*, **15** : 215-226.

Huda, A. K. S., Virmani, S. M. Sekaran, J. G. (1985). Simple model for predicting sorghum grain yield using environmental factors. *Jour. Ind. Soc. Ag. Statistics*, **Vol. XXXVII No. 2** :184-191.

LeDuc, S., Sakamoto, C., Strommen, N. and Steyeart, L. (1980). Some problems associated with using climate crop yield model in an operational system; an overview, *Biometeorology*, **7** part 2 : 104-114, suplement of *International Journal of Biometeorology* **24**.

Neter, J. and Wasserman, W. (1974). Applied Linear Statistical Models : Regression, Analysis of variance and Experimental Designs, Richard D. Irwin. Inc. Homewood Illions, 60430 USA.

Snedecor, G. W. and Cochran, W. G. (1967). Statistical Methods (6th Ed.). Oxford & IBH Publishing Co., Calcutta.

Thompson, L. M., (1969). Weather and technology in production of corn in the U.S. cornbelt, *Agronomy Journal*, 61 : 453-456.