Reprinted from the Indian Journal of Meteorology and Geophysics, Vol. 15, No. 2, April 1964.

633·6:66·012:551·5

A Curvilinear Study of Yield with reference to Crop-characteristics — Sugarcane

M. GANGOPADHYAYA and R. P. SARKER

Meteorological Office, Poona

(*Received 25 July 1962*)

ABSTRACT. Curvilinear technique has been utilised to find out the relation of yield of sugarcane with the crop-characteristics, namely, height, girth and the number of canes per clump after stabilization. The study indicates that the yield, in general, increases with height and girth and may become stationary after these characteristics exceed certain values. It also shows that there should, in general, be an optimum number of canes per clump after stabilization so as to give maximum yield. This value varies from 2 to 3. A set of curves have been determined by this method which can be used to predict yield about two or three months before crop is ready for harvest. The significant improvement in the correlations of yields estimated from these curves with the actuals as compared with linear multiple correlation is encouraging. The curvilinear regression accounts for about 80—90 per cent of the variations in yield whereas the linear regression can account for 35—70 per cent of the variations.

1. Introduction

As the crop-characteristics like germination, tillering, elongation and ripening etc, and consequently the ultimate yield of a crop are influenced by the weather elements it experiences during its life-cycle, it may be possible to work out a prediction formula for yield from the weather elements. But such a prediction formula would probably call for the meteorological elements right up to the time of harvest. An alternative is to use the plant itself as an integrator of weather effects and to base a prediction formula for yield on the easily recognizable and measurable state of the crop during its growth (Keen 1940). Since it is not quite justified to assume a linear relationship between yield and any of the crop-characteristics an attempt has been made in the present study to find out the curvilinear relations between the yield of sugarcane and the crop characteristics, taken one by one, while eliminating the net effects of the others and to forecast the yield using these results.

If $f_2(X_2), f_3(X_3), f_4(X_4)$........etc are the effects of the crop characteristics X_2, X_3, X_4 etc on the yield X_1, then the predicted yield is given by

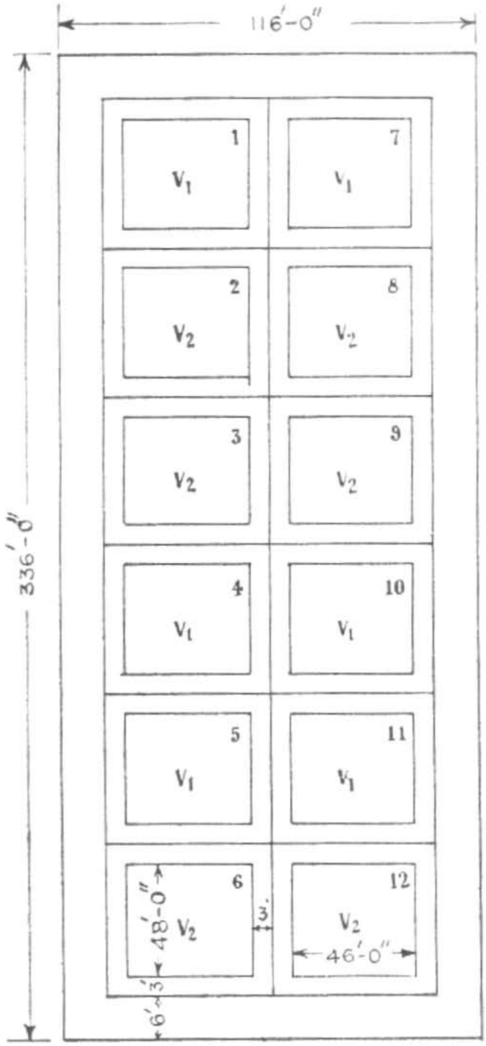
$$X_1 = f_2(X_2) + f_3(X_3) + f_4(X_4) + \ldots \qquad (1)$$

The partial regression curves $f_2(X_2)$.... etc between crop characteristics and yield have been found by the graphical methods of successive approximations (Ezekiel and Fox 1959) in preference to fitting mathematical expressions as the former involves no prior assumption as to the shapes of the curves.

2. Observations

The data made use of in the present study have been collected under the All India Co-ordinated Crop-Weather Scheme on Sugarcane. At each of the stations co-operating in the scheme two varieties of sugarcane are grown in 12 plots (six for each variety), each measuring 46 ft × 48 ft. The general layout is shown diagrammatically in Fig. 1.

It will be seen from the layout that all around the entire block a border of 6 ft width has been kept where sugarcane is grown. This has been done to eliminate the border effect on the plots in which growth observations are taken. As a further precaution against border effects each plot has border of 3 ft width wherein is grown the sugarcane of the same variety as is under study in that plot. This eliminates any possible effect of one variety upon the other.



Lay out of periodical growth observation plots, V_1 & V_2 are different varieties

Fig. 1

For observation, each plot is divided into two sub-plots by an imaginary line parallel to the rows of the plants. Periodical crop-observations are taken from three "sampling units" of plants chosen at random from each sub-plot. A "sampling unit" consists of two clumps of plants lying within a distance of 8 ft measured along any row. Thus for each variety there are 36 sampling units comprising of 72 clumps of plants. To ensure randomness each sample is chosen afresh for each and every observation using the system of random numbers developed by Fisher and Yates (1943) and measurements on plants are taken by placing the sampling rod, accordingly. Average of the six samples so taken at random from a plot is fairly representative of the plot.

The following developmental observations of sugarcane are recorded on the clumps selected—

(*a*) Germination counts daily from the date of appearance of first seedling till the germination appears to be complete, *i.e.*, for about 3 weeks. The plants that come up within a sampling unit, *i.e.*, within that length of an 8-ft rod are counted daily and then knowing the number of eye-buds planted per acre, the germination percentage is computed.

(*b*) Total number of canes in the clump, *i.e.*, the mother cane plus the tillers.

(*c*) *Leaf*—(*i*) Number of fully expanded green leaves in the tallest cane of the clump, (*ii*) Measurements of the breadth of each of the leaves of the above cane at the widest point and the length from tip to the collar.

(*d*) *Height of the cane*—The height of the tallest cane of the clump is measured from the surface of the ground (proper adjustment being made for the depth of the furrow or the height of the ridge with reference to the surface of the ground) to the highest visible transverse leaf mark.

(*e*) *Circumference*—The circumference of the tallest cane of the clump is measured by a cloth tape at 1/4, 1/2 and 3/4 of the height.

(*f*) *Brix reading*—This reading is taken with the help of a hand refractometer by extracting juice with a juice sampler. This observation is taken once every fortnight from about three months before harvest and the maximum of these values as determined by a smooth curve is taken.

(*g*) *Weight of canes*—Weight of canes per clump and also of canes falling within the sampling unit are taken at the time of harvest.

Observations on items (*b*) to (*e*) are taken weekly. The details of the procedure for taking these observations are indicated in the *Agrimet. Tech. Circular* No. 50 of the India Meteorological Department.

The data collected as above are now available in respect of Poona (Lat. 18°32′ N, Long. 73°51′E, 563 metres a.s.l.), Shakarnagar (Lat. 18°38′N, Long. 77°45′ E, 381 metres a.s.l.), and Samalkot (Lat. 17°03′N, Long. 82°13′E, 9 metres a.s.l.) for 8 to 14 years only. Data for other stations are of much shorter series.

3. Crop-characteristics used

The crop-characteristics that are expected to have a bearing on the ultimate yield of sugarcane are (*i*) germination percentage, (*ii*) maximum tillers and tiller mortality, (*iii*) the elongation, (*iv*) the girth and (*v*) the brix-reading which is an index of maturity.

A preliminary examination has shown that the germination percentage and the brix-reading do not have any appreciable influence on the final yield. Instead of the two elements in (*ii*) above the number of shoots (tillers plus the mother cane) per clump after stabilization has been taken. The crop-characteristics thus considered for the present study are—

(*i*) Height when it attains its maximum (X_2),

(*ii*) No. of canes per clump after stabilization (X_3),

(*iii*) The girth when it becomes stationary (X_4).

4. Procedure

For the freehand drawing of the curves some conditions have been imposed on each curve after keeping in view the logical nature of relations between yield and growth factors. They are—

(*i*) Yield will generally increase with height and may become asymptotic after the height exceeds some value. So X_1 will generally be an increasing function of X_2.

(*ii*) For maximum yield there will be an optimum number of shoots per clump, for if the number of shoots exceeds a certain value, yield will decrease due to competition among the plants. So the curve between X_1 and X_3 will have a single maximum.

(*iii*) Yield will generally increase with girth and may become asymptotic if the girth exceeds some value. So X_1 will generally be an increasing function of X_4.

Based on these conditions, graphical approximations have been made successively to find the curvilinear relations between yield and crop-characteristics for each variety and for each station. In the following paragraphs the procedure adopted in this paper for curvilinear study is elaborated for variety POJ-2878 at Poona. The results of similar studies with respect to various varieties at other stations have been summarised in a separate section.

TABLE 1

Crop-Characteristics

Station : Poona Variety : POJ 2878

Year	Height (cm) X_2	No. of canes per clump X_3	Girth (cm) X_4	Yield (tons/acre) X_1	Estimated yield X_1'	X_1-X_1' $=Z'$
1946-47	425	2·8	9·5	47·0	52·7	—5·7
1947-48	395	2·2	9·0	58·3	54·7	3·6
1948-49	347	2·1	8·5	43·9	41·5	2·4
1949-50	347	1·7	8·8	46·6	46·4	0·2
1950-51	352	2·1	8·6	44·6	42·8	1·8
1951-52	324	1·8	8·6	38·3	42·4	—4·1
1952-53	327	2·1	8·9	45·4	43·9	1·5
1953-54	308	2·2	8·5	33·4	38·3	—4·9
1954-55	305	2·3	9·1	45·0	43·2	1·8
1955-56	355	2·6	9·1	49.2	45·1	4·1
1956-57	355	3·1	9·4	46·0	45·3	0·7
1957-58	315	2·5	9·0	37·3	41·9	—4·6
1958-59	345	2·6	8·5	40·0	38·8	1·2
1959-60	316	2·6	8·6	40·6	37·3	3·3

Step I—"First approximation" net regression curves for variety POJ2878 at Poona

The yield and crop-characteristics data for variety POJ 2878 at Poona are given in Table 1.

Using Table 1, the linear multiple regression equation

$X_1=a_{1\cdot 234}+b_{12\cdot 34}\ X_2+b_{13\cdot 24}\ X_3+b_{14\cdot 23}\ X_4$

works out as

$X_1=.070\ X_2-5\cdot 236\ X_3+9\cdot 434\ X_4-51\cdot 967$ (2)

with significant multiple correlation $R_{1\cdot 234}=0\cdot 825$. Estimates X_1' from (2) and the residuals $Z'=X_1-X_1'$ are also given in the two extreme right columns of Table 1.

Step II

To find the regression curves between X_1 and X_2, the dependent variable X_1 is adjusted for the deviation from the means of all independent variables except X_2, *i.e.*, for X_3 and X_4 in this case, and a scatter diagram is drawn between these adjusted values of X_1 and the X_2. This procedure amounts to plotting X_2 values as abscissae and the residual Z' values as ordinates with the net regression line

$$X_1=0\cdot 070\ X_2-5\cdot 236\ M_3+9\cdot 434\ M_4-51\cdot 967$$
$$=0\cdot 070\ X_2+19\cdot 828 \quad (3)$$

as zero base; M_i being used to denote the mean value of X_i (i=1, 2, 3 and 4). The residuals are then averaged for selected group values of X_2 and are similarly plotted. A free hand curve is then drawn through the several group averages, as far as possible, as is consistent with a smooth curve, keeping in mind the limiting condition to the shape. This curve is the first approximation to the curvilinear function $X_1=f_2(X_2)$.

In exactly the same manner the first approximation curves for the functions $X_1=f_3(X_3)$, $X_1=f_4(X_4)$ are drawn and are given in Figs. 3 and 4.

TABLE 2

Residuals from linear regression Z'	Residuals from first approximate curve Z''
−5·7	−7·2
3·6	0·3
2·4	2·4
0·2	−0·7
1·8	1·1
−4·1	−2·3
1·5	1·0
−4·9	−1·2
1·8	4·8
4·1	1·1
0·7	−0·5
−4·6	−4·0
1·2	1·2
3·3	5·2

Step III—Estimates of X_1 from the first approximation curves

Denoting the first approximate curves by $f_2'(X_2)$, $f_3'(X_3)$ and $f_4'(X_4)$, the estimates X_1'' of X_1 from the first approximation curves are worked out by the equation

$$X_1'' = a'_{1\cdot 234} + f_2'(X_2) + f_3'(X_3) + f_4'(X_4) \quad (4)$$

where,

$$a'_{1\cdot 234} = M_1 - \frac{\Sigma[f_2'(X_2) + f_3'(X_3) + f_4'(X_4)]}{n} \quad (5)$$

The residuals $Z'' = X_1 - X_1''$ are then computed and given in Table 2 side by side with Z'.

It would be seen from Table 2 that the new residuals are in general smaller than the previous ones and there are 8 cases in which the new residuals are smaller and 4 in which they are larger than the previous ones. An accurate comparison can be had by comparing the standard errors of estimates from the two sets of residuals. The standard error of the first set of estimates is 3·29 tons/acre while that for the second set is 3·12 tons/acre. Apparently the new estimates come nearer the observed values, on the average, than did the first set of estimates.

TABLE 3

Z''	Z'''
−7·2	−3·0
0·3	1·9
2·4	0·5
−0·7	−0·5
1·1	−1·0
−2·3	−3·6
1·0	−0·5
−1·2	−2·1
4·8	4·0
1·1	1·8
−0·5	1·9
−4·0	−4·2
1·2	0·6
5·2	4·6

Step IV—Second approximate net regression curves

The first approximating curves from Figs. 2 to 4 are first drawn. Each of the last residuals Z'' is then plotted as a deviation just as before, except that the residuals are now plotted as deviations from the regression curves, instead of from the regression lines. Averaging the residuals values, the 2nd approximate curves are drawn, exactly as before and these are given in Figs. 5 to 7. The new estimates X_1''' are worked out from

$$X_1''' = a''_{1\cdot 234} + f_2''(X_2) + f_3''(X_3) + f_4''(X_4) \quad (6)$$

and the new residuals $Z''' = X_1 - X_1'''$ are given in Table 3. The residuals on the average have decreased and there are 8 cases in which they have decreased and 6 in which they increased. The standard error of the 3rd set of estimates has considerably decreased to 2·59 tons/acre as compared to 3·12 tons/acre of the second set of estimates. Apparently the third set of estimates come nearer the actuals than the 2nd set of estimates.

Step V — Further successive approximations

Using the residuals from the 2nd approximations and proceeding exactly as before the 3rd approximation curves are drawn and the estimates and residuals are computed. The process is continued until the

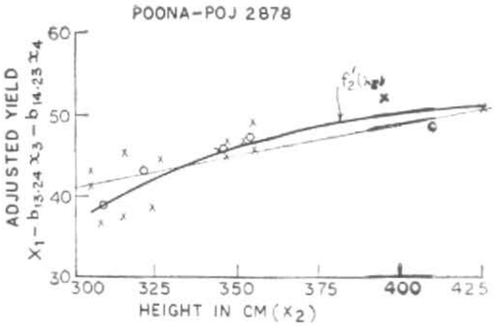


FIG. 2

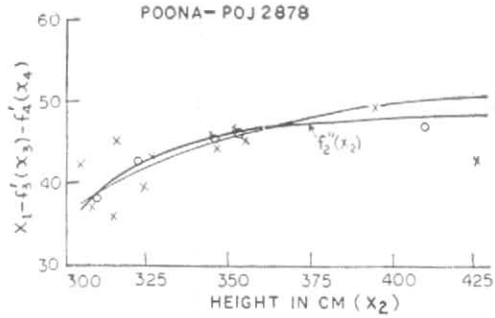


FIG. 5

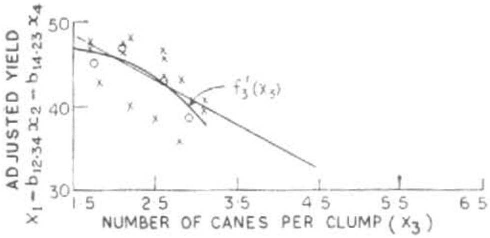


FIG 3

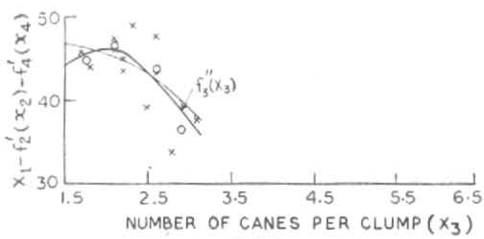


FIG. 6

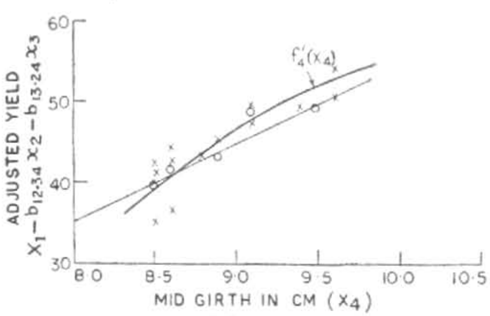


FIG 4

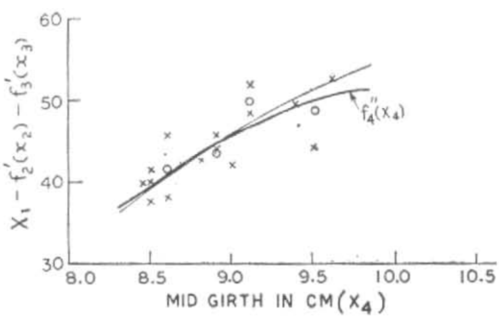


FIG. 7

Figs. 2—7

standard errors of estimate show a steady value or reach a minimum. In this particular case the standard error was the minimum at the 3rd approximation stage and the final curves are given in Figs. 8 to 10 and the residuals $Z''''=X_1-X_1''''$ are given in Table 4.

It would be seen from the above table that the residuals (Z'''') from 3rd approximation have decreased in 7 cases and increased

TABLE 4

Z'''	Z''''
—3·0	—1·7
1·9	2·1
0·5	0·2
—0·5	—0·2
—1·0	—1·6
—3·6	—2·4
—0·5	—1·9
—2·1	—1·1
4·0	3·7
1·8	—0·2
1·9	2·2
—4·2	—5·1
0·6	0·6
4·6	4·9
S.E. 2·59	S.E. 2·53

TABLE 5

Average yield of sugarcane crop with varying height keeping influences of girth and number of canes constant

Height X_2	Readings from final curves $f_2(X_2)$	Constant M_1–$M_{f(2)}$	Average yield $F_2(X_2)$
305	36·0	0·32	36·3
315	39·8	0·32	40·1
325	43·0	0·32	43·3
335	44·7	0·32	45·0
345	46·4	0·32	46·7
355	47·1	0·32	47·4
365	47·5	0·32	47·8
375	47·7	0·32	48·0
385	47·9	0·32	48·2
395	47·9	0·32	48·2
405	47·9	0·32	48·2
415	47·9	0·32	48·2
425	47·9	0·32	48·2

in 6 as compared to those from the 2nd approximations and the standard error of estimates have decreased to 2·53 tons/acre from 2·59 tons/acre. So the estimates from the 3rd approximation curves come nearer the observed values than the previous estimates.

5. Inferences from the curves

The final regression curves shown in Figs. 8 to 10 represent the net relation between yield and each crop-characteristic with the net variation associated with the other crop-characteristics held constant.

Following are the inferences—

(*i*) Yield slowly increases with height and remains practically stationary as the height exceeds 375 cm,

(*ii*) Yield increases as the number of canes (shoots plus mother cane) per clump increases and decreases as the number exceeds 2·1. The optimum number of canes per clump is 2·1.

(*iii*) Yield increases with girth and remains practically stationary as the mid-girth exceeds 9·5 cm.

6. Use of the curves

The curves given in Figs. 8 to 10 have been utilised to estimate the yield in the following way.

The yield (X_1) was estimated from the most important factor height (X_2) by the equation

$$X_1'=F_2(X_2)=f_2(X_2)—M_{f(2)}+M_1$$
$$=f_2(X_2)—43\cdot65+43\cdot97=f_2(X_2)+0\cdot32 \quad (7)$$

[$M_{f(2)}$ is the mean of the values obtained from the final curve $f_2(X_2)$]

The values estimated from Eqn. (7) for different values of X_2 are given in Table 5.

The estimated values are then adjusted for the other crop-characteristics. The *correction* for the number of canes per clump (X_3) is made by the equation—

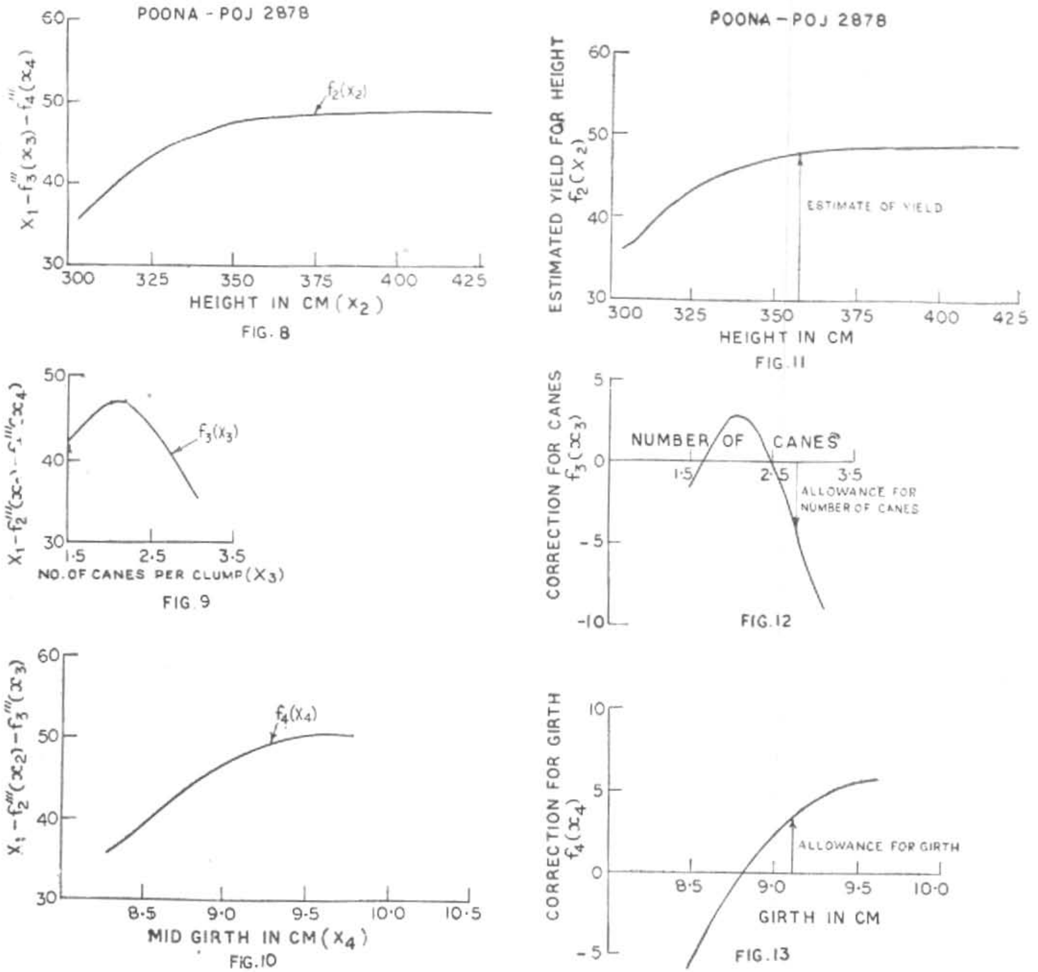
$$X_1'=F_3(x_3)=f_3(X_3)—M_{f(3)}$$
$$=f_3(X_3)—44\cdot09 \quad (8)$$

These *corrections* for the different values of the number of canes are given in Table 6.

Corrections for girth are found by the equation–

$$X_1'=F_4(x_4)=f_4(X_4)—M_{f(4)}$$
$$=f_4(X_4)—44\cdot43 \quad (9)$$

These corrections are given in Table 7.



Figs. 8—13

The yield values may now be estimated based on all the crop-characteristics by the equation—

$$X_1' = F_2(X_2) + F_3(x_3) + F_4(x_4) \qquad (10)$$

The final results of our curvilinear study, as represented in Tables 5 to 7 have been represented graphically in simple forms in the Figs. 11 to 13.

With the help of these curves given in Figs. 11 to 13, the yield can be predicted with a fair degree of accuracy after the three growth characteristics considered have

TABLE 6

Correction to yield due to differences in number of canes

No. of canes X_3	Readings from final curve $f_3(X_3)$	Constant $-M_{f(3)}$	Correction to expected yield $F_3(x_3)$
1·5	42·4	−44·09	−1·7
1·7	44·6	−44·09	0·5
1·9	46·3	−44·09	2·1
2·1	46·9	−44·09	2·8
2·3	45·8	−44·09	1·7
2·5	44·0	−44·09	−0·1
2·7	41·2	−44·09	−2·9
2·9	38·0	−44·09	−6·1
3·1	35·3	−44·09	−8·8

TABLE 7

Correction to yield for different values of girth

Values of girth X_4	Readings from final curve $f_4(X_4)$	Constant $-M_{f(4)}$	Correction to expected yield $F_4(x_4)$
8·5	39·0	−44·43	−5·4
8·6	40·9	−44·43	−3·5
8·7	42·8	−44·43	−1·6
8·8	44·2	−44·43	−0·2
8·9	45·6	−44·43	1·2
9·0	46·6	−44·43	2·2
9·1	47·8	−44·43	3·4
9·2	48·5	−44·43	4·1
9·3	49·2	−44·43	4·8
9·4	49·9	−44·43	5·5
9·5	50·0	−44·43	5·6
9·6	50·0	−44·43	5·6

stabilised. This occurs two to three months before the crop is ready for harvesting. The estimated yields from these curves for the fourteen years' data utilised here were all within 10 per cent of actuals except two in which the deviations were 12 and 13 per cent of the actuals and most of the estimates were within 5 per cent. The correlation between the estimates from the final curves and the actuals has come out to be 0·903 whereas it was 0·825 for the estimates from the linear regression, thus indicating that the curvilinear multiple regression equation accounts for about 82 per cent of the variation in the yield as compared to about 68 per cent in the case of linear multiple regression.

7. Results of the study on varieties grown at other places

So far we have detailed the procedure of curvilinear study to bring out the results of yield and crop-characteristics for variety POJ-2878 at Poona. The results of similar study on the other cases are given below—

(*a*) *Variety CO-419 at Poona*—The results are based on only 9 years' data. The relations between the crop-characteristics and yield are given in Figs. 14 to 16. Following are the inferences—

(*i*) Yield increase with height and remains practically stationary for the height beyond 340 cm.

(*ii*) Yield remains almost stationary for the number of canes upto 2·7 per clump and then increases with number of canes and decreases very steeply as the number of canes per clump exceeds 3·0. The optimum number of canes per clump for yield is 3·0.

(*iii*) Yield increases with girth and remains stationary for the girth beyond 9·5 cm.

The final curves for estimating yield from height and for making further corrections for number of canes and girth are given in Figs. 17 to 19. The estimated yield from these curves for the 9 years were all within 5 per cent of the actuals. Also the correlation between the estimates from the final curves and the actuals has come out to be 0·911 whereas it was 0·576 only for the estimates from the linear regression. This indicates that the curvilinear regression accounts for about 83 per cent of the variation in yield as compared to only 33 per cent by the linear regression.

(*b*) *Variety POJ-2878 at Shakarnagar (Black Soil)*—The results are based on only 9 years' data. The relations between yield and the crop-characteristics are given in Figs. 20 to 22. Following are the inferences—

(*i*) Yield increases with height

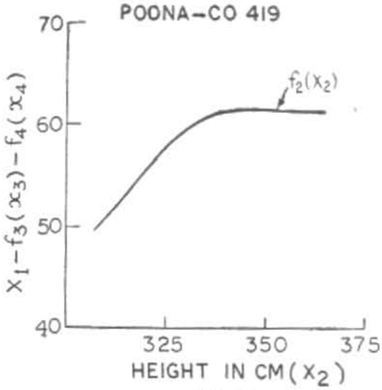


FIG. 14

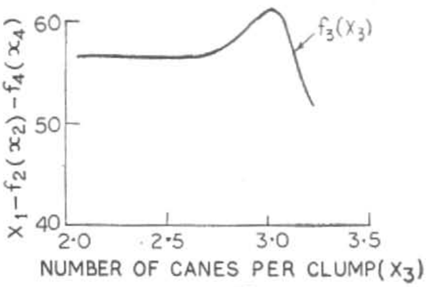


FIG. 15

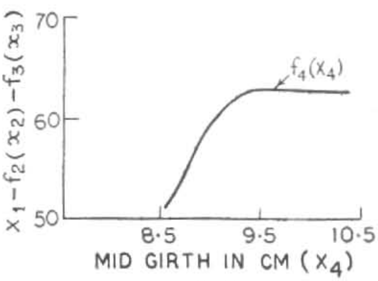


FIG. 16

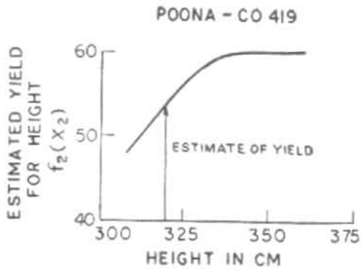


FIG. 17

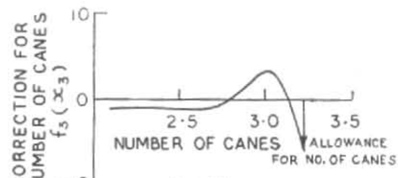


FIG. 18

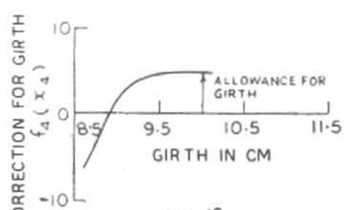


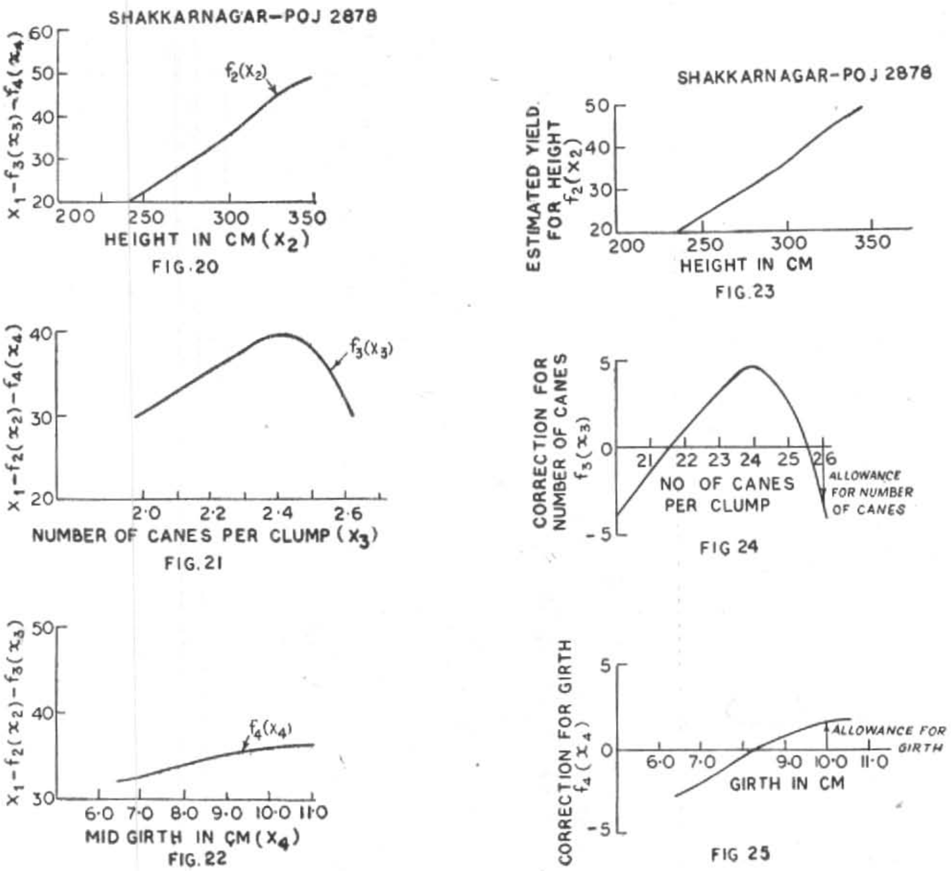
FIG 19

Figs. 14—19

(*ii*) Yield increases with the number of canes and decreases again when the number of canes per clump exceeds 2·4. The optimum number of canes per clump for yield is 2·4

(*iii*) Yield very slowly increases with girth

The final curves for estimating yield from height and for making further corrections for number of canes and girth are given in Figs. 23 to 25. The yields estimated from these curves for the 9 years were all within 10 per cent of the actuals except for one year in which the estimated yield was within



Figs. 20—25

15 per cent of the actual. Also correlation between the estimates from the final curves and the actuals has come out to be 0·957 whereas it was 0·859 for the estimates from the linear regression. This suggests that about 92 per cent of the total variation in yield can be accounted for by the curvilinear regression whereas the linear regression could account for only 74 per cent.

(c) *Variety CO-419 at Shakarnagar* (*Black Soil*)—The results are based on only 8 years' data.

In this case the linear multiple correlation coefficient between yield and the crop-charcteristics worked out to be 0·973 and all the estimates from the linear regression were within 4 per cent of the actuals. A further curvilinear study did not improve the result.

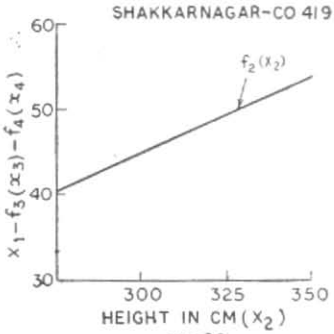


FIG. 26

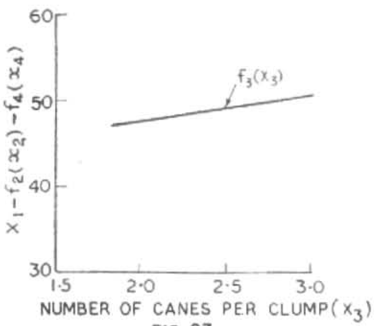


FIG. 27

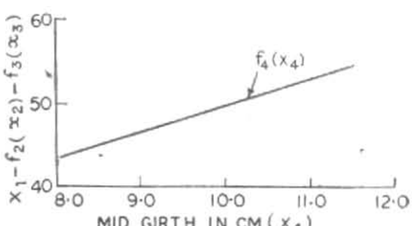


FIG. 28

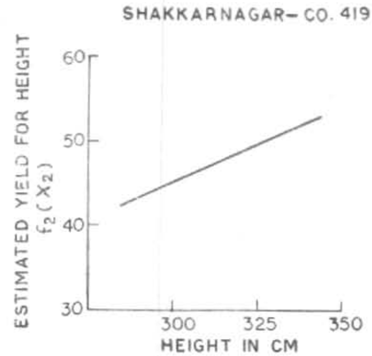


FIG. 29

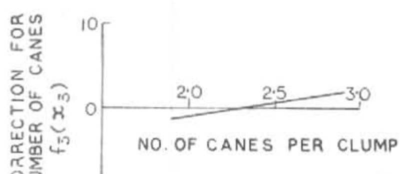


FIG. 30

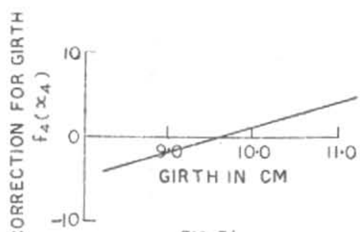
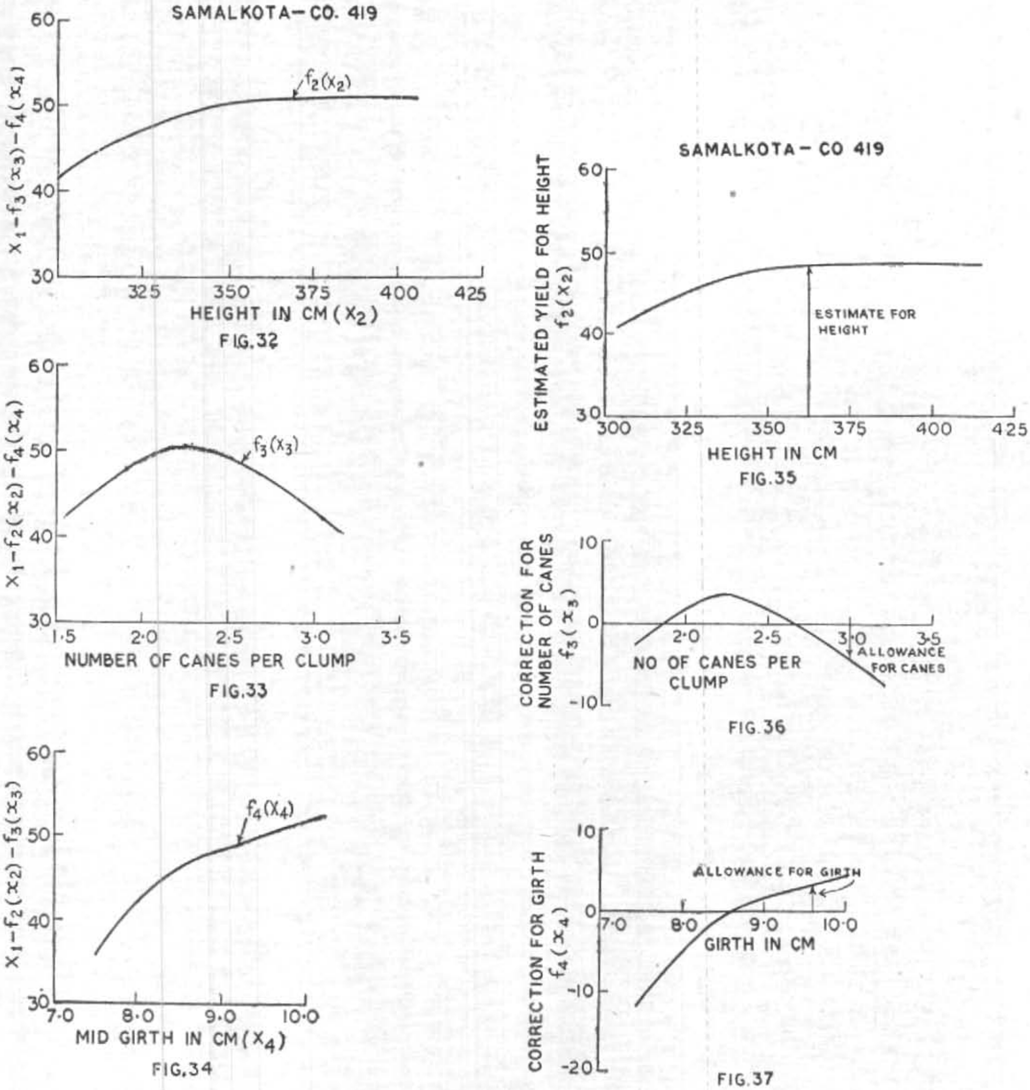


FIG. 31

Figs. 26—31

Apparently, with the limited available data, the relations of yield with the crop-characteristics in their corresponding ranges could be best represented by straight lines. These relations are represented in Figs. 26 to 28. The final curves for estimating yield from height and for making corrections for other two crop-characteristics are given in Figs. 29 to 31. The linear regression accounts for about 95 per cent of the variation in yield.

(*d*) *Variety CO-419 at Samalkota*—The results are based on 10 years' data only. The



Figs. 32—37

relations between yield and the crop-characteristics are given in Figs. 32 to 34. Following are the inferences—

(*i*) Yield slowly increases with height and remains practically constant as the height exceeds 360 cm

(*ii*) Yield increases with the number of canes and then decreases as the number of canes per clump exceeds 2·3. The optimum number of canes per clump for yield is 2·3.

(*iii*) Yield increases with girth

The final curves for estimating yield from height and for making further corrections for the other two crop-characteristics are given in Figs. 35 to 37. The estimated yields for the 10 years from these curves are all within 6 per cent of the actuals except one in which case, the estimate was within 14 per cent of the actual. The correlation betweeen the estimates from the final curves and the actuals worked out to be 0·881 wheres it was 0·711 for the estimates from the linear multiple regression. The curvilinear regression thus accounts for about 78 per cent of the variation in yield whereas the linear regression accounts for only about 50 per cent.

8. Conclusion

The curvilinear study of yield and crop-characteristics has brought out how the yield of sugarcane crop at different stations are related to the crop-characteristics. It has also brought out what should be the optimum number of canes per clump after stabilisation for giving maximum yield. Such information on optimum value could not be had from a linear regression study.

The curvilinear study has also enabled to find out some curves which can be used to predict the yield much more accurately than by linear regression method as indicated by the high correlations of the yield estimated from these curves with the actuals. With a long series of data the prediction curves would certainly give still more reliable results.

9. Acknowledgement

The authors have great pleasure to record their grateful thanks to Shri J. R. Banerjee for help in computation.

REFERENCES

Ezekiel, M. and Fox, Karl A.	1959	*Methods of Correlation and Regression Analysis*, pp. 204-248.
Fisher, R. A. and Yates, F.	1943	*Statistical Tables for Biological, Agricultural and Medical Research*, pp. 90-95.
Keen, B. A.	1940	Weather and Crops –*Quart. J. R. met. Soc*,, **66**.

M/P(N)6DGOB—280—2-7-64—GIPS