A STUDY ON CURVILINEAR RELATIONSHIP OF YIELD AS PRODUCT OF FUNCTIONS OF CROP CHERACTERISTICS SUGARCANE

BY A. N. BASU
Meoterological Office, Poona
(Received in October, 1966)

1. INTRODUCTION

Curvilinear technique has already been used by Gangopadhyay and Sarker (1964) to find out the relation of yield of sugarcane with the crop-characteristics namely, height, mid-girth and the number of canes per clump. There they have considered the simplest possible relation between yield and the function of the crop-characteristics namely,

$$y=f_1(x_1)+f_2(x_2)+f_3(x_3) \qquad \ldots(1)$$

where

$$f_1(x_1),\ f_2(x_2),\ f_3(x_3)$$

are the effects of x_1, x_2, x_3 respectively on yield y ; x_1, x_2, x_3 being the height, mid-girth and number of canes per clump.

It may not be quite justified to assume that yield is the sum of functions of the crop-characteristics and each of these functions is a function of single crop-characteristics only. That is, yield may also depend upon the joint effect of the crop-characteristics. In the present study attempt has been made to include the joint functions, such as

$$f_{12}(x_1, x_2),\ f_{23}(x_2, x_3),\ f_{31}(x_3, x_1),\ f_{123}(x_1, x_2, x_3),\ f_{112}(x_1^2, x_2)\ldots$$

to predict yield. For this, the predicted yield has been taken as

$$y=f_1(x_1)\, f_2(x_2)\, f_3(x_3) \qquad \ldots(2)$$

In this case, it may be transformed to the relation of the type of equation (1) by using the relation

$$\text{Log } y = F_1(\log x_1) + F_2(\log x_2) + F_3(\log x_3) \quad \ldots (3)$$

or

$$Y = F_1(X_1) + F_2(X_2) + F_3(X_3) \quad \ldots (4)$$

where

$$Y = \log y,\ X_1 = \log x_1,\ X_2 = \log x_2 \text{ and } X_3 = \log x_3.$$

The partial regression curves $F_1(X_1)$, $F_2(X_2)$ and $F_3(X_3)$ of Y on X_1, X_2 and X_3 respectively have been found by graphical methods of successive approximations (Ezekiel and Fox 1959, Gangopadhyay and Sarker, 1964).

2. Variable used

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

Gangopadhyaya and Sarker (1964) have shown by working out correlation coefficients between the yield and crop characteristics that maximum elongation (height), mid-girth when stationary and number of canes per clump after stabilization contribute significantly to the yield and that the germination percentage and brix-reading do not have any appreciable influence on the yield. So the three variables, *viz.*, maximum elongation (height), mid-girth when stationary and number of canes per clump after stabilizations have been used in the present study.

3. Procedure

The data collected at Poona from 1946-47 to 1962-63, under the Crop-Weather Scheme—Sugarcane have been used. No data after 1962-63 are available due to failure of irrigation following the breach of the Khadakwasla dam. Table 1 shows the data of the two varieties CO-419 and POJ-2878, grown at Poona. The procedure adopted is described in the following paragraphs using the data of the variety POJ—2878.

"First Approximation" net regression curves :

Table 2 shows $\log x_1$, $\log x_2$, $\log x_3$ and $\log y$. We denote this $\log x_i$ as X_i and $\log y$ as Y. The multiple regression equation of Y on X_1, X_2 and X_3

$$Y = a_{Y.123} + b_{Y1.23}\, X_1 + b_{Y2.31}\, X_2 + b_{Y3.12}\, X_3$$

TABLE 1

Crop-Characteristics V_1—*CO—419*

Station—Poona V_2—*POJ—2878*

Crop year	*Hight in cm.* (x_1)		*Mid-girth in cm.* (x_2)		*No. of canes per clump* (x_3)		*Yield in tons acre* (y)	
	V_1	V_2	V_1	V_2	V_1	V_2	V_1	V_2
1946—47	—	425	—	9·5	—	2·8	—	47·0
47—48	310	395	9·6	9·6	2·8	2·2	53·5	58·3
48—49	206	347	8·8	8·5	2·3	2·1	42·4	43·9
49—50	342	347	8·6	8·8	2·1	1·7	54·7	46·6
50—51	335	352	8·9	8·6	2·6	2·1	60·0	44·6
51—52	305	324	9·2	8·6	2·0	1·8	59·5	38·3
52—53	335	327	9·3	8·9	2·3	2·1	60·1	45·4
53—54	—	308	—	8·5	—	2·2	—	33·4
54—55	313	305	9·4	9·1	3·0	2·3	61·3	45·0
55—56	—	355	—	9·1	—	2·6	—	49·2
56—57	360	355	10·0	9·4	3·2	3·1	58 9	46·0
57—58	325	315	8·8	9·0	3·1	2·5	54·3	37·3
58—59	353	345	8·6	8·5	3·0	2·6	54·6	40·0
59—60	316	316	8·7	8·6	2·8	2·6	48·0	40·6
60—61	353	365	9·3	9·6	3·0	2·9	55·9	47·4
61—62	241	231	8·4	8·3	2·4	2·2	32·6	29·0
62—63	282	287	8 6	8·5	2·9	2·8	44·3	33·1

has been found to be

$$Y = -1{\cdot}68483 + 0{\cdot}70960\ X_1 + 1{\cdot}70678\ X_2 - 0{\cdot}26605\ X_3 \quad \ldots(5)$$

with significant correlation

$$R_{Y.(X_1X_2X_3)} = 0{\cdot}9069$$

The significance of the three regression coefficients 0·70960, 1·70678 and −0·26605 have been tested. It has been seen that the first two coefficients are highly significant and the third is significant at 5% level of significance.

Estimates Y' of Y from (5) and the residuals $z'=Y-Y'$ are given in Table 2. The net regression line of Y on X_1 is

$$Y=-1\cdot68483+0\cdot70960\ X_1+1\cdot70678\ \overline{X}_2-0\cdot26605\ \overline{X}_3$$

or $$Y=-0\cdot16532+0\cdot70960\ X_1 \quad \ldots(6)$$

TABLE 2

Estimated yield and residuals

$Y=\log y$	X_1	X_2	X_3	Y'	$Z'=Y-Y'$
1·67210	2·62839	0·97772	0·44716	1·73006	−0·05796
1·76567	2·59660	0·98227	0·34242	1·74314	0·02253
1·64246	2·54033	0·92942	0·32222	1·61837	0·02409
1·66839	2·54033	0 94448	0·23045	1·66850	−0 00011
1·64933	2·54654	0·93450	0·32222	1·63145	0·01788
1·58320	2·51054	0·93450	0 25527	1·62373	−0·04053
1·65706	2·51455	0·94939	0·32222	1 63416	0·02290
1·52375	2·48855	0·92942	0·34242	1·57625	−0·05250
1·65321	2·48430	0·95904	0·36173	1·61866	0·03455
1·69197	2·55023	0·95904	0·41497	1·65128	0·04069
1·66276	2·55023	0·97313	0·49136	1·65500	0 00776
1·57171	2·49831	0·95424	0·39794	1 61078	−0·03907
1·60206	2·53782	0·92942	0·41497	1·59192	0·01014
1 60853	2·49969	0·93450	0·41497	1·57354	0·03499
1·67578	2·56229	0·98227	0·46240	1·68687	−0·01109
1·46240	2 36361	0·9[illegible]908	0·34242	1 46996	−0·00756
1·51983	2·54788	0·92942	0·44716	1·52662	−0·00679

where $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$, and $\overline{Y}$ represent the means of X_1, X_2, X_3 and Y respectively. This line (6) has been shown in Fig. 1 and the residuals are plotted with X_1 as abscissa and corresponding Z' as ordinate with the net regression line as the initial line. The residuals are then

FIRST APPROXIMATE NET REGRESSION CURVE OF YIELD ON HEIGHT

POONA — POJ 2878

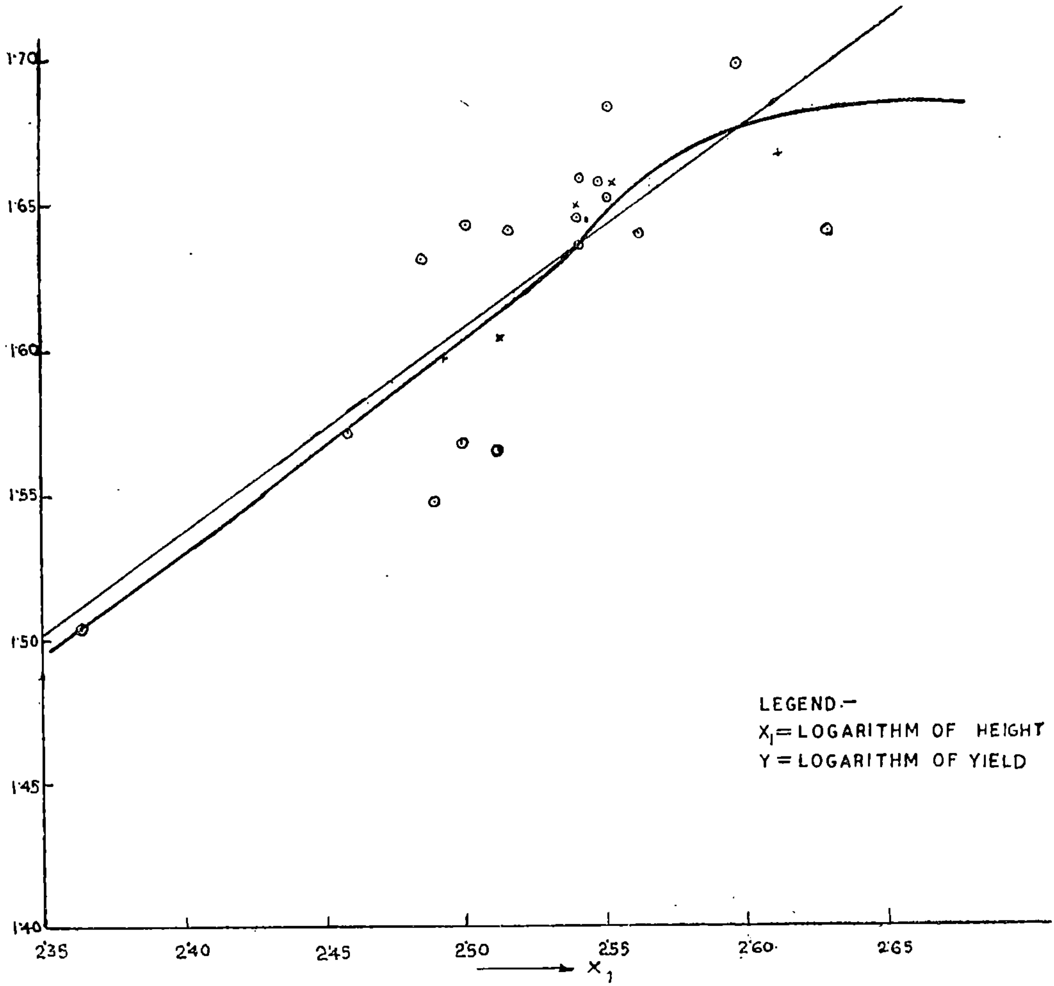

FIG-1

SECOND APPROXIMATE NET REGRESSION CURVE OF YIELD ON HEIGHT

POONA-POJ2878

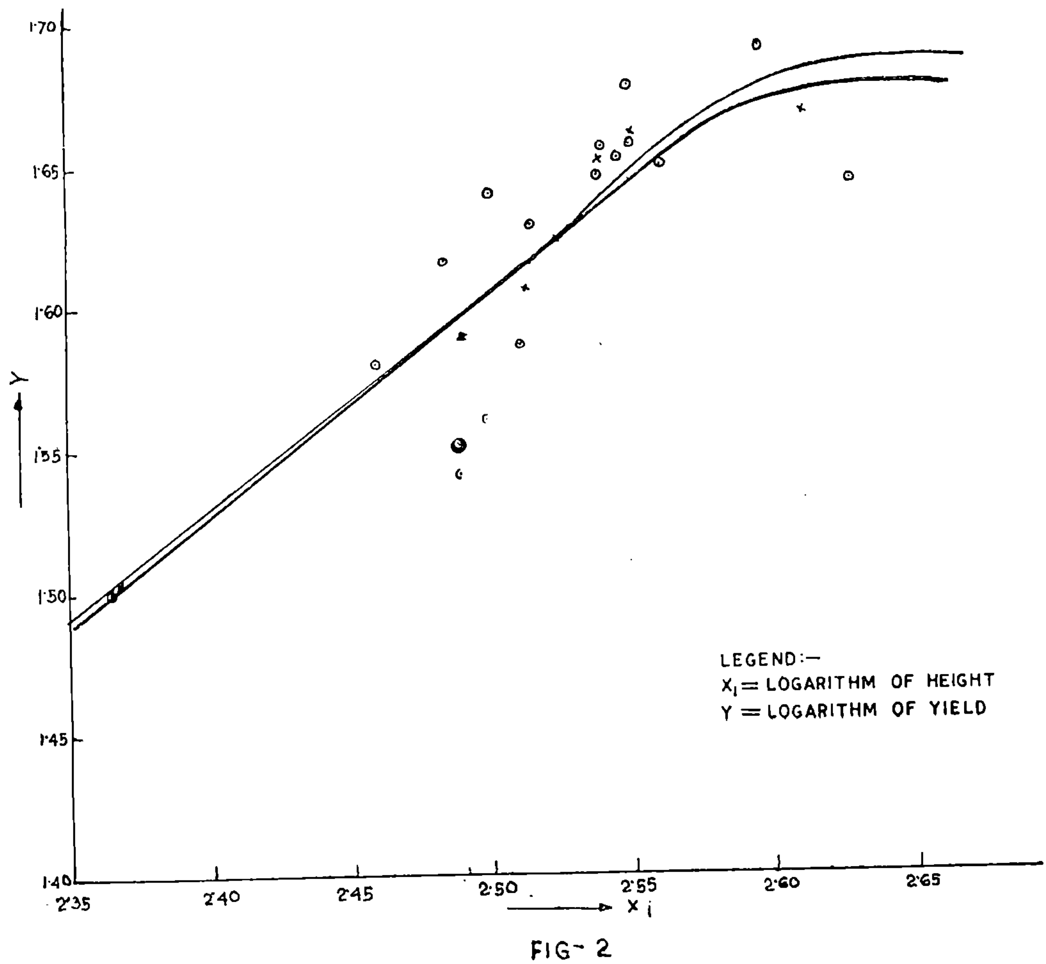

FIG- 2

averaged for selected group values of X_1. These averages are then plotted in the same manner. A free hand curve is then drawn through the several group averages, as far as possible, as is consistent with the smooth curve. This curve $Y_1=F_{11}(X_1)$ is the first approximation to the curvilinear function

$$Y=F_1(X_1)$$

Similarly the first approximation curves for the functions

$$Y=F_2(X_2) \quad \text{and} \quad Y=F_3(X_3) \text{ are}$$

$$Y=F_{21}(X_2) \quad \text{and} \quad Y=F_{31}(X_3).$$

TABLE 3

Residuals after Successive approximations

z'	Z''	Z'''
−0·058	−0·043	−0·039
0·023	0·014	0·018
0·024	0·018	0·021
0·000	0·016	0·017
0·018	0·007	0·010
−0·041	−0·026	−0·024
0·023	0·013	0·014
−0·053	−0 055	−0·054
0·035	0·024	0·019
0·041	0·029	0·023
0·008	0·008	0·013
−0·039	−0·043	−0·051
0·010	0·009	0·004
0·035	0·039	0·031
−0·011	−0·008	−0·008
−0·008	−0·004	0·001
−0·007	0·006	0·004

Estimates of Y from the first approximate regression curves :

The estimates Y'' of Y from the first approximate regression curves are found out from the equation

$$Y''=a'_{Y.123}+F_{11}(X_1)+F_{21}(X_2)+F_{31}(X_3) \qquad \ldots(7)$$

where

$$a'_{Y.123}=\overline{Y}-\frac{1}{n}\sum_{1}^{n}[F_{11}(X_1)+F_{21}(X_2)+F_{31}(X_3)] \qquad \ldots(8)$$

The residuals $Z''=Y-Y''$ are then computed. These are given in Table 3. It will be seen from Table 3 that values of Z'' are in general smaller than Z'. Out of 17 cases there are twelve cases in which Z'' are smaller than z' and four cases where Z'' are larger than Z' and in one case the two are same. The standard error of the residuals due to curvilinear regressions is 0·03408 as compared with 0·03473 due to linear regression. This shows that the curvilinear regression equation gives better representation than the linear regression equation.

Second approximate net regression curves :

The first approximate curve from Fig. 1 is first drawn. Each of the residuals Z'' is then plotted against the X values as before. The residuals are, in this case, plotted as deviation from the curves. Grouping the residuals and finding out the averages second approximation curves are drawn as before. The second approximate regression equation of Y on X_1, X_2 and X_3 now becomes

$$Y'''=a''_{Y.123}+F_{12}(X_1)+F_{22}(X_2)+F_{32}(X_3) \qquad \ldots(9)$$

where $F_{12}(X_1)$...represent the second approximation of $F_1(X_1)$...... $a''_{Y.123}$ is calculated just as $a'_{Y.123}$ in equation (8) above. The residuals $Z'''=Y-Y'''$ are calculated. Out of the 17 cases only in 7 cases Z''' is larger than Z''. The standard error of Z''' is 0·03325 as compared with 0·03408 of Z''. This shows that second approximate regression curve fits better than the first.

Further successive Approximations :

In the same manner the residuals Z'''...are plotted as deviations from the net regression curves.

$$Y=F_{12}(X_1)$$

At each stage the standard error of residuals is calculated. The above process is continued until the standard error shows a steady value or reaches a minimum. In the present case further approximations do not improve the curve any more.

4. INFERENCES FROM THE CURVES (VARIETY POJ—2878) :

The final regression curves shown in Figs. 3 to 5 represent the net relationship between yield and each crop characteristics after

FINAL NET REGRESSION CURVE OF YIELD ON HEIGHT

POONA-POJ2878

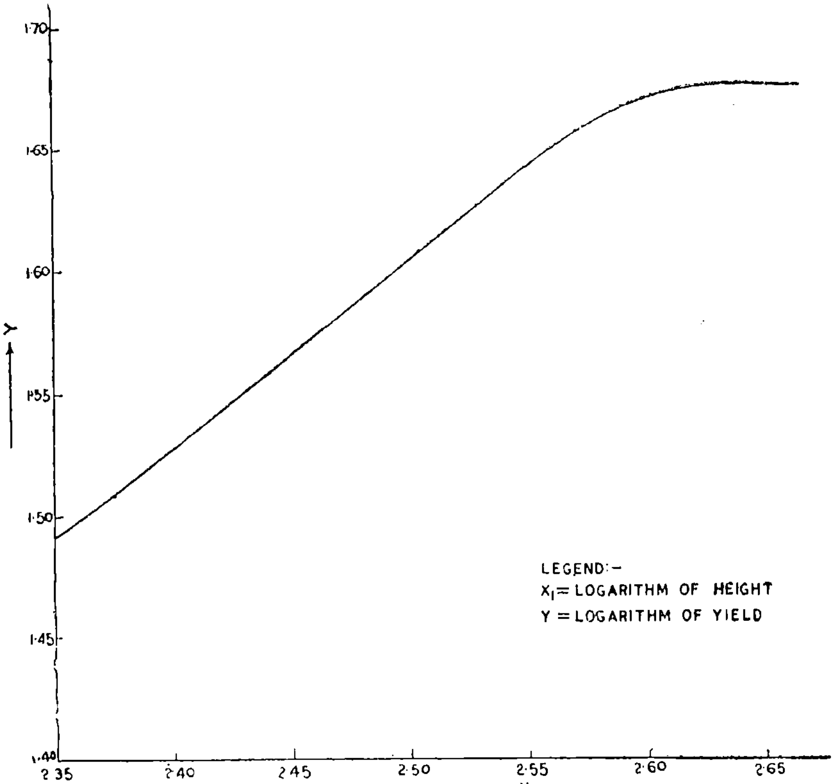

FIG-3

keeping the effect of other crop characteristics constant. Following inferences are drawn

(*i*) Yield increases with height and remains practically stationary as the height exceeds 417 cm.

(*ii*) Yield increases with height and remains practically stationary when the mid-girth exceeds 9·6 cm.

(*iii*) Yield increases very slowly as the number of canes increased and decreases as the number exceeds 2·2. The optimum number of canes per clump is 2·2.

FINAL NET REGRESSION CURVE OF YIELD ON MID-GIRTH

POONA—POJ 2878

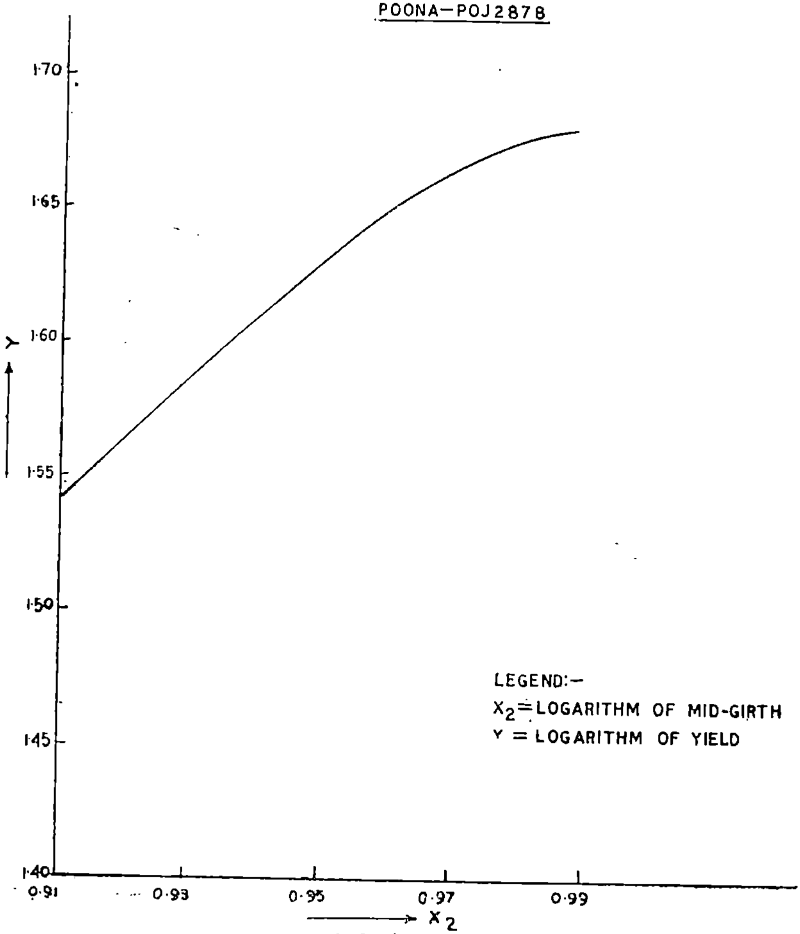

FIG - 4

5. USE OF CURVES (VARIETY POJ—2378)

The final net regression curves of Y on X_1, X_2 and X_3 after successive approximations are obtained. The regression equation is

$$Y = A_{Y.123} + F_1(X_1) + F_2(X_2) + F_3(X_3) \qquad \ldots(10)$$

where $A_{Y.123} = \overline{Y} - \frac{1}{n}\sum \left[F_1(X_1) + F_2(X_2) + F_3(X_3)\right] \qquad \ldots(11)$

FINAL NET REGRESSION CURVE OF YIELD ON No. OF CANES PER CLUMP

POONA–POJ2878

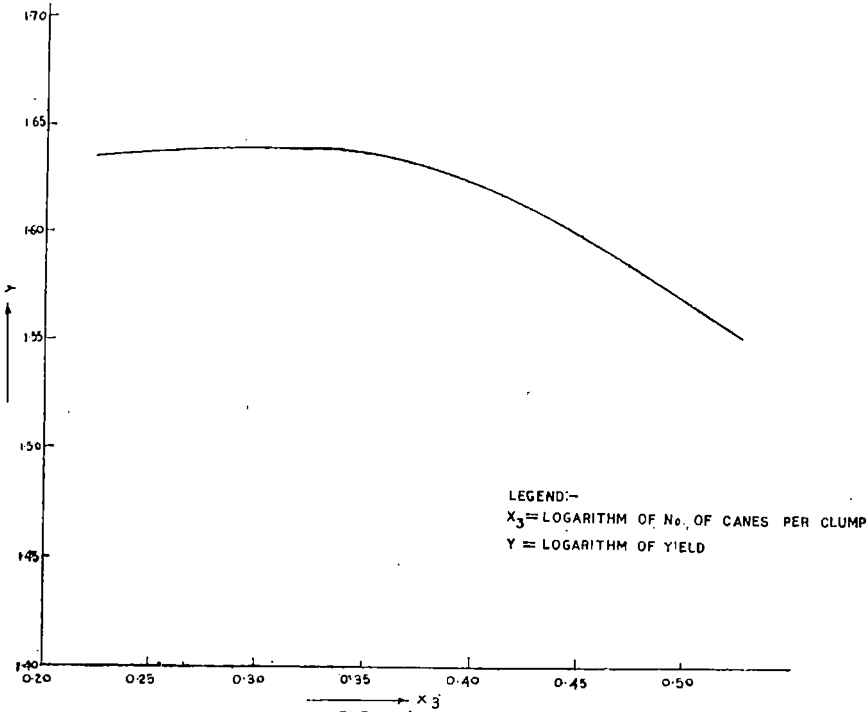

FIG-5

Summation extending over all the individuals—

From the net regression curves $Y=F_1(X_1)$, $Y=F_2(X_2)$ and $Y=F_3(X_3)$ the values of Y are read corresponding to independent variables (X_i). The values are shown in Table 4. Then the value of $A_{Y.123}$ is evaluated from the equation (11). From the equation (10), Y_e, the expected values of Y are calculated which are also shown in Table 4. From Y_e, the expected yield Y_c is found.

From this study, yield can be predicted with a considerable degree of accuracy after the crop-characteristics stabilize, *i.e.*, two to three months before the crop is ready for harvest. The expected yield with the percentage of variation from the actuals are shown in Table 4. All the estimated yields are within 7 per cent of the actuals except three in which the deviations are 9, 13 and 13 per cent. In 12 cases out of 17 cases the estimates are within 5 per cent of the actuals. The correlation coefficient between the estimates and actuals has come out as 0·9343. Thus the regression equation accounts for 87 per cent of the variation in the yield.

TABLE 4

Estimated Yield

$F(X_1)$	$F(X_2)$	$F(X_3)$	$A_{Y.123}$	Y_e	Antilog of Y_e i.e. estimated yield Y_e	Percentage of variation from the actuals
1·676	1·675	1·600	—3·240	1·711	51·4	9 4
1·672	1·678	1·638	—3·240	1·748	56·0	3·9
1·636	1·584	1·639	—3·240	1·619	41·6	5·2
1·636	1·617	1 636	—3·240	1·649	44·6	4·1
1·643	1·597	1·639	—3·240	1·639	43·6	2·2
1·613	1·597	1·637	—3·240	1·607	40·5	5·7
1·616	1 628	1·639	—3·240	1·643	44·0	3·1
1·596	1·584	1·638	—3·240	1·578	37 8	13·2
1·592	1·647	1·636	—3·240	1·635	43·2	4·0
1·616	1·647	1·617	—3·240	1·670	46·8	4·9
1·646	1·669	1·575	—3·240	1·650	44·7	2·8
1·603	1·637	1·623	—3·240	1·623	42·0	12·6
1·634	1·584	1·617	—3·240	1·595	39·4	1·5
1·604	1·597	1·617	—3·240	1·578	37·8	6·9
1 654	1·678	1·592	—3·240	1·684	48·3	1·9
1·501	1·562	1·638	—3·240	1·461	28·9	3·4
1·572	1·584	1·600	—3·240	1·561	32·8	0·9

6. Results using data of variety CO-419 grown at Poona :

The same method has been applied here with 14 years' data. The net regression curves are shown in Figs. 6-8. And following inferences are drawn from the curves :—

(*i*) Yield increases with height and becomes stationary as the height exceeds 372 cm.

(*ii*) Yield increases with mid-girth and remains stationary when mid-girth exceeds 9·5 cm.

(*iii*) Yield remains stationary with the number of canes and decreases as the number of canes exceeds 2·3. The optimum number of canes per clump is 2·3.

The estimated yield calculated from the regression equations are all within 7 per cent of the actuals except two in which the deviations are 8·6 and 8·3 per cent of the actuals. The correlation coefficient between the estimated and actual yield is found to be 0·9402. Thus the regression equation accounts for 88 per cent of the variation in yield.

FINAL NET REGRESSION CURVE OF YIELD ON HEIGHT

POONA—CO 419

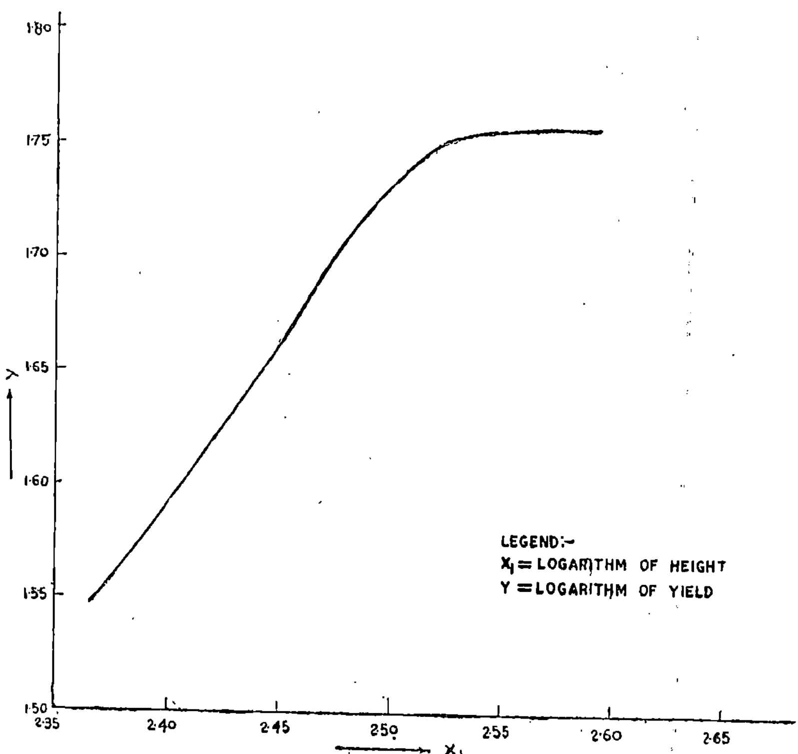

FIG. 6

FINAL NET REGRESSION CURVE OF YIELD ON MID-GIRTH

POONA-CO419

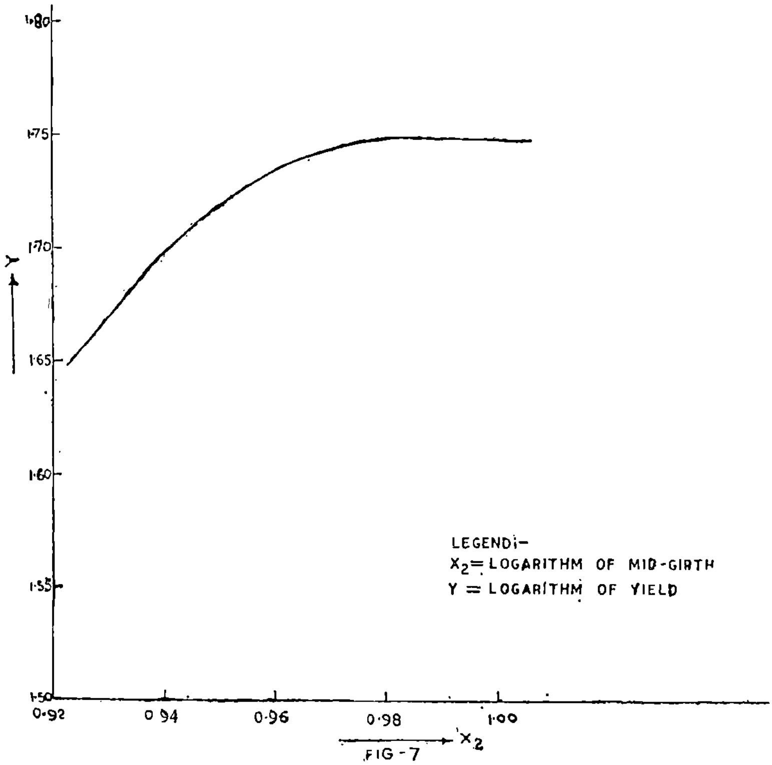

FIG-7

FINAL NET REGRESSION CURVE OF YIELD ON No. OF CANES PER CLUMP

POONA-CO419

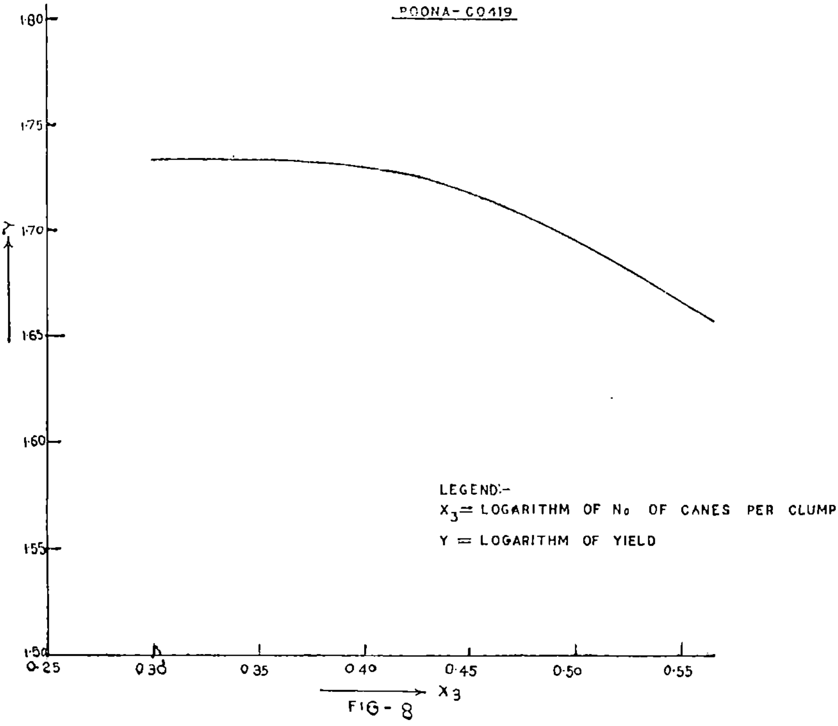

FIG-8

7. COMPARISION OF THE TWO METHODS :

It is desirable to verify how the results of the present study using product functions connecting yield and crop-characteristics compare with those made earlier by Gangopadhyaya and Sarker (1964) using additive functions connecting yield and crop-characteristics. The net regression curves assuming these two different kinds crop-characteristics-yield relations and using the same set of data as used by Gangopadhyaya and Sarker have been drawn. The results obtained from the different studies are shown in Table 5. The correlation

TABLE 5

Comparison of estimates of yield.

Observed Yield	*Estimated Yield*		*Deviations*	
	Log. scale	*Org. scale*	*Log. scale*	*Org. scale*
47·0	50·2	48·7	—3·2	—1·7
58·3	57·9	56·2	0·4	2·1
43·9	43·4	43·7	0·5	0·2
46·6	45·3	46·8	1·3	—0·2
44·6	45·7	46·2	—1·1	—1·6
38·3	39·0	40·7	—0·7	—2·4
45·4	45·4	47·3	0·0	—1·9
33·4	35·7	34·3	—2·3	—1·1
45·0	41·8	41·3	3 2	3·7
49·2	48·9	49·4	C·3	—0·2
46·0	45·3	43 8	0·7	2·2
37·3	41·1	42·4	—3·8	—5·1
40·0	39·6	39·4	0·4	0·6
40 6	36·5	35·7	4·1	4·9

coefficient between the estimated and actual yield is found to be 0·934 by assuming product functions in comparison with 0·903 found by previous method. The present regression equation seems to be a little better than the earlier one as it accounts for 87 per cent of the variation of yield while the earlier one accounts for about 82 per cent. Moreover, the assumption made in the present study is more generalised as this takes into account the effect of product and higher powers of the three crop-characteristics concerned. The computational labour involved in the present study does not compare unfavourably with that of the earlier study. Hence the present line of attack seems to be slightly superior to the earlier one.

ACKNOWLEDGEMENTS :

The author has great pleasure to thank Shri M. Gangopadhyaya, Director of Agricultural Meteorology, for suggesting the problem and taking keen interest for progress of the work. The author is also thankful to Smt. N.N. Khambete for her help in the computation.

REFERENCEES

Ezekiel, Mardecai and Karl A. Fox. (1959), "Methods of Correlation and Regression Analysis", pp. 204-275.

Fisher, R.A. and Yates, F. (1943), "Statistical Tables for Biological, Agricultural and Medical research".

Gangopadhyaya, M. and Sarker, R.P. (1964), "Ind. Jour. of Met. & Geophys." 15, 2, 201-214.