METHOD OF FITTING GROWTH CURVES

By V. K. Ramabhadran

Agricultural Meteorology Section, Meteorological Office, Poona.

Introduction

Any study of the life-history of a crop and the bearing which it has on the final yield should be based on quantitative measurements, made regularly on the growing crop. The best method of representing these growth observations, is to fit an appropriate curve to them. Pearl and Reed[1,2] have fitted 'growth curves' to many biological phenomena. They have used[3 4] the 'Logistic' curve defined as

$$dy/dx = ay\,(K - y) \tag{1}$$

(where K is the maximum value of y and 'a' a constant) in representing the growth in population of many countries. A curve of this type describes any phenomena that increases slowly in the beginning, rapidly in the middle and again slowly at the end. The 'curve' in fact is a combination of the two curves $dy/dx = y$ and $dy/dx = K - y$ which correspond respectively to the accelerating and inhibiting factors that influence any biological growth.

Integrating (1) we have

$$\log_e \frac{K - y}{y} = -\,aK\,(x - x_0), \tag{2}$$

where x_0 is the constant of integration. Equating the derivative of (1) to zero we get $y = K/2$ as the ordinate of the point of inflection and from (2) we see that 'x_0' is the abscissa.

Writing (2) in the form

$$y = \frac{K}{1 + e^{a_0 + a_1 x}}. \tag{3}$$

Pearl[1] uses the 'method of three equidistant ordinates' to get approximate values of the constants and applies the method of least squares in the final adjustment of the curve to the data. Henry

Schultz[5] attempts a similar adjustment by expanding the function

$$y = \frac{K_0 + \Delta K}{1 + e^{(a_0 + \Delta a_0) + (a_1 + \Delta a_1) x}}$$

(where K_0, a_0, a_1 are the first estimates) as a Taylor's series to a first approximation. Sichel[6] has attempted the fitting of growth curves by the method of "frequency moments", the n-th frequency moment being defined by $J_n = \sum_{x_1}^{x_k} (u_x)^n$, where u_x is the observed value of the dependent variable at time x. Yule[7] taking (2) in the form

$$y_x = \frac{K}{1 + e^{-aK(x - x_0)}} \tag{4}$$

applies the 'method of sum of reciprocals' to estimate the unknowns. He has further shown that since

$$y_{x_0+\delta} = \frac{K}{1 + e^{-aK\delta}} = K - \frac{K}{1 + e^{aK\delta}} = K - y_{x_0-\delta}$$

the curve (4) is symmetrical about the point of inflection (x_0, $K/2$). This means that the decline in growth starts after half the growth is completed and that the growth is retarded as gradually as it is promoted. Consequently if the growth (dy) in each interval be plotted against x, we should obtain a symmetrical curve. Lack of symmetry of such a curve suggests that the fundamental equation (1) should be modified for a skew material.

The Height Curve and the Curve of Ripening

Starting from the fundamental equation (1) we can get two curves, one of which will represent the 'elongation' of a crop and the other may be used to describe the ripening of sugarcane.

Pearl[9,8] has derived the asymmetrical form of the logistic by considering the equation

$$\frac{dy}{dx} = f(x) \, . \, y (K - y),$$

instead of equation (1). Integrating the above equation he has put the skew-logistic curve in the final form

$$y = d + \frac{K}{1 + e^{a_0 + a_1 x + a_2 x^2 + a_3 x^3}} \tag{5}$$

where $y = d$ and $y = K + d$ are the lower and upper asymptotes. We shall use this curve to represent the time-variation in the height (y) of a crop.

If instead of equation (1) we consider the equation

$$\frac{dy}{dx} = a (K - y) \tag{6}$$

we have on integration

$$\log_e \frac{1}{K - y} = ax + b, \tag{7}$$

where b is the constant of integration. This curve may also be taken as

$$y = K - e^{a_0 + a_1 x}$$

where we set

$$a_0 = -b$$

$$a_1 = -a.$$

Since $\frac{dy}{dx}$ vanishes when y equals K we see that $y = K$ is the asymptote of (7) and consequently a_1 should be negative. The equation (7) therefore represents a curve which is steep initially and flat at the end. The curve is used to represent the ripening of sugarcane.

Estimation of the Constants of (5) and (7) and Their Sampling Errors

In order to fit the curve (5) we first plot the observed 'y's (heights) and fix the lower and upper asymptotes approximately.

Let $y = d_0$ and $y = K_0 + d_0$ represent the lower and upper asymptotes respectively. Assuming $d = d_0 + \Delta$ and $K = K_0 + k$ (where Δ and k are corrections to d and K respectively) equation (5) when these values are substituted for d and K gives

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 = \log_e \frac{K_0 + d_0 - y + k + \Delta}{y - (d_0 + \Delta)}$$

$$= \log_e \left(\frac{K_0 + d_0 - y}{y - d_0}\right)\left(1 + \frac{k + \Delta}{K_0 + d_0 - y}\right)\left(1 - \frac{\Delta}{y - d_0}\right)^{-1}$$

$$= \log_e \left(\frac{K_0 + d_0 - y}{y - d_0}\right) + \left(\frac{k + \Delta}{K_0 + d_0 - y}\right) + \left(\frac{\Delta}{y - d_0}\right)$$

neglecting higher powers of k and Δ after the first. Forming the normal equations that minimise

$$\Sigma_x \left\{a_0 + a_1x + a_2x^2 + a_3x^3 - \log_e \frac{K_0 + d_0 - y}{y - d_0} - \frac{k + \Delta}{K_0 + d_0 - y} - \frac{\Delta}{y - d_0}\right\}^2$$

with respect to a_0, a_1, a_2, a_3, Δ and k, the estimates of these constants are obtained.

After obtaining the initial estimate of the constants we can improve them by correction factors derived from normal equations containing only the first differentials of the constants. Such a process incidentally makes also the error of estimate a minimum. Differentiating equation (5) we have

$$\Delta y = \Delta d + \frac{\partial y}{\partial K}\Delta K + \frac{\partial y}{\partial a_0}\Delta a_0 + \frac{\partial y}{\partial a_1}\Delta a_1 + \frac{\partial y}{\partial a_2}\Delta a_2 + \frac{\partial y}{\partial a_3}\Delta a_3.$$

Substituting for the various partial derivatives we get

$$\Delta y = \Delta d + \frac{y}{K}\Delta K - \frac{y^2 e^{\phi(x)}}{K}(\Delta a_0 + x\Delta a_1 + x^2\Delta a_2 + x^3\Delta a_3) \qquad (8)$$

where

$$\phi(x) = a_0 + a_1x + a_2x^2 + a_3x^3.$$

Forming the normal equations that minimise

$$\Sigma_x \left\{\Delta y - \Delta d - \frac{y}{K}\Delta K + \frac{y^2 e^{\phi(x)}}{K}(\Delta a_0 + x\Delta a_1 + x^2\Delta a_2 + x^3\Delta a_3)\right\}^2$$

with respect to the various corrections, we get the estimates. The corrected value of the constants will then be $a_3 + \Delta a_3$, $a_2 + \Delta a_2$, $a_1 + \Delta a_1$, etc. If the corrected values are used in (5) the fit is improved.

The sampling errors of the constants are obtained as follows:

Assuming that the graduation of a series of observed points by (5) may be regarded as a ' false position ' we may employ the equation of residuals, *viz.*, (8) to obtain the sampling errors of K, a_0, a_1, etc. We obtain the value of ΔK, Δa_0, Δa_1, etc., from the normal equations arising out (8) and thence determine the sampling error of the observational equation

$$\Delta y' = \frac{y'}{K}\Delta K - \frac{y'^2 e^{\phi(x)}}{K}(\Delta a_0 + x\Delta a_1 + x^2\Delta a_2 + x^3\Delta a_3)$$

(where $y' = y - d$) as $\sqrt{\frac{[vv]}{n-m}}$ in the standard notation.[10] The sampling errors of the unknowns are calculated from the sampling error of the observational equation by applying the formula

$$p_x = p\,[\alpha\alpha],$$

where

p_x = sampling error of the unknown x,

p = sampling error of the observational equation

and $[\alpha\alpha] = A/D$, D standing for the determinant of coefficients of the normal equations and A for the co-factor of the determinant corresponding to the unknown.

In order to fit the curve (7), we can fix the asymptote at $y = K_0$ from the graph of observed 'y's and set $K = K_0 + \Delta K$, where ΔK is small and higher powers may therefore be neglected. Taking the equation (7) in the form.

$$\log_e (K - y) = a_0 + a_1 x$$

we have by putting $K = K_0 + \Delta K$,

$$\log_e (K_0 + \Delta K - y) = a_0 + a_1 x$$

i.e.,

$$\log_e \left[(K_0 - y)\left(1 + \frac{\Delta K}{K_0 - y}\right)\right] = a_0 + a_1 x$$

i.e.,

$$\log_e (K_0 - y) + \frac{\Delta K}{K_0 - y} = a_0 + a_1 x$$

or

$$a_0 + a_1 x - \frac{\Delta K}{K_0 - y} = \log_e (K_0 - y) \qquad (9)$$

The normal equations corresponding to (9) give the value of a_0, a_1, and ΔK and their sampling errors also as shown above.

Application of the Curves (5) and (7) to Data

We may now proceed to the application of the curves to actual data. The data[11] considered in this paper are a part of the precision observations recorded on the 1938 Rice Crop at Karjat, one of the earliest stations in the Bombay Province. The values given in column (2) of Table I are the means of the height measurements made on the plants nearest to the observer in each bunch, the total number of bunches selected at random each week being 384 (32 in each of the 12 plots). The initial estimates of the constants are:

$$K = 85\cdot3 \;;\quad a_0 = 4\cdot36561 \;;\quad a_1 = -0\cdot64905$$
$$d = 11\cdot0 \;;\quad a_2 = 0\cdot07049 \;;\quad a_3 = -0\cdot00521$$

The values of y obtained from the corresponding equation

$$y = 11 + \frac{85.3}{1 + e^{4\cdot36561 - 0\cdot64905x + 0\cdot07049x^2 - 0\cdot00521x^3}}$$

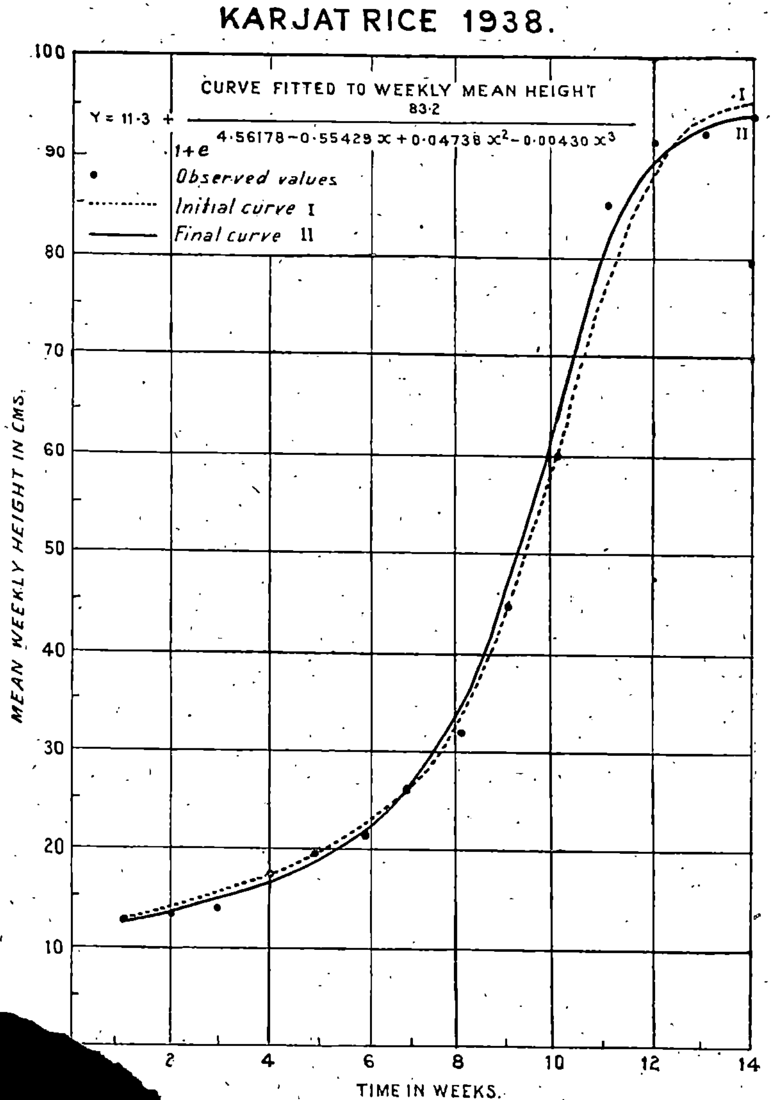
are given in column 2 of Table I and the graph is shown by the

TABLE I

Week No. beginning with the date of first observation	Weekly mean height in cms.	Calculated values from initial curve $K=85\cdot3$, $d=11\cdot0$, $a_0=4\cdot36561$, $a_1=-0\cdot64905$, $a_2=0\cdot07049$, $a_3=-0\cdot00521$	Calculated values from final curve $K=83\cdot2$, $d=11\cdot3$, $a_0=4\cdot56178$, $a_1=-0\cdot55429$, $a_2=0\cdot04738$, $a_3=-0\cdot00430$
1	12·9	12·9	12·7
2	13·2	14·0	13·5
3	13·8	15·4	14·5
4	17·1	17·1	15·9
5	19·6	19·3	18·0
6	21·4	22·2	21·1
7	26·7	26·7	26·1
8	31·5	33·7	34·3
9	44·3	44·6	47·1
10	59·7	59·6	63·6
11	85·4	76·3	79·7
12	91·8	88·3	89·3
13	92·2	94·0	93·1
14	93·8	95·8	94·2

dotted curve in Fig. 1. The least squares correction to these constants are:

$$\Delta K = -2\cdot08 \;;\quad \Delta a_0 = +0\cdot19617 \;;\quad \Delta a_1 = +0\cdot09476$$
$$\Delta d = +0\cdot31 \;;\quad \Delta a_2 = -0\cdot02311 \;;\quad \Delta a_3 = +0\cdot00091$$



…ing with the date of first observation)

FIG. 1

Constant	Value	S.E.
K ..	83·2	±·1272
a_0 ..	+ 4·56178	±·1072
a_1 ..	− 0·55429	±·0973
a_2 ..	+ 0·04738	±·0167
a_3 ..	− 0·00430	±·0009

Judging from the sampling errors the constants are significant.

The values calculated from the final equation

$$y = 11 \cdot 3 + \frac{83 \cdot 2}{1 + e^{4 \cdot 56178 - 0 \cdot 55429x + 0 \cdot 04738x^2 - 0 \cdot 00430x^3}}$$

are given in column 3 of Table I and the graph is shown by the continuous curve in Fig. 1. The root-mean-square deviation has decreased from 2·82 for the initial curve to 2·32 for the final curve, indicating the improvement. The point of inflection obtained as the intersection of the above curve with the curve

$$y = 52 \cdot 9 - 41 \cdot 6 \frac{\phi''(x)}{[\phi'(x)]^2},$$

where

$$\phi(x) = 4 \cdot 56178 - 0 \cdot 55429x + 0 \cdot 04738x^2 - 0 \cdot 00430x^3$$

is at $x = 9 \cdot 9$. That is, the growth is maximum at the 9·9th week. This value of x agrees with the observed data in which the maximum growth (from 59·7 cms. to 85·4 cms.) occurs in the 10th
After the point of inflection the weekly growth decreases
against the gradual increase before the point of infle
the acceleration is gradual while the inhibition is

The final choice of the curve (5) was m
curves were fitted. The appendix gives the d
The values calculated from the various equ

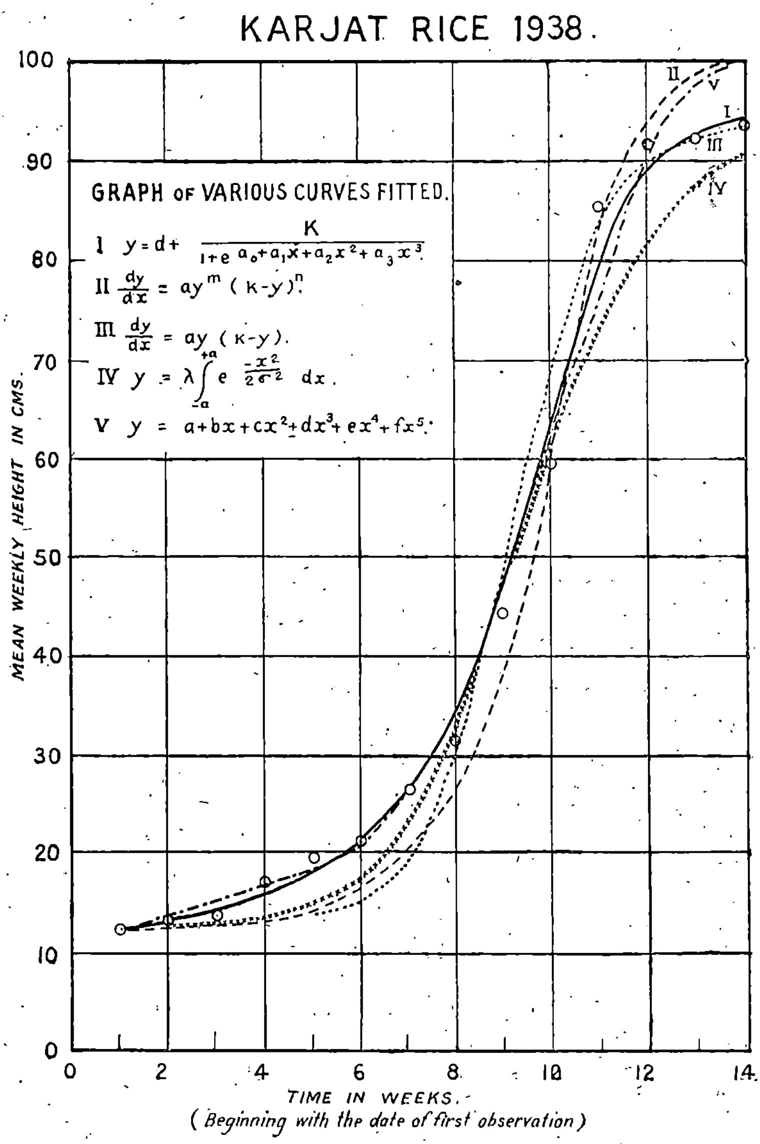


FIG. 2

logistic' are shown in Fig. 2. The root-mean-square deviations corresponding to each curve are given below;

No.	Curve fitted	Root-mean-square deviation
1	Skew-logistic	2·32
2	$\log_e \frac{K+d-y}{y-d} = a_0 + a_1 x + a_2 x^2 + a_3 x^3$	3·56
3	5th degree polynomial	4·05
4	Normal Ogive	4·42
5	$\frac{dy}{dx} = a y^m (K-y)^n$	4·45
6	$\frac{dy}{dx} = a y (K-y)$	5·12

Assuming that the root-mean-square deviation is an index of dispersion[14] of the observed from the calculated, it appears from the table that the skew-logistic curve is superior to the other curves. In order to demonstrate the applicability of the curve, the constants of the curves fitted together with the observed[12] and calculated heights for some more crops are given in Table II.

The data of ripening relate to the sugarcane crop at Poona.[12] Fortnightly measurements of the ripening of sugarcane are made with the help of the hand refractometer. The index thus obtained is called 'Brix Reading'. The value of the constants of the curve (7) for the 1946–47 data (variety P.O.J.) are $K = 21 \cdot 71$, $a_0 = 2 \cdot 4333$, $a_1 = -0 \cdot 4350$. The observed Brix (mean of readings taken on 72 canes selected at random) and the values calculated from the curve

$$y = 21 \cdot 71 - e^{2 \cdot 4333 - 0 \cdot 4350 x}$$

are given in Table III and in Fig. 3. The graph shows that the curve adequately represents the data. Though the set of observed values can be represented by some other curves, *e.g.*, a parabola of 2nd degree, it is desirable to consider in the first instance an equation with a biological background. The equation (6) in fact means that the ripening goes on at a definite rate which is at any moment proportional to the amount of ripening yet to be completed.

TABLE II

KARJAT RICE, 1935			PARBHANI COTTON, 1947			POONA SUGARCANE, 1947		
Week No.	Weekly mean height in cms.	Calculated value from skew-logistic curve with $K=85\cdot4$; $d=9\cdot2$ $a_0=4\cdot8511$; $a_1=-1\cdot1668$ $a_2=0\cdot1740$; $a_3=-0\cdot0104$	Week No.	Weekly mean height in cms.	Calculated value from skew-logistic curve with $K=55\cdot2$; $d=3\cdot0$ $a_0=4\cdot1204$; $a_1=-1\cdot0995$ $a_2=0\cdot0729$; $a_3=-0\cdot0028$	Week No.	Weekly mean height in cms.	Calculated value from skew-logistic curve with $K=438\cdot2$; $d=14\cdot7$ $a_0=5\cdot1286$; $a_1=-0\cdot5306$ $a_2=0\cdot0212$; $a_3=-0\cdot0004$
1	10·1	11·0	1	5·4	5·4	1	18·4	19·0
2	12·0	12·8	2	7·8	8·6	3	27·4	25·1
3	15·5	14·9	3	14·3	13·9	4	31·9	30·0
4	16·5	17·0	4	21·2	21·2	5	35·2	36·3
5	18·4	18·8	5	31·0	32·0	6	39·4	44·2
6	20·7	20·7	6	38·4	36·8	7	48·8	53·9
7	21·0	23·1	8	45·2	47·6	9	76·0	76·9
8	29·2	27·2	9	51·9	51·1	10	86·8	90·8
9	35·7	34·8	11	55·5	55·3	11	108·3	104·9
10	45·8	48·5	13	56·7	57·2	12	116·0	119·3
11	65·5	68·1				13	138·7	134·3
12	88·6	85·4				14	153·5	148·7
13	95·2	92·8				16	181·8	179·1
14	94·7	94·4				17	196·5	193·7
						19	226·7	222·9
						20	238·8	238·2
						22	261·7	269·6
						23	285·1	285·4
						24	321·7	301·6
						25	336·7	318·9
						27	344·1	353·0
						28	354·5	369·6
						29	375·0	384·7
						32	416·5	422·3
						34	439·4	437·9
						36	450·7	446·6
						38	449·5	450·7
						40	452·2	452·2

TABLE III

Time (x) in fortnights	Brix Reading y (observed)	Calculated value with $K = 21 \cdot 71$; $a_0 = 2 \cdot 4333$ $a_1 = -0 \cdot 4350$
1	14·61	14·32
2	16·60	16·95
3	18·74	18·62
4	19·22	19·72
5	20·65	20·42
6	21·35	20·88
7	21·38	21·24

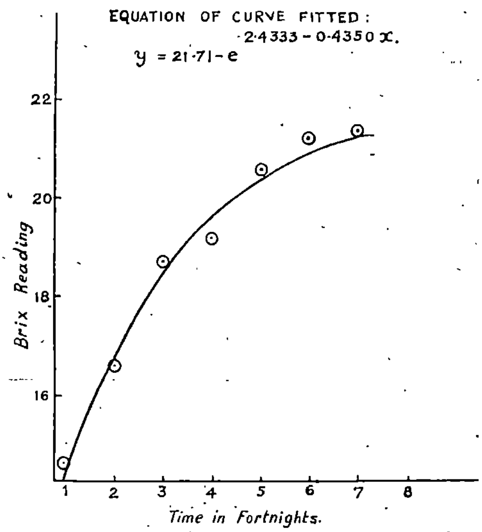


FIG. 3

The sampling errors of the constants together with the constants of the curve fitted to the ripening data of another year at Poona are given below:

Year	Variety	K	a_0	a_1
1946–47	P.O.J.	21·7 ±·06	2·4333 ±·2640	−0·4350 ±·0878
1947–48	P.O.J.	22·6 ±·06	2·4696 ±·1426	−0·3621 ±·0603
	C.O. 419	19·2 ±·24	2·6872 ±·1625	−0·4061 ±·0814

The sampling errors bring out the significance of the constants.

Conclusions and Summary

It is seen from Table I that the method of 'least squares approximation' improves the initial fit of the skew-logistic curve to the height data. It is found that this curve gives the 'best fit'. Hence a set of six constants K, d, a_0, a_1, a_2, a_3 represents the entire growth in height. The representation of the Ripening or the Brix Reading of sugarcane by (7) is appropriate and hence the three constants K, a_0, a_1 summarise the entire process of ripening of the cane. Such constants can therefore be termed 'Growth Constants'.

In order to study the influence of weather (which the crop experiences during its life-span) on growth phenomena like elongation, ripening, etc., the variation in these constants can be compared and studied in relation to the variation in the constants representing the distribution of weather factors during the growing period. Again since a growing crop integrates weather effects and also indicates its own future yield, we can use these 'growth constants' in order to work out a direct regression between the growth of a crop and its final yield. For such studies however records extending over a series of years will be essential.

The author is very glad to thank Dr. L. A. Ramdas, Director of Agricultural Meteorology, for suggesting the problem and for keen interest in its progress and Dr. U. S. Nair, Professor of Statistics, Travancore University, for elucidating some of the implications of the theoretical aspects. Thanks are also due to Messrs. P. S. Nair and N. S. Sastry for their help in computing the data.

References

1. Pearl, R. .. *Studies in Human Biology*, Chap. XXIV (Williams and Wilkins Co., Baltimore), 1924.
2. Reed, H. S. .. "Growth and variability in Helianthus," *Am. Jour. Bot.*, 1919, 6.

3. Pearl, R. .. "On the rate of growth of the population of the United States since 1790 and its mathematical representation," *Proc. Nat. Acad. Sci.*, 1920, **6**.

4. Wilson, E. B. .. "The Logistic or Autocatalytic grid," *ibid.*, 1925, **11**.

5. Schultz, H. .. "The standard error of forecast from a curve," *J. Am. Stat. Assoc.*, 1930, **25**, 139–85.

6. Sichel, H. .. "Fitting growth and frequency curves by the method of frequency moments," *J. R. S. S.*, 1947, **110**, Part IV.

7. Yule, G. U. ... "The growth of population and the factors which control it," *ibid.*, 1925, **88**.

8. Pearl, R. .. "On the mathematical theory of the population growth," *Metron*, 1923, **3**.

9. ——— and Reed, L. J. "Skew-growth curves," *Proc. Nat. Acod. Sci.*, 1925, 11.

10. Brunt, D. .. *The combination of Observations*, 1931, Chap. VI, 100–05.

11. Kalamkar, R. J., Kadam, B. S. Satakopan, V., and Gopal Rao, S. .. "Studies on precision observations on rise at Karjat," *Ind. J. Agri. Sci.*, 1943, **13**.

12. .. *Report on the Co-ordinated Crop-Weather Scheme*, 1947–48.

13. Krichewsky .. Physical Department, Paper No. 8, Govt. Press, Cairo, 1922.

14. Pearl, R. .. *Medical Biometry and Statistics*, 1940.

APPENDIX

The following curves were fitted before selecting the skew-logistic to represent the height:

(i) Since it was felt whether the skewness could have arisen by the accelerating factor dominating over the 'inhibition' factor, the following differential equation was formulated:

$$\frac{dy}{dx} = \text{constant (accelerating factor)}^m \times \text{(inhibiting factor)}^n$$

$$= ay^m (K - y)^n \qquad\qquad m \neq n.$$

The constants a, m, n are obtained as follows[13]:

$$\frac{dy}{dx} = z \text{ (say)} = ay^m (K - y)^n = \phi (y) \text{ (say)}.$$

If M_0, M_1, M_2 are the first three moments of $z = \phi (y)$, we have setting $y = Ku$,

$$a \int_0^K y^m (K - y)^n \, dy = aK^{m+n+1} \int_0^1 u^m \, (1 - u)^n \, du$$

$$= aK^{m+n+1} \, B (m + 1, n + 1),$$

(where B is the Beta function) and consequently

$$M_0 = aK^{m+n+1} \, B (m + 1, n + 1)$$

$$M_1 = aK^{m+n+2} \, B (m + 2, n + 1)$$

$$M_2 = aK^{m+n+3} \, B (m + 3, n + 1)$$

Let $\lambda = \frac{M_1}{M_0}$ and $\mu = \frac{M_2}{M_1}$ then m and n are given by the equations

$$m + 1 = \frac{\lambda (K - \mu)}{K (\mu - \lambda)}.$$

$$m + n + 2 = \frac{K - \mu}{\mu - \lambda}$$

Having obtained m and n from the above equations, we calculate 'a' from

$$a = \frac{M_0 \, \Gamma (m + n + 2)}{K^{m+n+1} \, \Gamma (m + 1) \, \Gamma (n + 1)}.$$

For the 1938 height data the curve is

$$\frac{dy}{dx} = 1.5934\, y^{1.7494}\, (K - y)^{0.8570}$$

showing clearly how the skewness in growth is attributable to the accelerating factor having a larger index—nearly double that of the inhibiting factor.

(ii) The curve $\frac{dy}{dx} = ay\,(K - y)$ was also fitted as a special case of the above when $m = n = 1$.

(iii) The Normal Ogive curve $y = \lambda \int_{-a}^{+a} e^{\frac{-x^2}{2\sigma^2}}\, dx$.

(iv) The 5th degree parabola $y = a + bx + cx^2 + dx^3 + ex^4 + fx^5$.

(v) The regression polynomial $\log_e \frac{K + d - y}{y - d} = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ fitted by the method of least squares.