MEDIUM-TERM AND LONG-TERM PREDICTION MODEL OF ANNUAL TEA CROP PRODUCTION USING HEURISTIC LEARNING ALGORITHM

By

A. S. Chaudhuri

Jalpaiguri Govt. Engg., College, Jalpaiguri, West Bengal

(Received August, 1980)

Summary

This work relating to the medium-term and long-term prediction model of annual production of Indian tea has been divided into two parts. In the first part an annual model of tea crop production for medium-term (5 to 7 years) prediction has been obtained. Different types of models of polynomials of increasing complexity have been tested. The polynomial which gives minimum of a selection criterion has been found. It is found that the annual tea crop production is a non-stationary process. It is observed that the law of annual tea crop production varies with time.

In the second part a long-term equilibrium model for annual tea crop prediction is suggested. Indian tea production has a growth rate of 3.17 per cent. This gives rise to a doubling time around 2000 AD. The growth rate has its inherent positive and negative feedbacks, In the finite system in the long run negative feedback is bound to dominate. To avoid this type of disaster a limit on growth after 2000 AD and eventually leading to an equilibrium condition of production is suggested. Accordingly a model for annual tea crop production for long-term prediction is obtained.

1. Medium-term prediction model of annual production of Indian Tea

With the theory of self-organisation[2-6] commonly known as group method of data handling it has been possible to formulate mathematical models for complex processes with prediction optimization.

The concept of self-organisation can be illustrated as follows : when the model complexity gradually increases the computer finds by shifting the different models, the minimum of a selection eriterion which the computer has been ordered to look for. Thus the computer indicates to the operator the model of optimum complexity.

The purpose of this part of the work is to obtain an easily usable model of optimum complexity for medium-term (5 to 7 years) prediction of annual production of Indian Tea.

2. A PREDICTION MODEL FOR ANNUAL TEA CROP PRODUCTION FOR MEDIUM-TERM PRODUCTION

We have yearly production data for Indian Tea[10] from 1900 to 1976. For the purpose of obtaining the model data from 1945 to 1976 were used. Data are shown in table 1.

Power density spectra versus cycle per annum characteristic[9] of the data does not show any harmonicity in the process. So it is obvious that the process does not contain any sinusoidal deterministic part.

It is observed from the physical nature of the agricultural system that if the land is constrained and if the variables of crop production are optimized the production of crop ultimately comes to a steady state value. The process can be very approximately identified to follow the pattern of a second order finite difference equation of the following nature :

$$Y_{k+1}=f\,(Y_k,\ Y_{k-1},\ t_k) \qquad ...(1)$$

We write,

$Y_{k+1}=y$, tea crop production for the $k+1$-th year

$Y_k=x_1$, tea crop production for the k-th year

$Y_{k-1}=x_2$, tea crop production for the k-1-th year

$t_k=x_3$, time instance for the k-th year.

So, $y=f\,(x_1,\ x_2,\ x_3)$ ...(2)

The function $f(\ .\ )$ is sought in the class of quadratic polynomials on the basis of a table of polynomials of gradually increasing complexity of three variables as shown in table 2 with the help of the theory of self-organisation of different mathematical models.

The model of optimum complexity is selected on the basis of minimum of integral square error criterion. Integral square error is defined as [2]

$$i^2=\frac{\sum_{i=1}^{N}(Y_{tab}(i)-Y_{d.m}(i))^2}{\sum_{i=1}^{N}(Y_{tab}(i))^2}$$

TABLE 1

Annual production of Indian Tea

Year	*Production in million Kgs.*
1945	229.038
1946	246.062
1947	254.801
1948	262.092
1949	265.365
1950	275.475
1951	290.789
1952	283.618
1953	278.777
1954	295.519
1955	307.704
1956	308.719
1957	310.802
1958	325.225
1959	325.955
1960	321.077
1961	354.397
1962	346.735
1963	346.413
1964	372.485
1965	366.374
1966	375.983
1967	384.759
1968	402.489
1969	393.588
1970	418.517
1971	435.468
1972	455.996
1973	471.952
1974	489.475
1975	487.137
1976	512.441

TABLE 2

Gradually increasing complexity of polynomials of three variables

General form af Polynomial: $y=a_0+a_1x_1+a_2x_1^2+a_3x_2+a_4x_1x_2+a_5x_2^2+a_6x_3+a_7x_1x_3+a_8x_2x_3+a_9x_3^2$

a_1	x_1^2	x_2	x_1x_2	x_3^2
$y_1=a_0+a_1x_1$	$y_2=a_0+a_1x_1^2$	$y_4=a_0+a_1x_2$	$y_8=a_0+ax_1x_2$	$y_{256}=a_0+a_1x_3^2$
	$y_3=a_0+a_1x_1+a_2x_1^2$	$y_5=a_0+a_1x_1+a_2x_2$	$y_9=a_0+a_1x_1$	$y_{257}=a_0+a_1x_1+a_2x_3^2$
		$y_6=a_0+a_1x_1^2+a_2x_2$	$+a_2x_1x_2$	—
		$y_7=a_0+a_1x_1+a_2x_1^2+a_3x_2$	$y_{10}=a_0+$ ...	—
			$y_{11}=a_0+$ ...	—
			$y_{12}=a_0+$ ...	$y_{511}=a_0+a_1x_1+a_2x_1^2+a_3x_2$
			$y_{13}=a_0+$ ...	$+a_4x_1x_2+a_5x_2^2+a_6x_3+a_7x_1x_3$
			$y_{14}=a_0+$ ...	$+a_8x_2x_3+a_{9x}^2$
			$y_{15}=a_0+$ ...	General rule of total number of combination is 2^n-1.
				Where n is the number of terms containing the variables in the general form of polynomial.

where $Y_{tab}(i)$, $i=1, 2,..., N$ years are the tabulated values of the variable in the interpolation region and $Y_{d.m}(i)$ are the values of the variable obtained from the model.

Data points are divided into two sequences A and B. Points of the sequences are selected alternatively. Time instances are taken as $t_1+1.0$ for 1947, $t_2=2.0$ for 1948 and so on.

The models from table 2 are tested comprising of three variables x_1, x_2 and x_3 for all the data points, data points of sequence A and data points of sequence B respectively.

The model of annual tea crop production is obtained as

$$y=535.455-1.51486\ x_1+0.002317\ x_1^2 \\ -0.19305\ x_2+6.693114\ x_3+0.6657\ x_3^2 \qquad ...(4)$$

In finite difference form,

$$Y_{k+1}=535.455-1.51486\ Y_k+0.002317\ Y_k^2 \\ -0.19305\ Yk_2-1+6.693114\ t_k \\ +0.06657\ t_k \qquad ...(5)$$

Corresponding minimum value of the integral square error is 0.000158.

Table 3 shows the actual and the predicted values from model of tea production from 1960 to 1976 with extrapolation upto 1981 on the basis of the base year of 1976.

3. Prediction of Predictions

By changing the base year and predicting for years ahead of the base year a series of predictions can be obtained. It is observed that the predictions are diverging in nature. Since the prediction changes with the change of the base year, the process is non-stationary and the operational law guiding the process varies with time. As the prediction varies according to the prediction time, the base year, it is intuitively assumed that the prediction is also a function of the prediction time tp_o. The model is now expressed in the form of finite difference equation as stated below :

$$Y_{k+1}=f(Y_k, Y_{k-1}, t_k, tp_o\ k) \qquad ...(6)$$

Let

$$Y_{k+1}=y$$
$$Y_k=x_1$$
$$Y_{k-1}=x_2$$
$$t_k=x_3$$
$$tp_o\ k=x_4$$

So,

$$y=f(x_1, x_2, x_3, x_4) \qquad ...(7)$$

Here $tp_o\ 1=1$ for 1966, $tp_o\ 2=2$ for 1967 and so on.

TABLE 3

Actual and predicted values of annual tea production

Year	*Actual value in million kgs.*	*Predicted value from model in million kgs.*
1960	321.077	331.870
1961	354.397	346.217
1962	346.735	351.808
1963	346.413	355.032
1964	372.435	363.897
1965	366.374	372.736
1966	375.983	380.107
1967	384.759	387.030
1968	402.489	402.563
1969	393.588	405.962
1970	418.517	419.507
1971	435.462	429.004
1972	455.996	453.476
1973	471.952	456.769
1974	489.475	488.271
1975	487.137	494.411
1976	512.441	513.660
1977	—	535.067
1978	—	561.645
1979	—	595.526
1980	—	641.050
1981	—	707.300

A gradual increase in the complexity of polynomials of four variables is presented in table 4. The polynomial which gives minimum of the integral square error is obtained as,

$$Y_{k+1} = -52.31837 - 0.111366\, Y_{k-1} + 0.000294\, Y_k\, Y_{k-1} + 21.63834\, t_k + 0.0024847\, Y_k\, t_k - 0.0031771\, Y_{k-1}\, t_k - 25.17740\, tp_o k + 0.050886\, Y_k\, tp_o k - 0.012901\, Y_{k-1}\, tp_o k \quad \ldots(8)$$

Corresponding value of integral square error is 0.0000003106.

TABLE 4

Gradually increasing complexity of polynomials for four variables

General form of polynomial : $y=a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_3+a_5x_1x_3+a_6x_2x_3+a_7x_4+a_8x_1x_4+a_9x_2x_4+a_{10}x_3x_4$

x_1	x_2	x_1x_2	x_3	x_1x_3	x_3x_4
$y_1=a_0+a_1x_1$	$y_2=a_0+a_1x_2$	$y_4=a_0+a_1x_1x_2$	$y_8=a_0+a_1x_3$	$y_{16}=a_0+a_1x_1x_3$	$y_{512}=a_0+a_1x_3x_4$
	$y_3=a_0+a_1x_1+a_2x_2$	$y_5=a_0+a_1x_1+a_2x_1x_2$	$y_9=a_0\times a_1x_1+a_2x_3$	$y_{17}=a_0+\cdots$	$y_{513}=a_0+a_1x_1+a_2x_3x_4$
		$y_6=a_0+a_1x_2+a_2x_1x_2$	...	...	...
		$y_7=a_0+a_1x_1+a_2x_2+a_3x_1x_2$	...	...	...
			$y_{15}=a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_3$	...	$y_{1023}=a_0+a_1x_1+a_2a_2+a_3x_1x_2+a_4x_2+a_5x_1x_3+a_6x_2x_3+\cdots+a_{19}x_3x_4$

This model can be used for prediction of tea crop for 5 to 7 years ahead of time. This model exhibits over-fitting tendencies for prediction of tea crop for more than 7 years ahead of time. This is because of the fact that the effect of different negative feedbacks operating in the process become evident only in the long run.

In the second part of the work we are suggesting an equilibrium state model whtch can be successfully used for longterm prediction.

4. Long-term equilibrium state prediction model for annual production of Indian Tea

When the data for the last ten years from 1967 to 1976 are analyed it appears that the average yearly growth rate of Indian tea is about 3.17 per cent. This gives rise to a doubling time of crop production of about 22 years i.e. the tea production will be doubled from its present value of 512.441 million Kgs in 1976 to 1024.882 million Kgs in 1998. It may be recognised that this growth rate is a complicated process and is likely to change slowly rather than quickly. The figure 1 shows the annual Indian tea production both in the interpolation and extrapolation regions. The trend has been extrapolated heuristically.

It is observed that in the interpolation region the growth rate is almost exponential in nature. This exponential growth rate is a product of many interacting positive feedback loops. For example, to sustain the growth rate of production consumption of chemical fertilisers and pesticides will increase proportionately and more irrigation will be necessary. In the long run all these have a negative feedback effect. Use of increased chemical fertilisers will reduce the inherent soil fertility. Use of increased pesticides will pollute the environment and as a result it will disturb the ecological balance.

It was observed [1] that positive feedback loops operating without any severe constraints generate exponential growth. But any physical system is finite. There must be constraints which can act to stop the exponential growth. These constraints are negative feedback loops. The negative feedbacks become stronger and stronger as the growth approaehcs the ultimate limit, or the carrying capacity of the system's environment. Finally the negative feedback loops balance or dominate the positive ones and the growth comes to an end.

It is observed [8] that our finite world is capable of supporting a population twice the present figure. And the doubling of population will come around 2000 A.D. Policy should be made so as to

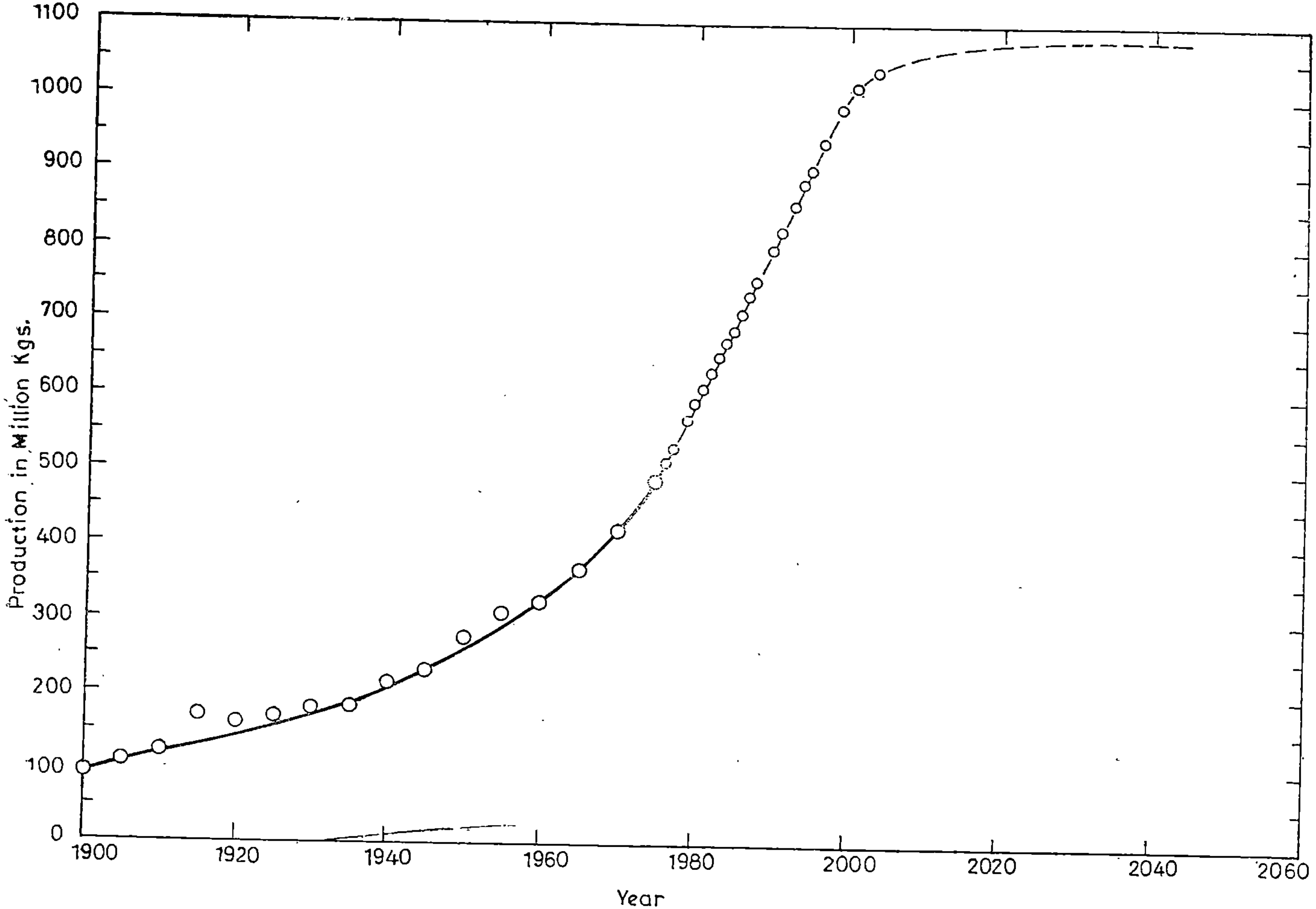

Fig. 1. Annual Tea Production in Interpolation & Extrapolation regions.

achieve an equilibrium state in population around 2000 A.D. Keeping this veiw in mind we must strive for restricted growth to the tune of doubling the production sometimes around 2000 A.D. Tea Research Association of India has also suggested to achieve doubling of production around 2000 A.D. [7]

5. Development of the Equilibrium State Model

A second order finite difference equation model isused for long-term prediction of annual tea crop production. Imposing a limit to growth after 2000 A.D. the crop produbtion has been extrapolated beyond 2000 (A.D.)

The model postulated as

$$Y_{k+1} = f(Y_k, Y_{k-1}, t_k)$$

Polynomials of gradually increasing complexity for three variables are shown in table 2. Model of optimum complexity which gives minimum of integral square error criterion is found as

$$\begin{aligned} Y_{k+1} = & -16.6933 + 0.25448\ Y_k + 0.001239\ Y_k^2 \\ & + 0.861295\ Y_{k-1} \\ & - 0.001339\ Y_k\ Y_{k-1} + 0.098488\ t_k \end{aligned} \quad \ldots(10)$$

The corresponding value of integral square error is, 0.00009597.

5. Conclusion

In the first part of the work the model has been obtained for medium-term prediction (5 to 7 years) of annual Indian tea crop production. It is found that annual tea crop production follows a time varing law. Consequently a model which is time variant is obtained. For long term prediction this model is found to exhibit overfitting tendencies. To avoid the overfitting an equilibrium state model has been postulated. Imposition of limit on growth in this model does not mean stagnation in tea industry. After doubling of production around 2000 A.D. it is suggested to center our research activities for quality tea instead of increasing the quantity. It is hoped that this policy would help in a harmonical balance in ecological conditions.

Acknowledgement

The author extends his appreciation to the referee of this paper for his valuable suggestions and comments. The partial financial support of Tea Research Association of India is gratefully acknowledged. The assistance and co-operation of the computer centre of Kiev Polytechnical Institute, USSR are thankfully appreciated. The auther thanks Academician A.G. Ivakhnenko who taught him with great patience the theory of computer aided self-organisation of mathematical models.

REFERENCES

[1] Forrester Jay W., : *World Dynamics*, Wiight-Allen Press Inc., Cambridge, Massachusetts, USA, 1971.

[2] Ivakhnenko A.G., : "Long-term prediction and control of complex processes," *Tekhnika, Kiev*, 1975 (in Russian).

[3] Ivakhnenko A.G., & et.el., : "Long-term prediction of random processes by GMDH algorithms using the unbiasedness and balancc of variables criteria, Part 4," *Soviet Automatic Control.* Vol. 9, July-August 1976.

[4] Ivakhnenko A.G., : "Present state of the theory of computer aided self-organisation of mathematical models," *Soviet Automatic Control*, Vol. 8, September-October, 1975.

[5] Ivakhnenko A,G., : "Heuristic self-organisation in problems of engineering cybernetics," *Automatica*, Vol. 6, pp,207-219, 1970.

[6] Ivakhnenko A.G., : "The group method of data handling in long range forecasting," *Technological forecasting and social change*, Volume 12, Numbers 2/3, August 1978, Elsevier, New York.

[7] Jain N. J., : "Towards-higher productivity of mature tea, Keynote address," 28th Tocklai Conference 24.11.1977.

[8] Meadows Donella H., et.el., : "The limit to growth," *Patomac Associates*, New American Library, New York, 1972.

[9] Mustafi A. & Chaudhuri A.S., : "Use of multilayer group method of data handing for predicting tea crop production." to be published in the *Journal of Indian Society of Agricultural Statistics*.

[10] *Tea Statistics*, : J. Thomas and Co. Pvt. Ltd., Calcutta, 1877.