CHAPTER IV

ELECTRICAL ENERGY CONSUMPTION MODEL WITH INTERACTING PARAMETERS BY A LEARNING IDENTIFICATION ALGORITHM

CHAPTER IV

ELECTRICAL ENERGY CONSUMPTION MODEL WITH INTERACTING PARAMETERS BY A LEARNING IDENTIFICATION ALGORITHM

4.0 Introduction

The levels of energy consumption reveal the levels of cultural and economic development of different countries. Techno-economic and socio-economic parameters of energy system are interrelated. The energy system has distinct cybernetic features. Venikov has pointed out that deep-lying feedback paths exist in energy system. Thus it is understandable that significant interaction exists between energy consumption and different techno-economic parameters.

This paper presents a mathematical description of annual energy consumption with population, gross national product, gross domestic saving and gross domestic capital formation as exogenous variables in the form of a polynomial of optimum complexity with the help of a method of applied cybernetics commonly known as multilayer group method of data handling algorithm (GMDH). The method has the potentiality of identifying, implicitly in a learning identification environment, the interactions and feedbacks of interactions of different input parameters on the output of the process.

To give a mathematical description of the annual energy consumption as a function of a set of exogenous variables interrelated with one another through deep-lying feedback paths is a complex process. Modern Control theory based on differential or difference equations is not adequate to describible the process. In view of this difficulty, the method of modelling applied here uses a technique of self-organisation. This GMDH algorithm of self-organisation involves generation and comparison of different regression polynomials by using all possible combinations of input variables and selection therefrom of the best possible ones according to the criterion of minimum integral square error defined in equation (4.1.11). The GMDH technique is found to simulate adequately the input-output relationship of the complex process of annual electrical energy consumption as a function of a set of input variables (input set is not exhaustive).

4.1.0 Brief Description of GMDH

The multilayer group method of data handling involves the use of regression polynomials as the basic means of investigation of complex dynamical systems. The relevant polynomial is a regression equation which connects a value of an output variable with past or current values of output and input variables. The regression analysis in this case helps in evaluating the co-efficients of the polynomial by using the criterion of minimum integral square error. The polynomials are then treated in the same manner as that used for selection of seeds in agriculture as per a unique mathematical concept propagated and established by Ivakhnenko [36,37].

The salient features of GMDH as applicable in the case of multilayer selection process used in the present work are now briefly described here :

The process can be described by

$$\theta = f(x_1, x_2, \ldots, x_n) \tag{4.1.1}$$

and it involves the construction of several layers of partial descriptions using two input variables at a time, e.g., the first layer can be represented as

$$y_i = f(x_j, x_k) \tag{4.1.2}$$

for $j = 1, 2, \ldots, n$

with $k = 1, 2, \ldots, n \; (j \neq k)$

and $i = 1, 2, \ldots, m$ where $m = \binom{n}{2}$

Likewise the second layer can be represented as

$$z_{i'} = g(y_{j'}, y_{k'}) \tag{4.1.3}$$

for $j' = 1, 2, \ldots, m$

with $k' = 1, 2, \ldots, m \; (j' \neq k')$

and $i' = 1, 2, \ldots, p$ where $p = \binom{m}{2}$

and so on, it being noted that m and p are the numbers of pairwise combinations of the first and second layers respectively. The first step concerns the selection of input variables on the basis of strong correlation (defined in equation (4.1.7).

The co-efficients of the first layer of partial description are calculated by solving a system of normal Gaussian equations. The left hand side of the equations are set equal to the values of output at every point. After finding the values of the co-efficients, the values of the intermediate variebles are obtained. Then using the observed data the integral square error is determined for each of the variables. Only those variables which give low error (self selection threshold) are selected for subsequent use. These variables are retained and other variables are discarded. In the second layer of selection, the co-efficient of the partial description of the layer are calculated and the accuracy is checked again to select the accurate intermediate variables of the layer ; $z_1, z_2, \ldots z_p$. The process of selection continues so long as the integral square error on the whole data set comes to a minimum and then starts increasing in the next layer or, the integral square error approaches an asymptotic minimum.

The polynomial description of the process is obtained in the form of partial description of the intermediate variables of different layers. Eliminating the intermediate variables, the complete polynomial description of the process is obtained in the form of Gabor-Kolmogrov type polynomial as

$$\theta = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j x_i x_j$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_i a_j a_k x_i x_j x_k + \ldots \qquad (4.1.4)$$

Use of GMDH in formulating the electrical energy consumption process :

4.1.1 Input Data :

The input data from the Indian techno-economic scenario consisted of the annual energy generated (gross) in Mn.Kwh, population in million, gross national product at factor cost (Rs. Crores) at 1970-71 prices, gross domestic saving and gross domestic capital formation as percent of gross domestic product at market prices from 1960-61 to 1980-81. The complete data are shown in Table 4.1.1 . All the relevant data X(k) for the k-th year were rationalised as

$$x(k) = \frac{X(k) - X_{min}}{X_{max} - X_{min}} \qquad (4.1.5)$$

TABLE 4.1.1 INPUT DATA

YEAR	ENERGY GENERATED (GROSS) IN Mn. Kwh	POPULATION (MILLION)	GROSS NATIONAL PRODUCT AT FACTOR COST AT 1970-71 PRICES IN Rs. CRORES	GROSS DOMESTIC SAVING AS PER CENT OF GROSS DOMESTIC PRODUCT AT MARKET PRICES	GROSS DOMESTIC CAPITAL FORMATION AS PER CENT OF GROSS DOMESTIC PRODUCT AT MARKET PRICES
1960-61	20123	442.4	25424	12.7	16.9
1961-62	22957	452.2	26293	13.1	16.3
1962-63	26237	462.0	26834	14.5	17.1
1963-64	30212	472.1	28210	14.4	16.6
1964-65	33133	482.5	30399	13.6	16.2
1965-66	36826	492.3	28791	15.7	18.2
1966-67	40612	504.2	29081	16.3	19.7
1967-68	46215	515.4	31590	13.2	16.5
1968-69	51642	527.0	32460	14.1	15.4
1969-70	56544	538.9	34518	16.4	17.1
1970-71	61211	551.2	36452	16.8	17.8
1971-72	66384	562.5	37000	17.2	18.4
1972-73	70516	575.9	36599	16.2	16.9
1973-74	72796	588.3	38410	19.2	20.0
1974-75	76678	600.8	38857	18.2	19.1
1975-76	85926	613.2	42571	20.0	19.9
1976-77	95615	625.8	43124	22.0	20.4
1977-78	98928	638.4	46354	21.2	19.7
1978-79	110130	651.0	49343	24.4	24.6
1979-80	112820	663.6	46966	22.5	22.9
1980-81	119227	682.8	50507	22.2	24.2

4.1.2 Formulation of the process equation

The annual electrical energy consumption can be represented by a general form of process equation as

$$y(k) = f\left[y(k-1),\ y(k-2), \ldots, x_1(k), x_1(k-1), \ldots, x_2(k), x_2(k-1), \ldots, x_3(k), x_3(k-1), \ldots \right] \qquad (4.1.6)$$

where k, k-1, k-2,..., refers respectively to the current day, one day preceding the current day, two days preceding the current day and so on and $y(.)$, $x_1(.)$, $x_2(.)$, $x_3(.)$ and $x_4(.)$ are the rationalised data for annual electrical energy consumption, population, gross national product, gross domestic saving and gross domestic capital formation respectively.

The arguments having correlations with $y(k)$ are then selected for inclusion in the process equation of the basis of the correlation functions for the time shift λ, in years, defined as

$$\psi_{yx}(\lambda) = \frac{\sum_{i=1}^{N-\lambda} \left[\left(y(i) - \frac{1}{N} \sum_{k=1}^{N} y(k)\right) \left(x(i+\lambda) - \frac{1}{N} \sum_{k=1}^{N} x(k)\right) \right]}{\left[\sum_{i=1}^{N-\lambda} \left(y(i) - \frac{1}{N} \sum_{k=1}^{N} y(k)\right)^2 \sum_{j=1+\lambda}^{N} \left(x(j) - \frac{1}{N} \sum_{k=1}^{N} x(k)\right)^2 \right]^{\frac{1}{2}}} \qquad \ldots (4.1.7)$$

where N is the number of data points.

After such selection of arguments as having correlation with annual energy consumption the process equation becomes

$$y(k) = f[\, y(k-1), x_1(k), x_1(k-1), x_1(k-2), x_2(k), x_2(k-1), x_3(k) \,] \quad \ldots \; (4.1.8)$$

or denoting these respective arguments as $y(k) = y, y(k-1) = x_1'$, $x_1(k) = x_2'$, $x_1(k-1) = x_3'$, $x_1(k-2) = x_4'$, $x_2(k) = x_5'$, $x_2(k-1) = x_6'$ and $x_3(k) = x_7'$, the process equation becomes,

$$y = f\,(x_1', x_2', x_3', x_4', x_5', x_6', x_7') \qquad (4.1.9)$$

Table 4.1.2 shows the correlation co-efficients of different exogenous variables with the annual electrical energy consumption for different lagged instances in years.

4.1.3 First layer of selection

There are $\binom{7}{2} = 21$ possible combinations of selecting two arguments at a time out of seven. For every such combination, the partial regression equation is written as

$$y_a = \alpha_{0a} + \alpha_{1a} x_b' + \alpha_{2a} x_c' + \alpha_{3a} x_b' x_c' + \alpha_{4a} x_b'^2 + \alpha_{5a} x_c'^2 \qquad (4.1.10)$$

where $a = 1, 2, \ldots, 21$, while b and c are indices for all 21 combinations. And this lead to 21 systems of normal

TABLE 4.1.2

CORRELATION COEFFICIENTS OF DIFFERENT EXOGENOUS VARIABLES WITH THE ANNUAL ELECTRICAL ENERGY CONSUMPTION FOR DIFFERENT LAGGED INSTANCES IN YEARS.

Time instant (Year)	Annual Energy Consumption - Annual Energy Consumption	Annual Energy Consumption - Population	Annual Energy Consumption - Gross National Product	Annual Energy Consumption - Gross Domestic Saving	Annual Energy Consumption - Gross Domestic Capital Formation
0	1	0.995851	0.992315	0.947459	0.850142
1	0.982801	0.979509	0.973108	0.936665	0.848097
2	0.932136	0.936809	0.935295	0.876886	0.805719
3	0.834372	0.823633	0.832906	0.801000	0.742843
4	0.699787	0.680482	0.694260	0.683842	0.640432
5	0.516821	0.496483	0.542507	0.527507	0.492223
6	0.319394	0.277528	0.347854	0.400121	0.411760
7	0.130834	0.082781	0.144170	0.238482	0.411783
8	-0.059705	-0.108712	-0.042876	0.072807	0.260831
9	-0.244068	-0.298346	-0.237582	-0.174616	0.086300
10	-0.411211	-0.468869	-0.402052	-0.338780	-0.150841

Gaussian equations with matrices of the order 6 x 6.
The co-efficients α 's are then estimated by solving normal equation systems constructed from the data set. For estimating the co-efficients it is assumed that the equation error is very small, being distributed with zero mean, constant variance and also uncorrelated with the inputs. The second assumption is that for the construction of the model the inputs and outputs are known exactly without any measurement error.

The accuracy of every variable y_a is calculated by using the entire data set. From all variables seven more accurate ones are chosen which give low values of integral square error criterion as

$$ISE = \frac{\sum_{i=1}^{N} \left[y_{observed}(i) - y_{modelled}(i) \right]^2}{\sum_{i=1}^{N} \left[y_{observed}(i) \right]^2} \qquad (4.1.11)$$

4.1.4 Selection of other layers

Seven intermediate variables of y_a layer chosen from the first layer give 21 combinations of two arguments of y_a layer. Again in the second layer these becomes

$$z_a = \beta_{oa} + \beta_{1a} y'_b + \beta_{2a} y'_c + \beta_{3a} y'_b y'_c + \beta_{4a} y'^2_b + \beta_{5a} y'^2_c$$

... (4.1.12)

where a = 1,2,..., 21 while b and c are indices of all 21 combinations. Calculation of the co-efficients of β and estimation of the accuracy of z_a are repeated as in the case of y_a.

The seven z_a variables are then chosen for the next layer u_a.

$$u_a = \gamma_{oa} + \gamma_{1a} z'_b + \gamma_{2a} z'_c + \gamma_{3a} z'_b z'_c + \gamma_{4a} z'^2_b + \gamma_{5a} z'^2_c \qquad (4.1.13)$$

In this way each layer is tested for accuracy by using the entire data set and on the basis of minimum integral square error criterion explained earlier. For all layers, variables on the left hand side of the equations are kept equal to the value of the output variable.

4.2.0 Illustration

It was observed that as the layer increases the integral square errors converge asymptotically to a very small value. To make the model suitable for practical application the integral square error of 3.408346E-04 at z_5 was considered for the point of termination of formation of process equation.

The changes of integral square error for different layers are shown in Fig.4.2.1. The integral square errors for different combinations in different layers are shown in Table 4.2.1 in ascending order.

The annual energy consumption in India has been identified by the polynomial as

$$y = z_5$$

$$z_5 = 7.4900222 - 0.3 + 1.428809 y'_1 - 0.496869 y'_6$$
$$- 35.759734 y'_1 y'_6 + 17.285701 y'^2_1 + 12.524885 y'^2_6$$

$$y'_1 = -0.031427 + 1.470818 x'_2 - 0.602856 x'_6 + 6.259670 x'_2 x'_6$$
$$- 3.220067 x'^2_2 - 2.814044 x'^2_6$$

$$y'_6 = 0.074196 + 0.934731 x'_4 - 0.131836 x'_7$$
$$+ 0.063213 x'_4 x'_7 + 0.089602 x'^2_4 + 0.156838 x'^2_7 \quad (4.2.1)$$

Fig.4.2.2(a) and 4.2.2(b) shows the observed values and errors between the observed and the modelled values of annual electrical energy consumption in India in rationalised unit.

Modelling errors are found to have a variance of 1.000014 and mean of 2.780174×10^{-4}, and are almost found to be uncorrelated for $i \neq j$. [89] The software developed in BASIC language is given in the Appendix as A2.1.

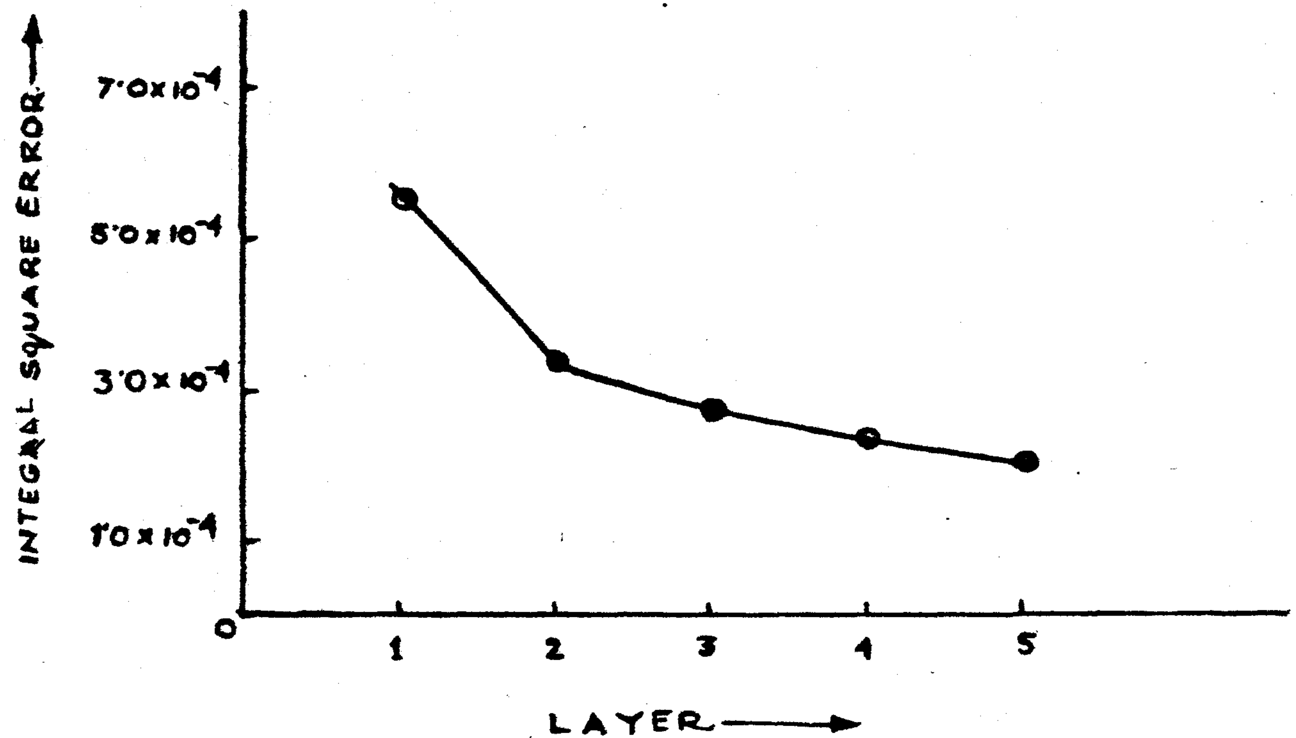

FIG. 4.2.1 INTEGRAL SQUARE ERROR IN DIFFERENT LAYERS.

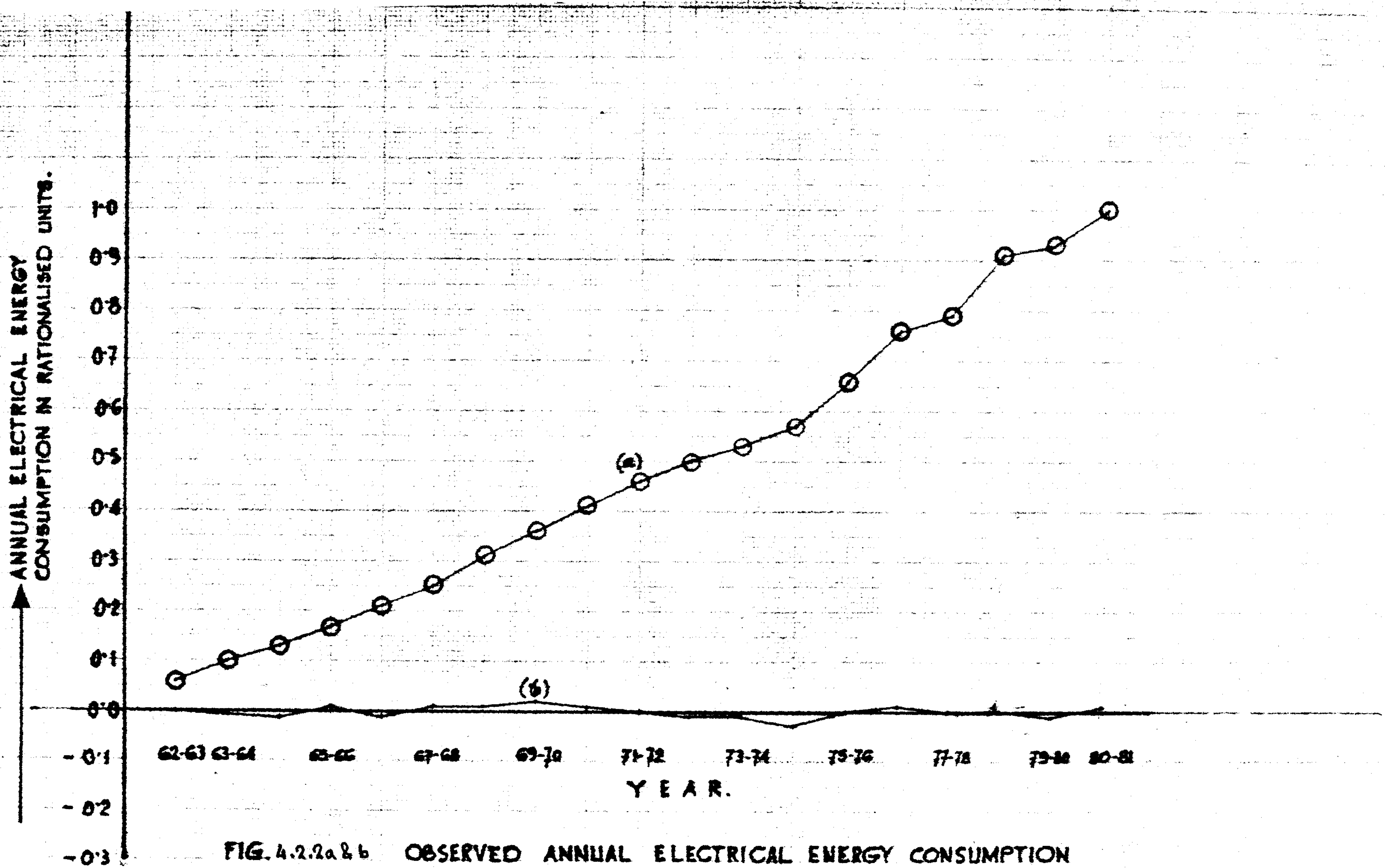

FIG. 4.2.2a & b OBSERVED ANNUAL ELECTRICAL ENERGY CONSUMPTION IN RATIONALISED UNITS & ERRORS BETWEEN THE OBSERVED & THE MODELLED VALUES.

TABLE 4.2.1

THE INTEGRAL SQUARE ERRORS FOR DIFFERENT COMBINATIONS IN DIFFERENT LAYERS.

Layer 1		Layer 2	
Combination	Integral Sq. Error	Combination	Integral Square Error
$x_2'x_6'$	5.593734 E-04	$y_1'y_6'$	3.408346 E-04
$x_3'x_6'$	5.714863 E-04	$y_1'y_5'$	3.427570 E-04
$x_4'x_6'$	5.796394 E-04	$y_1'y_4'$	3.428152 E-04
$x_2'x_7'$	6.330834 E-04	$y_2'y_4'$	3.887865 E-04
$x_3'x_7'$	7.097150 E-04	$y_3'y_6'$	3.898408 E-04
$x_4'x_7'$	7.248394 E-04	$y_3'y_4'$	3.903861 E-04
$x_3'x_5'$	7.431770 E-04	$y_2'y_5'$	3.916334 E-04

TABLE 4.2.1 (Continued)

Layer 3		Layer 4	
Combination	Integral Sq. Error	Combination	Integral Square Error
z_4z_7	2.808629 E-04	U_2U_4	2.432557 E-04
z_4z_5	2.996592 E-04	U_2U_3	2.454077 E-04
z_2z_5	3.053902 E-04	U_2U_5	2.482261 E-04
z_2z_6	3.058650 E-04	U_2U_6	2.492578 E-04
z_3z_5	3.076447 E-04	U_2U_8	2.537689 E-04
z_3z_6	3.082091 E-04	U_2U_7	2.569498 E-04
z_2z_7	3.198816 E-04	U_5U_7	2.738455 E-04