CHAPTER VI

FORECASTING OF DAILY DISSOLVED OXYGEN LEVELS OF A NON TIDAL RIVER BY SELF ORGANISATION OF MATHEMATICAL MODELS.

CHAPTER VI

FORECASTING OF DAILY DISSOLVED OXYGEN LEVELS OF A NON TIDAL RIVER BY SELF ORGANISATION OF MATHEMATICAL MODELS.

6.0 Abstract :

A simple dynamic model of the dissolved oxygen levels of a nontidal river has been obtained. The model structure has been identified with the help of a learning identification algorithm known as combinatorial group method of data handling technique. The model has been verified by simulation against field data. It is found that the major variations observed in the measured data have been adequately represented in the model.

6.1. Introduction :

The realistic modelling of dissolved oxygen levels (D.O) is important for controlling the water quality of a river. This paper presents a simple model for forecasting daily D.O. levels of a nontidal river. The D.O. levels are a major index of the general state of the aquatic environment. The important engineering decisions in water quality control relate to the determination, regulation, maintenance and control of the level of D.O. that is consistent with the multiple uses of the natural water bodies. Furthermore, it has been observed that even the temporary fluctuations of the D.O. level below the critical values can cause a long term and possibly irreversible damage to the ecosystem.

This work introduces a technique of technical cybernetics classified as the theory of self organisation of mathematical models and commonly known as the combinatorial group method of data handling algorithms. With the help of this method it has been possible to formulate mathematical models of complex processes with prediction optimisation.

6.2. The Concept of Self-Organisation :

When the model complexity gradually increases the computer finds by shifting the different models, the minimum of a selection criterion 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 applicable model of obtimum complexity for daily forecasting of D.O. levels of a nontidal river. The model has been verified by field measurement of dissolved oxygen collected over a 80-day period from the river Cam in Eastern England.

Model Development :

The daily measured data for D.O. concentration of a stream are

$P(i) \quad i = 1, 2 \ldots, \quad N1$, i being the time instants in days

The mean value is given by,

$$\overline{P}(N1) = \frac{1}{N1} \sum_{i=1}^{N1} p(i) \qquad \ldots \qquad (6.1)$$

Auto covariance of the data at lag instant k is given by

$$GA(k) = \frac{1}{N1-k} \sum_{i=1}^{N1-k} \left[p(i) - \overline{p}(N1) \right] \left[P(i+k) - \overline{p}(N1) \right] \quad \ldots (6.2)$$

Where $k = 0, 1, 2, \ldots, M$ and $M < \frac{1}{4} N1$

The normalized co-efficients of covariance are given by

$$RM(k) = \frac{GA(k)}{GA(0)} \qquad \ldots \qquad (6.3)$$

The estimate of the normalised power density spectra for the data are given by,

$$PS(w_h) = \frac{2}{\pi} \sum_{k=0}^{M} E_k \, RA(k) \, \cos w_h k \quad \ldots \quad (6.4)$$

where $w_h = 2\pi f_h$ and $f_h = \frac{h}{2M}$, $0 \leq f_h \leq 0.5$

and $h = 0, 1, 2, \ldots, M$

E_k has been defined as the weight for window correction and may be taken as

$$E_k = \begin{cases} 1.0 & \text{for } 0 < k < M \\ 0.5 & \text{for } k = 0, M \end{cases} \quad \ldots \quad (6.5)$$

These raw estimate of power spectral density are smothed by using Hamming Window to obtain the final estimates of the power spectrum. The smoothed estimates of the ordinates of the power spectrum are given by

$$S(w_h) = 0.54 \, PS(w_h) + 0.46 \, PS(w_1) \quad \text{for } h = 0$$

$$S(w_h) = 0.23 \, PS(w_{h-1}) + 0.54 \, PS(w_h) + 0.23 \, PS(w_{h+1})$$

$$\text{for } o < h < M \quad \ldots \quad (6.6)$$

$$S(w_h) = 0.54 \, PS(w_h) + 0.46 \, PS(w_{h-1}) \quad \text{for } h = M$$

The periodicity in terms of fundamental and its harmonics can be estimated from the power spectral density versus frequency characteristics.

The non-stationary process model with deterministic periodicity can be expressed as

$$P(k) = U(k) + V(k) \quad \ldots\ldots\ldots \quad (6.7)$$

Where

$U(k)$ is the deterministic trend components,

$V(k)$ is the stationary stochastic components

$k = 1, 2, \ldots,$ are time instances.

The deterministic periodic components can be represented as

$$U(k) = \sum_{r}^{R} \left(a_r \cos (r.2\pi f_o k) + b_r \sin (r 2\pi f_o k) \right) \quad \ldots (6.8)$$

Where f_o is the fundamental cycle per instance of time. the stochastic components V(k) can be expressed in autoregressive form as

$$V(k) = \sum_{n=1}^{N} C_{k-n} \; V(k-n) + \eta(k) \qquad \ldots \qquad (6.9)$$

Where $n = 1, 2 \ldots, N$ the prehistory interval and $\eta(k)$ the error sequence. The error sequence $\eta(k)$ is modelled in the moving average form as,

$$\eta(k) = \sum_{j=1}^{m} q_{k-j} \; \eta(k-j) + V(k) \qquad \ldots \qquad (6.10)$$

Where m is the prehistory interval of the moving average terms of $\eta(k)$. v(k) resembles to a white noise process with zero mean and constant variance

V(k) can be written as

$$V(k-n) = P(k-n) - U(k-n) \qquad \ldots \qquad (6.11)$$

Hence equation (6.9) can be written as

$$V(k) \sum_{n=1}^{N} C_{k-n} \; P(k-n) - \sum_{n=1}^{N} C_{k-n} \; U(k-n) + \eta(k) \qquad \ldots \qquad (6.12)$$

From equations (6.7) and (6.8) it follows that,

$$P(k) = \sum_{n=1}^{N} C_{k-n} \; P(k-n) + \sum_{n=0}^{N} \sum_{r=0}^{R} [d^{r}_{k-n} \cos (r 2\pi f_o (k-n)) + g^{r}_{k-n} \sin (r 2\pi f_o (k-n))] + \eta(k) \qquad \ldots \qquad (6.13)$$

In matrix from the process P(k) can be expressed as

$$P(k) = a^{T}(k-1) \; Z(k-1) + \eta(k) \qquad \ldots \qquad (6.14)$$

$$\text{Where } a^{T}(k-1) = [C_{k-1}, \ldots, C_{k-n}, \sum_{n=0}^{N} (d^{0}_{k-n}, \ldots, d^{R}_{k-n}, g^{0}_{k-n}, \ldots, g^{R}_{k-n})] \ldots (6.15)$$

and

$$Z(k-1) = [P(k-1), \ldots, P(k-N), 1, \ldots, \quad Sin(R2\pi f_0(k-N)), \ldots, Cos(R2\pi \; f_0(K-n)]^T \quad \ldots (6.16)$$

In functional form

$$P(K) = f\Big(P(K-1), P(K-2), \ldots P(K-N), Sin(2\pi f_0 K), Cos(2\pi f_0 K), Sin(2\pi f_0(K-1)), Cos(2\pi f_0(K-1)) \ldots Sin(2\pi f_0(K-N)), Cos(2\pi f_0(K-N)) \ldots Sin(R2\pi f_0(K-N), Cos(R2\pi f_0(K-N))\Big) + \eta(K) \quad \ldots (6.17)$$

For simplicity let us assume that $N = 2$ and $R = 1$

Then

$$P(K) = f\Big(P(K-1), P(K-2), Sin(2\pi f_0 K), Cos(2\pi f_0 K), Sin(2\pi f_0(K-1)), Cos(2\pi f_0(K-1)), Sin(2\pi f_0(K-2)), Cos(2\pi f_0(K-2))\Big) + \eta(K) \quad (6.18)$$

Let $P(K) = y$, $P(K-1) = x_1$, $P(K-2) = x_2$, $Sin(2\pi f_0 K) = x_3$
$Cos(2\pi f_0 K) = x_4$, $Sin(2\pi f_0(K-1)) = x_5$, $Cos(2\pi f_0(K-1)) = x_6$
$Sin(2\pi f_0(K-2)) = x_7$ and $Cos(2\pi f_0(K-2)) = x_8$

and $\eta(k)$ as residual ξ

The equation (6.18) becomes

$$Y = f(x_1, x_2, \ldots, x_8) + \xi \quad \ldots \quad (6.19)$$

The estimate of y as $\hat{y}$ is defined as

$$\hat{y} = f(x_1, x_2, \ldots\ldots, x_8) \quad \ldots \quad (6.20)$$

The function f(.) in equation (6.20) is sought in the class of quadratic polynomials on the basis of a table of polynomials of gradually increasing complexity of eight variables as shown in table 6.2.1 with the theory of self organisation of mathematical models.

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

$$ISE = \frac{\sum_{i=1}^{N1} (y_o(i) - y_m(i))^2}{\sum_{i=1}^{N1} (y_o(i))^2} \qquad \ldots \quad (6.21)$$

Where $y_o(i)$, i=1, 2, ..., N = days, are the observed values of the output variable in the interpolation region and $Y_m(i)$ are the values of the variables obtained from the model. For the model of optimum complexity the residual has almost the property of a random variable with correlation co-efficient.

$$\phi_{\xi\xi}(\lambda) = 1.0 \quad \text{for } \lambda = 0$$
$$\phi_{\xi\xi}(\lambda) = 0 \quad \text{for } \lambda \neq 0$$

where λ is the shift of instance

The correlation co-efficient is defined as

$$\phi_{\xi\xi}(\lambda) = \frac{\sum_{i=1}^{N1-\lambda} \left(P(i) - \frac{1}{N1} \sum_{i=1}^{N1} P(i)\right) \left(P(i+\lambda) - \frac{1}{N1} \sum_{i=1}^{N1} P(i)\right)}{\left[\sum_{i=1}^{N1-\lambda} \left(P(i) - \frac{1}{N1} \sum_{i=1}^{N1} P(i)\right)^2 \sum_{J=1+\lambda}^{N1} \left(P(j) - \frac{1}{N1} \sum_{i=1}^{N1} P(i)\right)^2\right]^{\frac{1}{2}}} \qquad \ldots (6.22)$$

6.3 Illustration :

Fig 6.3.1a shows the D.O. level at the down stream end of 4.7 km stretch of the River Cam in eastern England, collected over 80 days period during the summer of 1972. Fig 6.3.2 shows the power density spectra versus cycle per month characteristics of the observed data. It is evident that the fundamental component of the periodicity, is 0.03125 cycle per day and the harmonics are negligible.

From the correlation co-efficient versus shift of instances of time as shown in Fig 6.3.3 it is obvious that the process follows a maximum prehistory interval of 5 times instances in days of which the most significant prehistory interval $N = 2$.

A selected number of polynomials were tested and the polynomial of optimum complexity was selected as

$$Y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_3 + a_5 x_1 x_3 + a_6 x_2 x_3$$
$$+ a_7 x_4 + a_8 x_1 x_4 + a_9 x_2 x_4 + a \ldots\ldots \qquad (6.23)$$

In finite difference form with periodicity the equation (6.23) can be expressed as

$$\begin{aligned} P(k) = {} & 0.229857 + 0.892091\, P(k-1) + 0.344082\, P(k-2) - 1.751167\, P(k-1) \\ & P(k-2) + 0.078986 \times \sin 2\pi\, (0.03125)\, k \\ & - 0.039061\, P(k-1) \quad \sin 2\pi (0.03125)\, k + 0.290850\, p(k-2) \\ & \sin 2\pi (0.03125)\, k + 0.007663 \quad \cos 2\pi\, (0.03125)\, k \\ & - 0.021738\, P\, (k-1) \quad \cos 2\pi\, (0.03125)\, k \\ & - 0.001492\, P\, (k-2) \quad \cos 2\pi (0.03125)\, k \end{aligned} \qquad \ldots (6.24)$$

Table 6.2.1

Gradually increasing complexity of polynomials of eight variables.

General from of Polynomial :

$y=a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_3+a_5x_1x_3+a_6x_2x_3+a_7x_4+a_8x_1x_4+a_9x_2x_4+a_{10}x_3x_4+a_{11}x_5+a_{12}x_1x_5+a_{13}x_2x_5$ #

$+a_{14}x_3x_5+a_{15}x_4x_5+a_{16}x_6+a_{17}x_1x_6+a_{18}x_2x_6+a_{19}x_3x_6+a_{20}x_4x_6+a_{21}x_5x_6+a_{22}x_7+a_{23}x_1x_7+a_{24}x_2x_7$ +

$+a_{25}x_3x_7+a_{25}x_4x_7+a_{27}x_5x_7+a_{28}x_6x_7+a_{29}x_8+a_{30}x_1x_8+a_{31}x_2x_8+a_{32}x_3x_8+a_{33}x_4x_3+a_{34}x_5x_8$ + $a_{35}x_6x_8+a_{36}x_7x_8$.

x_1	x_2	x_1x_2	x_3	x_1x_3	...	x_3x_4	....	x_7x_8
					...	y_{512} =	....	....
$y_1=a_0+a_1x_1$	$y_2=a_0+a_1x_2$	$y_4=a_0+a_1x_1x_2$	$y_8=a_0+a_1x_3$	$y_{16}=a_0$ +		$a_0+a_1x_3x_4$		
	$y_3=a_1+a_1x_1$	$y_5=a_0+a_1x_1$ +	$y_9=a_0+a_1x_1$	$a_1x_1x_3$		$y_{513}=a_0+a_1x_1$		
	$+a_2x_2$	$+a_2x_1x_2$	$+a_2x_3$	$a_{17}=a_0+....$		$+a_2x_3x_4$		
		$y_6=a_0+a_1x_2$ #						
		$+a_2x_1x_2$	$y_{15}=a_0+a_1x_1$					
		$y_7=a_0+a_1x_1$	$+a_2x_2+a_3x_1x_2$			$y_{1023}=a_0+a_1x_1$ #		$y_{36}=a_0$ + $(2-1)$
		$+a_2x_2+a_3x_1x_2$	$+a_4x_3$			$+a_2x_2+a_3x_1x_2$ +		$+a_1x_1+a_2x_2$ #
						$a_4x_3+a_5x_1x_3$ +		$+a_3x_1x_2$ +....
						.... $+a_{10}x_3x_4$		$+a_{36}x_7x_8$

The General rule of total number of combination is 2^n-1 where n is the total number of terms containing the variables in the general form of the polynomial.

The model has integral square error as 6.6325 E - 03 and mean error as 2.8280 E - 03. Fig 6.3.1b shows the error of prediction. Fig 6.3.4 shows the correlation co-efficient versus shift of instances of time of the errors of prediction. It is observed that errors are almost uncorrelated for $\lambda \neq 0$.

6.4 Conclusion :

The forecasting model has been found to simulate adequately the major variations of the dissolved oxygen levels observed in the measured data.

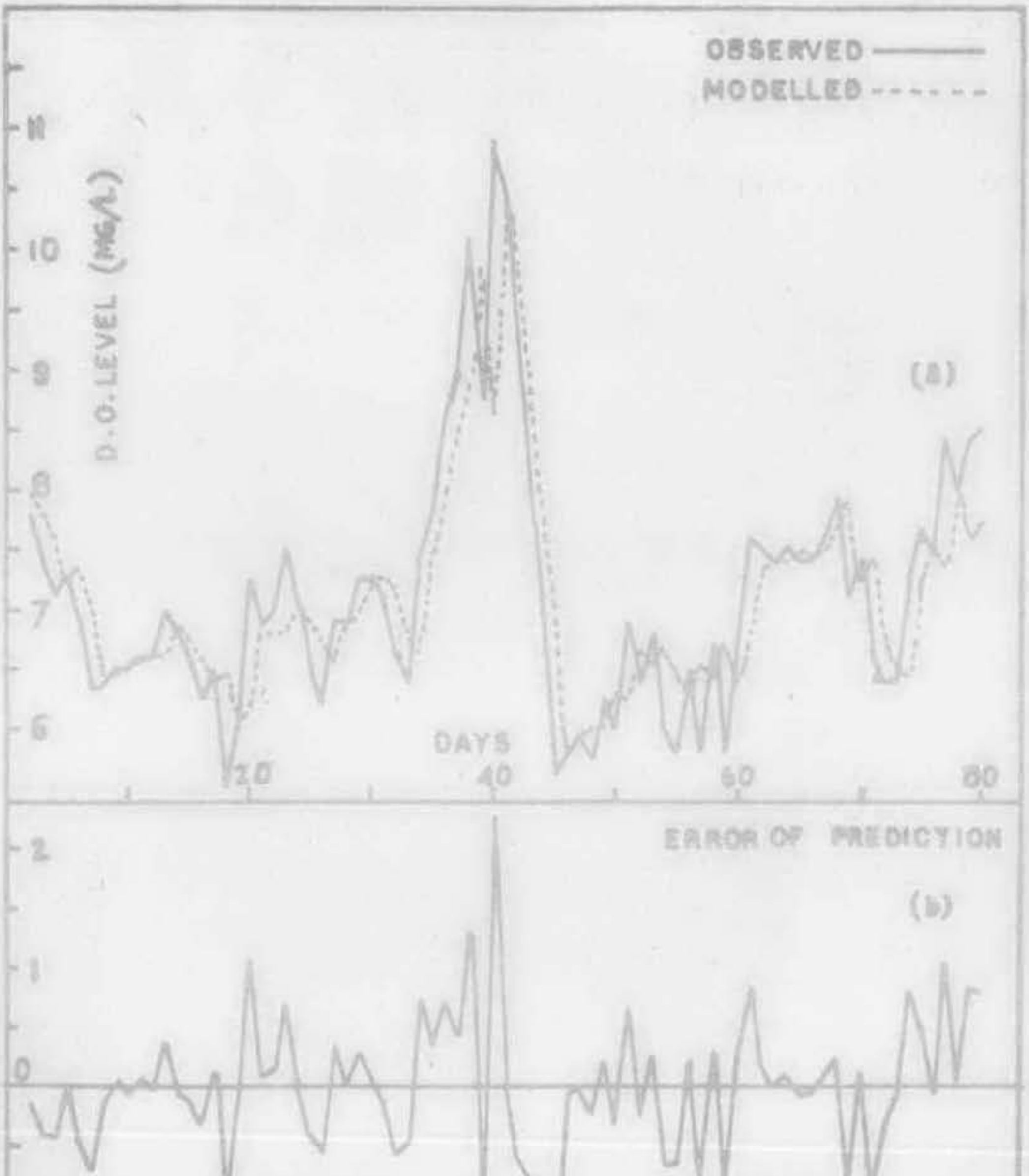

FIG 6.3.1 OBSERVED AND MODELLED VALUES OF D.O. LEVEL

POWER DENSITY SPECTRA OF D.O. LEVEL

FIG 6.3.2

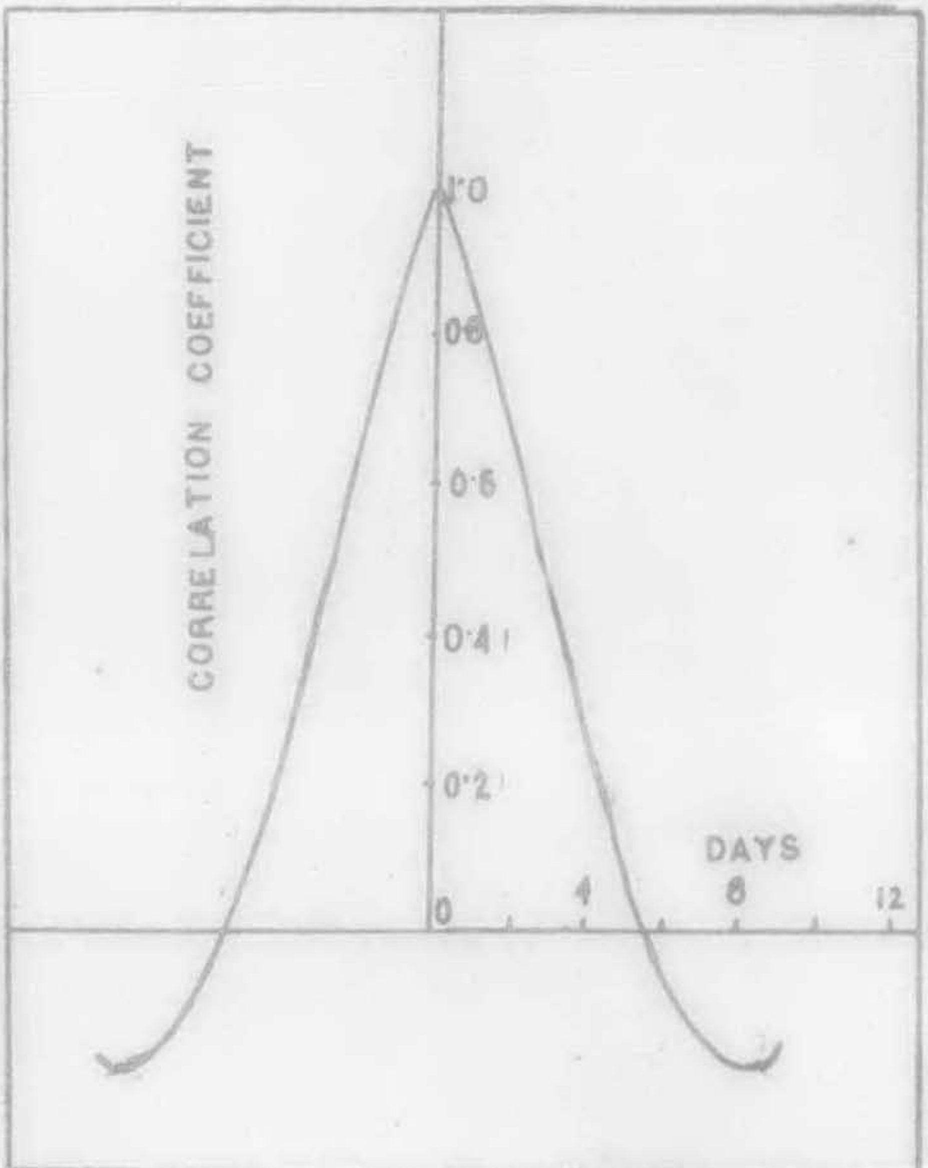

FIG 6.3.3 CORRELATION COEFFICIENT. D.O. LEVEL

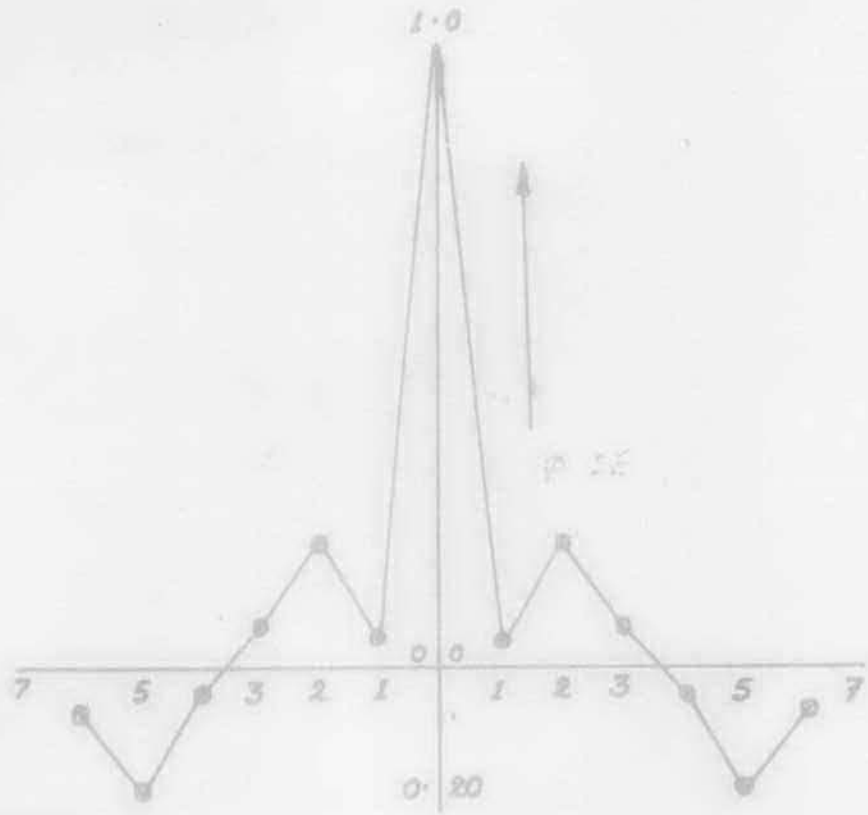

FIG. 6.3.4

CORRELATION COEFFICIENT VS SHIFT OF TIME FOR ERROR