



# LUND UNIVERSITY

## Maximum Likelihood Identification Applied to DO-BOD-ALGAE Models for a Freshwater Stream

Beck, Bruce

1974

*Document Version:*

Publisher's PDF, also known as Version of record

[Link to publication](#)

*Citation for published version (APA):*

Beck, B. (1974). *Maximum Likelihood Identification Applied to DO-BOD-ALGAE Models for a Freshwater Stream*. (Research Reports TFRT-3117). Department of Automatic Control, Lund Institute of Technology (LTH).

*Total number of authors:*

1

### General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00

MAXIMUM LIKELIHOOD IDENTIFICATION APPLIED  
TO DO-BOD-ALGAE MODELS FOR A FRESH-  
WATER STREAM

M.B. BECK

Report 7431(C) October 1974  
Lund Institute of Technology  
Division of Automatic Control

TILLHÖR REFERENSBIBLIOTEKET

UTLÅNAS EJ

## 1. INTRODUCTION.

The plentiful supply of nutrients for the growth of microorganisms in a reach of non-tidal river produces an aquatic environment in which so-called "blooms" of algae can be readily precipitated under favourable conditions. In general nutrients enter the stream in effluent discharges or from agricultural land drainage and surface runoff. It has been observed that during periods of prolonged, dry, sunny weather significant algal populations are established which not only increase the amplitude of diurnal oscillations and mean daily levels of dissolved oxygen (DO) but also increase the biochemical oxygen demand (BOD) in the river. Subsequently, with the onset of dull, cloudy weather the DO falls rapidly to low values, partly due to the cessation of photosynthetic production and partly a result of the oxygen demand created by mass algal deaths. Ultimately, if it is desired that the DO be maintained at suitably high levels the presence of excess nutrients and the stimulation of algal blooms constitute a control problem in water quality [7], [25].

The current discussion is an extension of studies on the identification and modelling of the dynamics of BOD-DO interaction [6]. The previous work included the collection of field data over an 80-day period during the summer of 1972 from a 4.7 km stretch of the River Cam in eastern England. An initial deterministic simulation led to the approximate validation of a dynamic model for DO and BOD with the introduction of pseudo-empirical terms to account for the effects of algae [8]; this involved the discrete low-pass filtering of data on the hours of sunlight incident on the system each day for the quantitative prediction of the algal population. However, considering the presence of random disturbances of the system and chance errors of measurement a more complete

tistical analysis of the model using a continuous-discrete version of the extended Kalman filter (see e.g. Jazwinski [16]) confirmed the preliminary identification of the model structure and allowed a more precise estimation of the parameters [9].

In this paper two other methods of identification are applied to the field data from the River Cam. The first is a black box maximum likelihood technique [3], for which it is assumed that the prediction of BOD and DO at the downstream end (i.e. output) of the reach of river are independent of each other and can be considered, therefore, as two multiple input-single output systems. The time-series models obtained in this manner agree with the a priori knowledge of the system and further indicate important features in the BOD and DO data which had hitherto remained ill-defined; it is found that the effects of the sunlight conditions (and hence, indirectly, the algal population) on the BOD and DO are asynchronous. Thus, assuming complete knowledge of the dynamics of all other interactions between BOD and DO a maximum likelihood method applied to stochastic differential equations with discrete-time observations enables the exploitation of the black box identification results in order to derive a more complete physical picture of the disturbances caused by algae.

Now, while it is possible to extend the original pseudo-empirical expression, what is really required is a more fundamental growth and decay model for algae. Bearing in mind the type of mathematical description used by Chen [10] in ecological modelling and the dynamics of micro-organism species in the activated-sludge and anaerobic digestion processes (see e.g. Andrews [1]), a Monod function [17] is hypothesised for algal population growth in which the sunlight conditions are assumed to be the rate limiting factor. Consequently, a combined DO-BOD-algae

model is proposed, in which the living and dead algal populations are described explicitly as separate state variables; except for the interaction of algae the DO-BOD model remains unaltered. Unfortunately, due to data restrictions\* the algal model can only be verified in a largely qualitative manner through the observed disturbances of the BOD and DO levels. The parameters are accordingly not uniquely identifiable without such necessary observations and, moreover, in view of the nonlinearities deterministic simulation is of merit in a verification study for the time-being.

---

\* i.e. the absence of observations on variables more directly related to the quantity of algae (e.g. chlorophyll-A, algal counts) and the absence of measurements of nitrogen- and phosphorus-bearing materials, which could also be growth-rate limiting.

## 2. THE A PRIORI MODEL FOR ALGAL POPULATION DYNAMICS.

The starting point for the present analysis is the model identified using the extended Kalman filter which is given by the set of lumped-parameter differential-difference equations [9],

$$\begin{aligned} \text{DO: } \dot{x}_1(t) = & - (K_1 + Q(t)/V)x_1(t) - K_2x_2(t) + \\ & + (Q(t)/V)u_1(t) + K_1C_S(t) + \\ & + K_3(I(t_k) - \bar{I}) + D_B \end{aligned} \quad (i)$$

$$\begin{aligned} \text{BOD: } \dot{x}_2(t) = & - (K_2 + Q(t)/V)x_2(t) + (Q(t)/V)u_2(t) + \\ & + K_4(I(t_k) - \bar{I}) + L_A \end{aligned} \quad (ii)$$

$$I(t_k) = I(t_{k-1}) + \frac{1}{\tau} \left[ u_3(t_k) (T(t_k) - \bar{T})/\bar{T} - I(t_{k-1}) \right] \quad (iii)$$

$$(I(t_k) - \bar{I}) = 0 \text{ for } I(t_k) < \bar{I} \quad (iv)$$

(the dot notation refers to differentiation with respect to time  $t$  (in days).)

Here  $x_1$  = concn. of DO at the downstream end (output) of the reach (mg/l)

$x_2$  = concn. of BOD at the downstream end (output) of the reach (mg/l)

$u_1$  = concn. of DO at the upstream end (input) of the reach (mg/l)

$u_2$  = concn. of BOD at the upstream end (input) of the reach (mg/l)

$u_3(t_k)$  = sunlight incident on the system during day  $t_k$  (hrs/day)

- $T$  = river water temperature ( $^{\circ}\text{C}$ )  
 $I(t_k)$  = "sustained sunlight effect" at day  $t_k$  \*  
 $\bar{I}$  = a threshold level for the sunlight effect \*  
 $\bar{T}$  = a mean river water temperature ( $^{\circ}\text{C}$ )  
 $K_1$  = reaeration rate constant for DO ( $\text{day}^{-1}$ )  
 $K_2$  = BOD decay rate constant ( $\text{day}^{-1}$ )  
 $K_3$  = coefficient for the sunlight effect in the DO equation \*  
 $K_4$  = coefficient for the sunlight effect in the BOD equation \*  
 $Q$  = volumetric flow-rate in the reach (cuft/day)  
 $V$  = mean volumetric hold-up in the reach (cuft)  
 $C_s$  = saturation concentration of DO in the reach (mg/l)  
 $\tau$  = a time constant for the discrete low-pass filter of the sunlight effect (days).  
 $D_B$  = rate of addition of DO to the reach by decomposition of bottom mud deposits ((mg/l)/day) (i.e. typically  $D_B \leq 0$  as described by eqn. I(i)).  
 $L_A$  = rate of addition of BOD to the reach by local surface runoff ((mg/l)/day).

The saturation concentration of DO,  $C_s$ , is generated from the following polynomial in  $T$ , the river water temperature ( $^{\circ}\text{C}$ ),

$$C_s(t) = 14.54 - 0.39T(t) + 0.01(T(t))^2 \quad (1)$$

The model given by eqn. I applies to a reach of river as defined in Figure 1.

---

\* No specific units are assigned to these quantities owing to the dimensional anomaly of eqn. I(iii).

The experimental field data from the River Cam and the derivation and identification of model I are presented in detail elsewhere [6], [8], [9]; however, it is pertinent to discuss the features of the sustained sunlight effect and its interaction with the DO and BOD equations. Firstly, the low-pass filter mechanism of eqn. I(iii), albeit a heuristic data manipulation, has a structure which would appear intuitively to agree with the true dynamics of an algal population. The time constant,  $\tau$ , chosen to be 4 (days) in previous work, implies that a prolonged period of persistent sunny weather is necessary for a population to establish itself and produce observable disturbances of the BOD and DO; in other words, the choice of  $\tau$  permits a certain degree of discrimination between the effects of isolated sunny days and sequences of consecutively bright days. However, low-pass filtering the sunlight data gives  $I(t_k) > 0$  for all  $t_k$  and thus the constraint of eqn. I(iv) is introduced so that the effects of algae are only discernible in the BOD and DO when  $I(t_k)$  is greater than an estimated threshold level,  $\bar{I}$ .

The temperature coefficient  $(T(t_k) - \bar{T})/\bar{T}$  in eqn. I(iii) is included for completeness, since the deterministic simulation responses of model I given in figure 2 and with the parameter values of table 1, are based

Table 1 - Parameter values for model I (from [9])				
$K_1$	0.17	$K_4$	0.32	$V$ 5.4 ( $10^6$ )
$K_2$	0.32	$\bar{I}$	6.0	$\tau$ 4
$K_3$	0.31	$\bar{T}$	8.0	$L_A$ 0.0
				$D_B = \begin{cases} -2.7 & \text{for } 0 \leq t \leq t_{19} \\ -0.4 & \text{for } t > t_{19} \end{cases}$

on this version of the model. However, omission of such a coefficient has almost negligible effect on the respon-



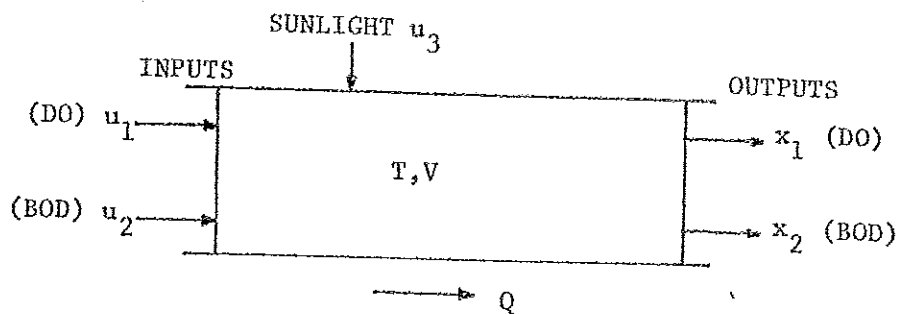


Figure 1 Schematic representation of a reach of river (all variables as defined for eqn I).

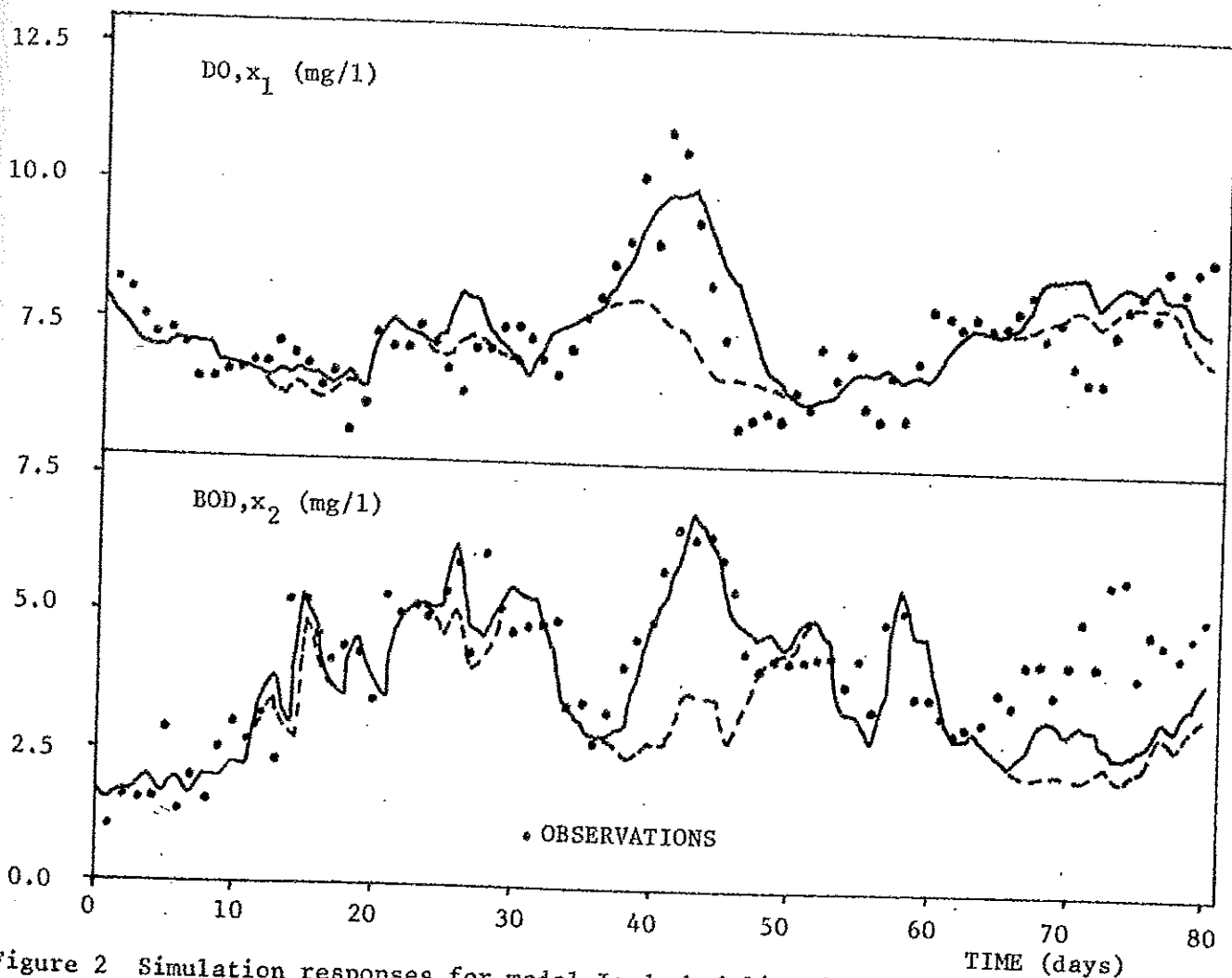


Figure 2 Simulation responses for model I; dashed line denotes responses with no algal effects included.

ses over the "critical" periods  $t_{36} \rightarrow t_{48}$  and from  $t_{60}$  onwards (approximately) where algal effects are significant.

Clearly, the sustained sunlight effect only describes those factors which were observed in the DO and BOD data and it does not quantify explicitly the dynamics of an algal population. Nevertheless, in addition to the River Cam experiment more recent work with a seventy-day period of observations from the Bedford Ouse has further validated this type of model for DO-BOD interaction [22], [23]. With more data on other variables available in the latter study it has been possible to improve the prediction of photosynthetic production of oxygen from algae by a term which utilises measurements of the chlorophyll-A content in the river water.

The usefulness of such a pseudo-empirical relationship as eqn. I(iii) is not in doubt if the objective is to model only the DO-BOD dynamics of a reach of river. Yet it would be more satisfying to clarify in mathematical terms the simplified pictographic relationships between BOD, DO and algae in figure 3 and to have some kind of quantitative view over the bloom conditions in an algal population. The following, therefore, is an initial step in such a direction; the key to the extended model is the observation that, unlike the single time constant,  $\tau$ , of eqn. I(iii), the interaction of the sustained sunlight effect with the BOD and DO is asynchronous.

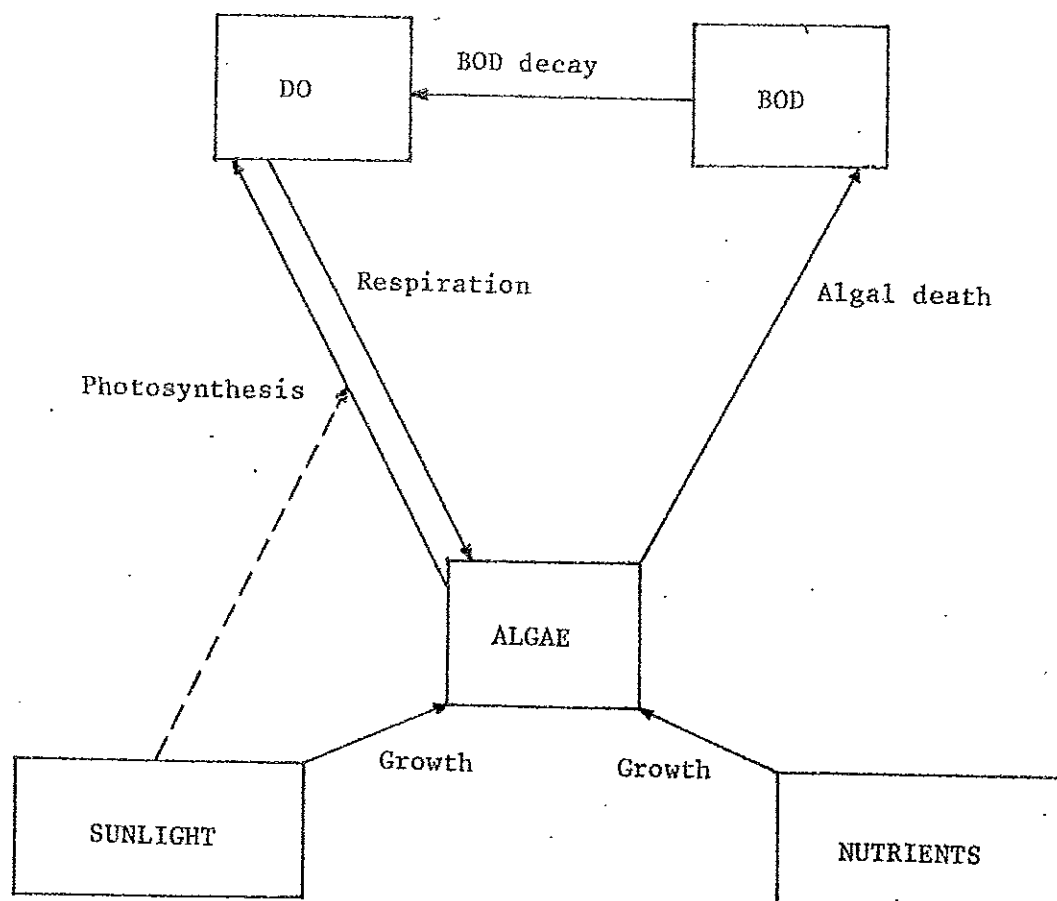


Figure 3 Some DO-BOD-algae inter-relationships.

### 3. MODEL IDENTIFICATION AND PARAMETER ESTIMATION.

The experimental data from the River Cam have now been examined with a variety of techniques, which number among them the Extended Kalman Filter (EKF), Maximum Likelihood (ML) applied to both black box and state-space models, and a Multivariable extension of the Instrumental Variable - Approximate Maximum Likelihood method (MIVAML). The results of such analyses are reviewed briefly and suitably organised according to the overall identification/parameter estimation/control system synthesis process.

In the first instance, from physico-chemical and biological reasoning it appeared that the dynamics of DO-BOD interaction could be described by a lumped-parameter second-order multivariable model [6]. Given such a priori knowledge of the system it was logical to exploit this information for identification purposes; in this sense, and because of the coupled nature of DO and BOD dynamics, it did not seem appropriate or necessary to start from a black box assumption. Thus, a deterministic simulation compared against the field data gives a first approximation of the model structure validity and parameter magnitudes [6],[8]. At this stage it is difficult to distinguish between those parts of the response errors which result either from structural errors or from measurement errors and stochastic disturbances of the system. Within the inherent statistical setting of the problem the recursive EKF scheme affords the possibility of "refining" both the model structure and the parameter estimates. The criteria employed to determine structural adequacy are open to question; in this case it is required that the

recursive estimates\* of the (assumed) time-invariant parameters display trajectories which are sensibly stationary once any initial transients have decayed away [6], [9]. In addition the statistical properties of the residual errors at the innovations process can be checked to ensure that they are reasonably consistent with the a priori assumptions about the nature of the stochastic effects in the system. Clearly, identification, as opposed to parameter estimation, is important since there is little to be gained from obtaining relatively efficient estimates of parameters in a model whose correct form has not been firmly established.

Since the original verification of the DO-BOD model was demonstrated a better perspective has been evolved on the position of the EKF in the general area of recursive methods of time-series analysis for model identification and estimation (see e.g. Young [24]). When the state-space structure of a model is incompletely known, the flexibility of the EKF is an advantage in the identification stage of the analysis. But for parameter estimation convergence cannot be guaranteed with the EKF; it is difficult to assess accuracy bounds and its very flexibility can lead to inefficiency when used to estimate the parameters that characterise the identified model structure. Consequently, a multivariable extension of the instrumental variable-approximate maximum likelihood method has been proposed for the parameter estimation phase of the analysis for state-space representations [26]. This latter, therefore, represents a secondary "refinement" of the parameter estimates and has been successfully applied to the field data from the River Cam and two reaches of the Bedford Ouse [23], [26].

---

\* Note that here the recursive form of the scheme is crucial to the identification process.

The application of ML techniques to the modelling of DO-BOD interaction reported herein is essentially a continuation of the identification process. It emerges from the EKF results that the coupling between the DO and BOD dynamics is relatively weak; hence it is not particularly restrictive to assume each to be an independent multiple input-single output system. And in any case the objective of the black box modelling analysis is primarily to clarify the time-dependency of the outputs on the input sunlight conditions; as the results indicate we shall interpret the parameter estimates as a measure of the statistical significance of such cause-effect relationships. This proves to be of great advantage and the physical implications of the results are then tested in an extension of the state-space model with an ML technique. The fact that some of the coefficients in the model are time-varying is not prohibitive since they refer to well-described dynamic features of DO-BOD interaction and it is superfluous to estimate them.

Thus, since we have digressed intentionally from the specifics of DO-BOD interaction, we might offer these comments on identification and estimation procedures. In an applications context absolute convergence and statistical "purity" of the estimates is not necessarily the ultimate goal. Moreover, this study confirms that black box and mechanistic models, together with the recursive or off-line approach, may be equally useful at some stage in the analysis; each has yielded a little more insight into the nature of the system.

Of course, which model would prove to be most useful for control purposes remains to be seen, although an initial feasibility study has reported results based on a state-space representation [6], [25]. It is possible, as we mention later, that a black box model has distinct advantages

but the question will only be resolved with the innovation of more reliable on-line instrumentation [7].

### 3.1. Maximum likelihood identification.

The method of maximum likelihood has been widely applied in identification and parameter estimation analyses; a detailed presentation of these techniques is unnecessary and the procedure is merely outlined as it applies to the two types of model structure used in this study.

#### Black box models.

The essential feature of a black box model is that it assumes no knowledge of physical relationships between a system's inputs and output other than that the inputs should produce observable responses in the output.

Thus, given the set of input/output data samples  $\{u_i(t_k), i = 1, 2, \dots, m; y(t_k); k = 1, 2, \dots, N\}$ , where  $u_i(t_k)$ ,  $i = 1, 2, \dots, m$ , are the  $m$  input signals and  $y(t_k)$  is the output signal, the identification problem is to find an estimate of the parameters of the system model [14]

$$A(z^{-1})y(t_k) = \sum_{i=1}^m B_i(z^{-1})u_i(t_k) + \lambda C(z^{-1})n(t_k) \quad (2)$$

in which  $n(t_k)$  is a sequence of independent, normal  $(0,1)$  random variables and  $z$  denotes the shift operator

$$z\{y(t_k)\} = y(t_{k+1}) \text{ etc.} \quad (3)$$

$A(z^{-1})$ ,  $B_i(z^{-1})$ ,  $i = 1, 2, \dots, m$ , and  $C(z^{-1})$  are the polynomials

$$\left. \begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \dots + a_n z^{-n} \\ B_i(z^{-1}) &= b_{i0} + b_{i1} z^{-1} + \dots + b_{in} z^{-n} \quad i = 1, 2, \dots, m \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \dots + c_n z^{-n} \end{aligned} \right\} \quad (4)$$

The residual errors of eqn. (2),  $\{\epsilon(t_k), k = 1, 2, \dots, N\}$ , defined by

$$C(z^{-1})\epsilon(t_k) = A(z^{-1})y(t_k) - \sum_{i=1}^m B_i(z^{-1})u_i(t_k) \quad (5)$$

are thus an independent and normal  $(0, \lambda)$  sequence. The logarithm of the likelihood function is now

$$L = -\frac{1}{2\lambda} \sum_{k=1}^N \epsilon^2(t_k) - N \log \lambda + \text{constant} \quad (6)$$

and the maximisation of  $L$  is equivalent to minimising the loss function

$$V(\theta) = \frac{1}{2} \sum_{k=1}^N \epsilon^2(t_k) \quad (7)$$

where  $\theta$  is the column vector of parameters in the model, eqn. (2),

$$\theta^T = [a_1, \dots, a_n, b_{10}, \dots, b_{1n}, b_{20}, \dots, b_{mn}, c_1, \dots, c_n] \quad (8)$$

(superfix  $T$  denotes the transpose of a vector or matrix). When  $\hat{\theta}$  has been found, such that  $V(\hat{\theta})$  is minimal, the



maximum likelihood estimate of  $\lambda$  is given by

$$\hat{\lambda}^2 = \frac{2}{N} V(\hat{\theta}) \quad (9)$$

Strictly speaking, the model eqn. (2) applies only to stable, linear, time-invariant systems. A more complete discussion of the model structure, the minimisation of the loss function  $V(\theta)$ , and the conditions for the estimates to be consistent, asymptotically normal and efficient are given in the original source references [3], [4], [14].

#### State-space models.

Alternatively, noting the structure of the a priori DO-BOD model, eqn. I, the identification and estimation procedure can be formulated as a problem of determining the parameters in the general stochastic differential equation

$$dx = Axdt + Budy + dv \quad (10)$$

A and B are the system and input matrices, respectively,  $x$  is the state vector and  $u$  the input vector; it is assumed that the initial state is a gaussian vector with mean value  $\bar{x}_0$  and covariance  $R_0$ .  $\{v(t), 0 \leq t \leq \infty\}$  is a Wiener process with incremental covariance  $R_1 dt$  and is also assumed to be independent of the initial state. Now let the observations,  $y$ , of the outputs of the system at discrete times  $t_0, t_1, \dots, t_N$ , be characterised by

$$y(t_k) = Cx(t_k) + Du(t_k) + e(t_k) \quad k = 0, 1, \dots, N \quad (11)$$

where C and D are matrices of appropriate dimensions. The

measurement errors  $e(t_k)$  are assumed to be independent and gaussian with zero mean and covariance  $R_2$ ; further, they are assumed to be independent of  $v(t)$  and the initial state.

In the particular case of the model described by eqns. (10) and (11), the logarithm of the likelihood function can be written as (c.f. eqn. (6))

$$L = -\frac{1}{2} \sum_{k=0}^N \varepsilon^T(t_k) R^{-1}(t_k) \varepsilon(t_k) - \frac{1}{2} \sum_{k=0}^N \log \det R(t_k) + \text{constant} \quad (12)$$

in which

$$\varepsilon(t_k) = y(t_k) - \hat{y}(t_k/t_{k-1}) \quad (13)$$

and  $\hat{y}(t_k/t_{k-1})$  denotes the conditional mean of  $y(t_k)$  given the set of observations up to and including  $y(t_{k-1})$ ;  $R(t_k)$  is the conditional covariance of the residual errors  $\varepsilon(t_k)$ , which, incidentally, are also the so-called innovations of the output process.

The computation of  $R(t_k)$ , and hence the likelihood function, are easily achieved in a recursive manner using the Kalman-Bucy filtering theory (see e.g. Åström [2]). The maximisation of  $L$  for the optimal estimation of a parameter vector  $\theta$  and other aspects of the identification procedure are discussed more fully in [5] and [12]. However, note that the implementation of this technique is more difficult than black box identification since more assumptions must be made about the system model structure and the noise statistics must be quantified by the covariance matrices  $R_0$ ,  $R_1$  and  $R_2$ .

#### 4. BLACK BOX MODELLING: THE ASYNCHRONOUS DISTURBANCE OF THE DO AND BOD DYNAMICS BY AN ALGAL POPULATION.

The black box analysis of the BOD and DO data was executed on a PDP-15 computer using a suite of programs specifically designed for this type of maximum likelihood estimation [14], [15]. The major concern is to identify the general time-dependency of the noisy observed outputs, DO and BOD, defined by figure 1 and

$$\left. \begin{array}{l} \text{DO: } y_1(t_k) = x_1(t_k) + e_1(t_k) \\ \text{BOD: } y_2(t_k) = x_2(t_k) + e_2(t_k) \end{array} \right\} k = 1, 2, \dots, N \quad (14)$$

upon the input sunlight conditions  $u_3$ . Although  $y_1$  and  $y_2$  show some slight dependence on other inputs, e.g. flow rate  $Q$  and temperature  $T$ , in practice most of the output variations can be adequately characterised by models of the general form,

$$\left. \begin{array}{l} y_1(t_k) = f_1\{y_1, u_1, u_3\} \\ y_2(t_k) = f_2\{y_2, u_2, u_3\} \end{array} \right\} \quad (15)$$

where the time arguments of  $u_3$  are to be identified.

The models suitably identified are all first-order and the estimates converged within twenty iterations of the process. Second-order models were attempted, but convergence for the DO was achieved only after some manipulation, such as assigning fixed zero values to several of the parameters, notably those of the  $C(z^{-1})$  polynomials. Second-order models for the BOD proved to be quite unsatisfactory with no convergence within forty iterations. This is not so surprising when one considers the physical characteristics of the reach of river from which the data are taken. The sampling interval of the data

$(t_k - t_{k-1})$  is 1.0 day, while the transportation delay in the reach varies approximately between 0.9 and 1.6 days over the observed 80-day period; thus, the outputs  $y_1(t_k)$  and  $y_2(t_k)$  would not be expected to be strongly dependent upon  $y_1(t_{k-2})$  and  $y_2(t_{k-2})$ , respectively, or, for the same reasons, upon  $u_1(t_{k-2})$  and  $u_2(t_{k-2})$ , respectively.

The difficulty of fitting second- or higher-order models, which would be a problem for the multiple delayed effects of the sunlight input  $u_3$ , is easily obviated by the generation of appropriately shifted time-series for the inputs [14], [15]. Other points to notice are that, since it was not possible to design an experiment with an artificial disturbance of the inputs, the system is not highly excited and the coefficients of the  $B_1(z^{-1})$  polynomials are accordingly poorly estimated. Two data points have also been adjusted: in the input BOD,  $u_2(t_{57})$  resulted from a thunderstorm and the effects of such a fast transient are not properly observable in the sampled output  $y_2(t_k)$ ; the DO output observation  $y_1(t_{34})$  has been linearly interpolated owing to a missing measurement. Finally, in view of a limited data record length the identification results are improved by the additional estimation of initial values  $y(t_0)$  of the outputs.

#### 4.1. The BOD model.

An initial test of the time-dependency of the output BOD,  $y_2$ , on the incident sunlight input is examined with the synthesis of several shifted sequences of  $u_3(t_k)$ ; the estimated coefficients are given in table 2. From the relatively poor estimates of  $b_{21}$ ,  $b_{31}$  and  $b_{51}$  it appears that  $y_2(t_k)$  can be assumed to be a function of  $u_3(t_{k-2})$  and  $u_3(t_{k-4})$  only. The results of such a final model structure are given in table 3 and figure 4, where

the pattern of the incident sunlight is also shown. The dynamics of the "peak" in the BOD between  $t_{36}$  and  $t_{48}$  are well-described; note also the similarity between the model error and the residuals  $\{\varepsilon(t_k)\}$  which implies that the stochastic effects in the system are largely those of measurement error (i.e.  $A(z^{-1})\alpha C(z^{-1})$ ).

If the identification should have a nice statistical interpretation, the residuals should be independent, stochastic variables. Figure 5 shows the autocorrelation

Table 2 - 1st-order model  $y_2(t_k)$ ;

$u_2(t_{k-1}), u_3(t_k), u_3(t_{k-1}), u_3(t_{k-2}), u_3(t_{k-3}),$   
 $u_3(t_{k-4})$

Coefficient	Estimate
$a_1$	$-0.749 \pm 0.066$
$b_{11}$	$0.102 \pm 0.043$
$b_{21}$	$-0.008 \pm 0.018$
$b_{31}$	$0.027 \pm 0.021$
$b_{41}$	$0.030 \pm 0.021$
$b_{51}$	$0.008 \pm 0.022$
$b_{61}$	$0.053 \pm 0.022$
$c_1$	$-0.506 \pm 0.133$
$\lambda$	$0.618 \pm 0.050$

Table 3 - 1st-order model  $y_2(t_k)$ ;

$u_2(t_{k-1}), u_3(t_{k-2}), u_3(t_{k-4})$

Coefficient	Estimate
$a_1$	$-0.751 \pm 0.062$
$b_{11}$	$0.102 \pm 0.042$
$b_{21}$	$0.048 \pm 0.015$
$b_{31}$	$0.060 \pm 0.020$
$c_1$	$-0.520 \pm 0.128$
$\lambda$	$0.627 \pm 0.051$

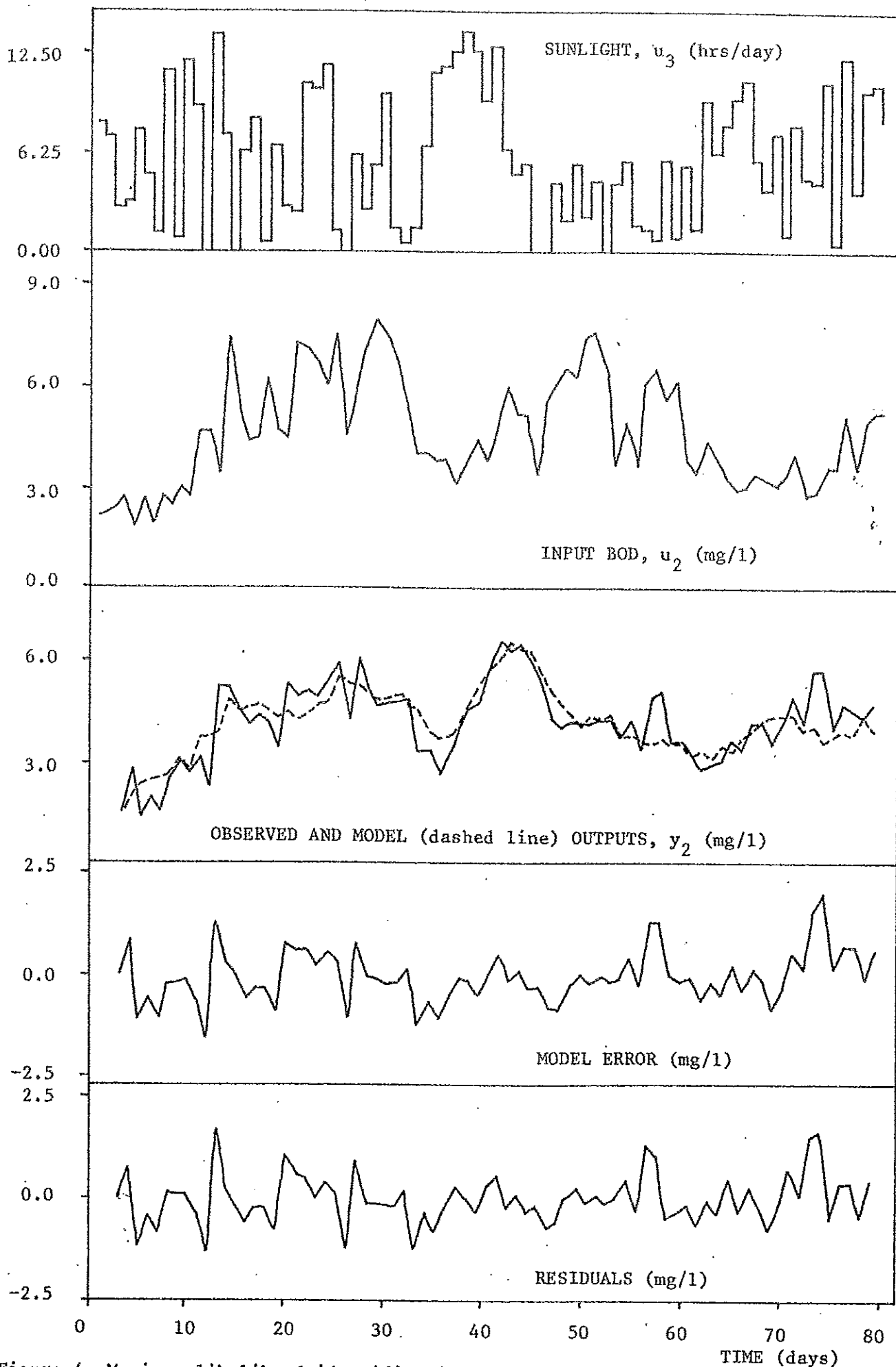


Figure 4 Maximum likelihood identification results for the BOD model of table 3.

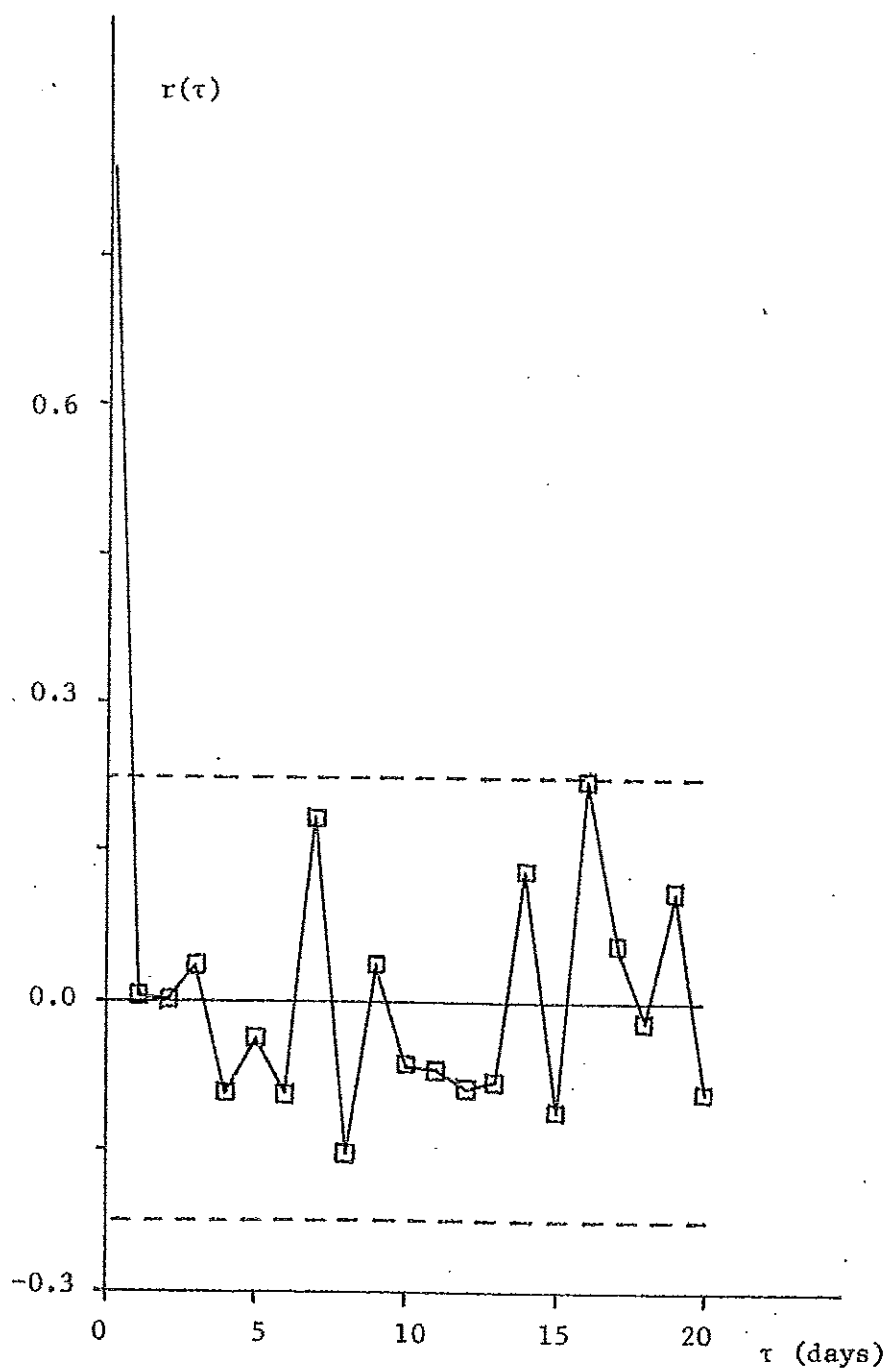


Figure 5 Sample autocorrelation function  $r(\tau)$  for ML BOD model of table 3; dashed lines denote limits for 95% confidence that  $r(\tau) = 0, \tau \neq 0$ .

function of  $\{\varepsilon(t_k)\}$  for the model and it can be seen that all points lie within the bounds for 95% confidence that the residuals are independent; similar checks (see [15]) using chi-squared goodness-of-fit tests also indicate that the residuals are satisfactorily normal and independent of the inputs  $u_2(t_k)$  and  $u_3(t_k)$ . Such examination of the residuals demonstrates whether the final model structure is in reasonable agreement with the original statistical assumptions made in the derivation of the general model form, eqn. (2).

#### 4.2. The DO model.

The approach to the identification of a DO model is the same as for the BOD model; in table 4 are given the estimates for the general test of the causal relationships between  $y_1$  and  $u_3$ . Note that the coefficient  $c_1$  is assumed equal to zero, in which case ML estimation reduces to a least squares analysis; it is permissible to do this in view of the poor estimation and instability of the  $C(z^{-1})$  polynomial. If the poor estimates  $b_{41}$ ,  $b_{51}$  and  $b_{61}$  and their associated time-series are assumed negligible, the estimation results for a final DO model are shown in table 5 and figure 6. As with previous identification results [9] there remains an inexplicable initial error  $(t_0 \rightarrow t_{19})^*$  in the model output; hence, incidentally, the quantification of  $D_B$  is given as time-dependent in table 1. It is also observable that there may be a trend in the

---

\* It has been proposed that this could be due to biased measurements of DO (and possibly BOD) [6], although recent MIVAML results suggest that flow conditions can account for this error [26]; any such significance of the flow-rate appears not to be identifiable in the present analysis.



Table 4 - 1st-order model  $y_1(t_k)$ ;
 $u_1(t_{k-1}), u_3(t_k), u_3(t_{k-1}), u_3(t_{k-2}), u_3(t_{k-3}),$   
 $u_3(t_{k-4})$ 

Coefficient	Estimate
$a_1$	$-0.740 \pm 0.074$
$b_{11}$	$0.161 \pm 0.052$
$b_{21}$	$0.056 \pm 0.017$
$b_{31}$	$0.040 \pm 0.017$
$b_{41}$	$-0.023 \pm 0.018$
$b_{51}$	$0.018 \pm 0.017$
$b_{61}$	$-0.004 \pm 0.017$
$c_1$	0
$\lambda$	$0.543 \pm 0.044$

Table 5 - 1st-order model  $y_1(t_k)$ ;
 $u_1(t_{k-1}), u_3(t_k), u_3(t_{k-1})$ 

Coefficient	Estimate
$a_1$	$-0.715 \pm 0.064$
$b_{11}$	$0.174 \pm 0.050$
$b_{21}$	$0.057 \pm 0.016$
$b_{31}$	$0.044 \pm 0.017$
$c_1$	0
$\lambda$	$0.554 \pm 0.045$

model error; however, correcting for this, or the inclusion of temperature as an input signal, produces marginal improvement in the model prediction and the statistics of the residuals. Nevertheless, the autocorrelation function of  $\{\varepsilon(t_k)\}$  for the model of table 5 is given in figure 7 and this and other statistical tests of the residuals are satisfactory.

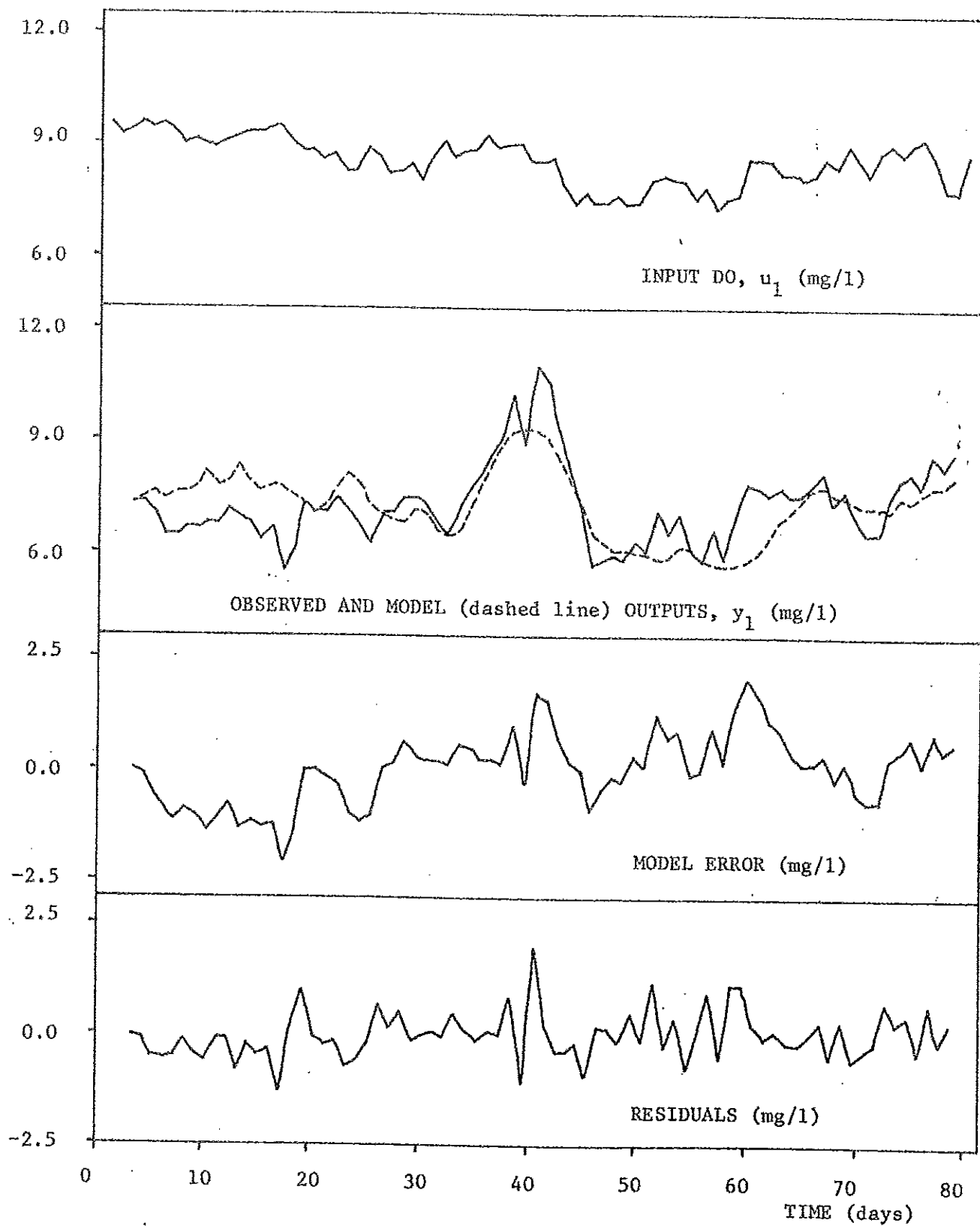


Figure 6 Maximum likelihood identification results for the DO model of table 5.

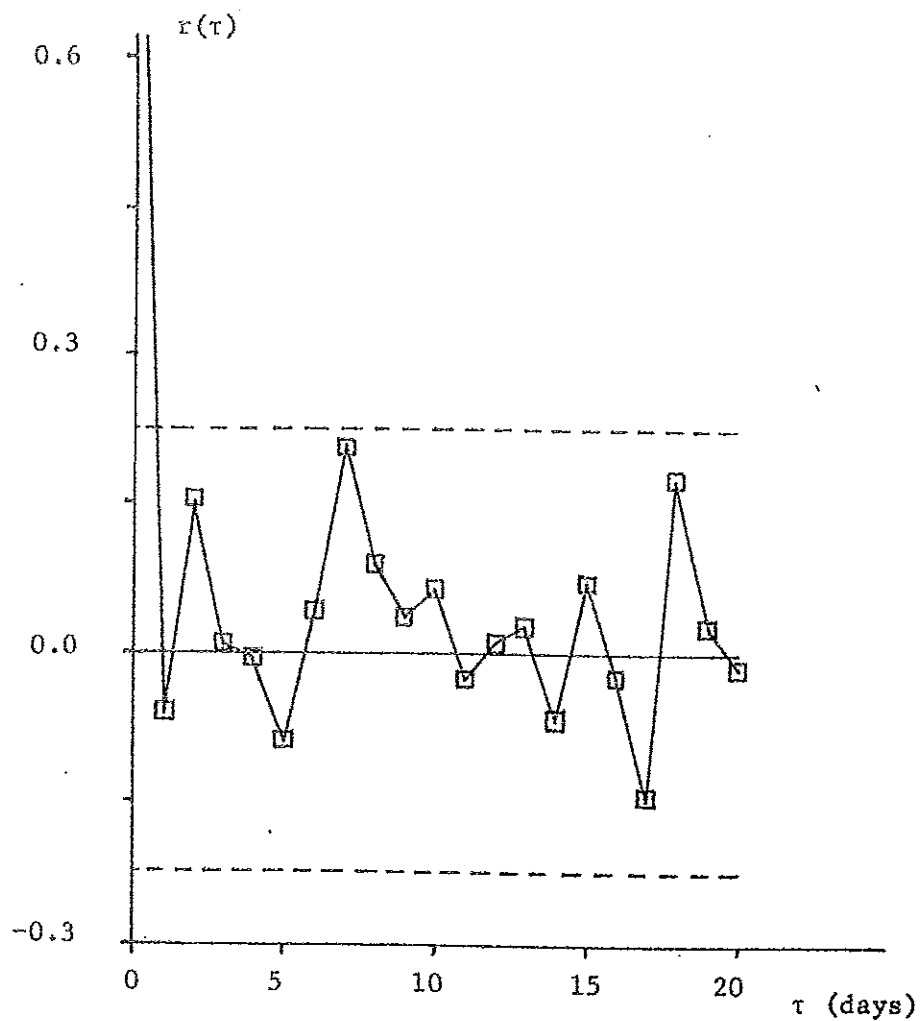


Figure 7 Sample autocorrelation function  $r(\tau)$  for ML DO model of table 5; dashed lines denote limits for 95% confidence that  $r(\tau) \approx 0, \tau \neq 0$ .

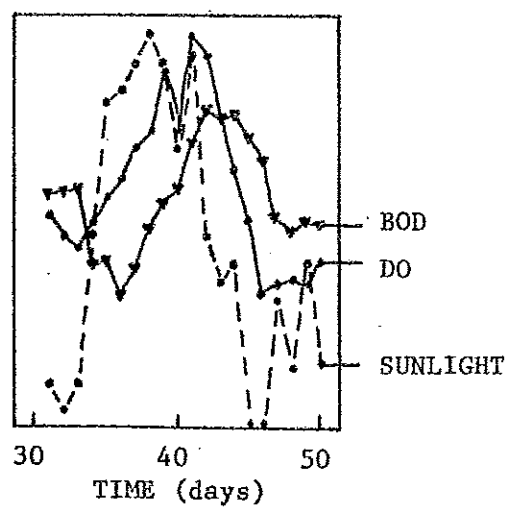


Figure 8 Observed data during the period of significant algal growth.

#### 4.3. Some implications of the results.

A certain degree of caution should be exercised in drawing physical interpretations from black box models but in the current study it is just such that leads to the final form of the explicit description for algal population dynamics. Expressed concisely in the form of eqn. (15) the foregoing results give

$$\left. \begin{aligned} \text{DO: } y_1(t_k) &= f_1\{y_1(t_{k-1}), u_1(t_{k-1}), u_3(t_k), \\ &\quad u_3(t_{k-1})\} \\ \text{BOD: } y_2(t_k) &= f_2\{y_2(t_{k-1}), u_2(t_{k-1}), u_3(t_{k-2}), \\ &\quad u_3(t_{k-4})\} \end{aligned} \right\} \quad (16)$$

and we can deduce that the dependence of DO and BOD on the sunlight input,  $u_3$ , is asynchronous and not simultaneous as implicitly assumed by model I. In other words, it appears that the production of DO in the reach at time  $t_k$  is a function of the incident sunlight during that day and the previous day ( $t_{k-1}$ ); in contrast, a BOD production results from longer delayed effects of approximately 2 to 4 days behind the current time. Certainly these interpretations agree with the observed dynamics of the field data shown in figure 8.

But apart from their relevance to algal population dynamics the models display satisfying properties in other respects. For instance, the order and magnitudes of the coefficients of the  $A(z^{-1})$  polynomials

$$\text{DO: } A_1(z^{-1}) = (1 - 0.715z^{-1});$$

$$\text{BOD: } A_2(z^{-1}) = (1 - 0.751z^{-1})$$

indicate that the estimated dynamics of the DO and BOD are essentially similar; we would expect this since the same term,  $(Q(t)/V)$ , is the dominant factor of the eigenvalues of eqns. I(i) and I(ii) in the state-space model. And, as an aside from the current discussion, a model for DO prediction on this basis (eqn. (16)) possesses desirable characteristics for control application since it is not dependent upon the inconvenience of measuring BOD, which is a five-day procedure by definition.

# 5. IDENTIFICATION RESULTS: (2) THE SYNTHESIS OF A MODIFIED STATE-SPACE MODEL.

The black box modelling results provide us with a good basis for the re-examination of a state-space model for DO-BOD interaction with algal disturbances. This section discusses an extension of the pseudo-empirical sustained sunlight effect of the a priori model, eqn. I, such that the asynchronous responses of DO and BOD to an algal population can be identified and their dynamics quantified.

If the maximum likelihood scheme outlined in section 3.1 is to be applied to a linear, continuous-time model, the discrete low-pass filter mechanism of the sustained sunlight effect, eqn. I(iii), should be represented in an equivalent continuous-time form, say

$$\frac{dI(t)}{dt} = \left[ -\frac{1}{\tau_1} \right] I(t) + \left[ \frac{g_1}{\tau_1} \right] u_3(t) \quad (17)$$

where  $\tau_1$  is a time constant (in days) and  $g_1$  is a gain coefficient. Additionally, from the observed dynamics of figure 8 it seems reasonable to suggest that while DO is related to  $I(t)$ , the BOD could be dependent upon the output  $I'(t)$  of a second filter "in series" with eqn. (21), i.e.

$$\frac{dI'(t)}{dt} = \left[ -\frac{1}{\tau_2} \right] I'(t) + \left[ \frac{1}{\tau_2} \right] I(t) \quad (18)$$

in which  $\tau_2$  is a time constant (in days). Thus, defining two further state variables

$$x_3(t) \triangleq I(t) ; \quad x_4(t) \triangleq I'(t)$$

it is possible to represent DO-BOD interaction as the following set of differential equations (from eqns. I(i), I(ii), (17) and (18)):

$$\begin{aligned}
 \dot{x}_1(t) &= - (K_1 + Q(t)/V)x_1(t) - K_2x_2(t) + \\
 &\quad + K_3(x_3(t) - \bar{I}_D) + (Q(t)/V)u_1(t) + \\
 &\quad + K_1C_s(t) + D_B \quad (i) \\
 \dot{x}_2(t) &= - (K_2 + Q(t)/V)x_2(t) + K_4(x_4(t) - \bar{I}_B) + \\
 &\quad + (Q(t)/V)u_2(t) + I_A \quad (ii) \\
 \dot{x}_3(t) &= - (1/\tau_1)x_3(t) + (g_1/\tau_1)u_3(t) \quad (iii) \\
 \dot{x}_4(t) &= - (1/\tau_2)x_4(t) + (1/\tau_2)x_3(t) \quad (iv)
 \end{aligned}
 \quad \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \text{II}$$

Here  $\bar{I}_D$  and  $\bar{I}_B$  are threshold levels\* referring to the individual effects of  $x_3$  on the DO, and of  $x_4$  on the BOD, respectively.

Thus finally, it is required that the parameter vector  $\theta$  be estimated in the model

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \theta_1 & 0 \\ 0 & \alpha_{22} & 0 & \theta_2 \\ 0 & 0 & -\theta_3 & 0 \\ 0 & 0 & \theta_4 & -\theta_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} \beta_{11} & 0 & 0 & \beta_{14} \\ 0 & \beta_{22} & 0 & \beta_{24} \\ 0 & 0 & \theta_5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \quad (19)$$

---

\* Note that no nonlinear constraint of the form eqn. I(iv) can be suitably accommodated in the structure of the linear model, eqn. II.

where

$$\begin{aligned}\alpha_{11} &= -(K_1 + Q(t)/V) & \beta_{11} &= \beta_{22} = Q(t)/V \\ \alpha_{21} &= -K_2 & \beta_{14} &= K_1 C_s(t) + D_B - K_3 \bar{I}_D \\ \alpha_{22} &= -(K_2 + Q(t)/V) & \beta_{24} &= L_A - K_4 \bar{I}_B\end{aligned}$$

and  $u_4(t) = 1.0$ , for all  $t$ , is a hypothetical signal such that the terms  $\beta_{14}$  and  $\beta_{24}$  can be included in the input matrix for computational convenience. The parameter vector  $\theta$  is defined by

$$\theta^T = [K_3, K_4, (1/\tau_1), (1/\tau_2), (g_1/\tau_1)]$$

and the values of table 1, with  $\bar{I}_D = \bar{I}_B = \bar{I}$  for simplicity, can be substituted for other elements of the matrices in eqn. (19); for the evaluation of  $\beta_{14}$ , we make the approximation,

$$C_s(t) = 10.0 \text{ for all } t$$

The values assumed for  $(K_3 \bar{I}_D)$  and  $(K_4 \bar{I}_B)$  are somewhat arbitrary and in any event the estimation of  $K_3$  and  $K_4$  will not be unique since it is dependent upon the choice of values for  $\bar{I}_D$  and  $\bar{I}_B$ . Nevertheless, the objective of the identification is primarily to estimate the dynamic components  $\tau_1$  and  $\tau_2$  and allow the gain coefficients  $g_1$ ,  $K_3$  and  $K_4$  to adjust such that the system responses are consistent with the observations

$$\begin{bmatrix} y_1(t_k) \\ y_2(t_k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t_k) \\ x_2(t_k) \\ x_3(t_k) \\ x_4(t_k) \end{bmatrix} + \begin{bmatrix} e_1(t_k) \\ e_2(t_k) \end{bmatrix} \quad (20)$$



The formulation of the estimation problem, eqns. (19) and (20), is partly a function of necessity and partly a matter of intention; it is not possible to estimate some of the  $\alpha$  and  $\beta$  parameters since they are time-varying, but, on the other hand, it is reasonable to assume that the DO and BOD dynamics are well known apart from their relationships with the algal effects. Hence, the definition of  $\theta$  is such that the efficiency of the algorithms should be directed towards the estimation of parameters associated with the dynamics of algal growth and decay.

The results of the maximum likelihood identification are shown in figure 9 with  $\theta$  estimated as

$$\hat{\theta}^T = [0.115, 0.146, 0.514, 0.704, 1.196]$$

From the estimates  $\hat{\theta}_3$  and  $\hat{\theta}_4$  the time constants of the two filters are given by  $\tau_1 = 1.95$  (days) and  $\tau_2 = 1.42$  (days); these values tend to confirm the inferences drawn from the black box modelling results. The variances of the residuals for the DO and BOD are 0.34 and 0.59, respectively, although the values are subject to the assumptions made in the quantification of the covariance matrices  $R_0, R_1, R_2^*$ .

---

\* Within the computational scheme it is possible to estimate elements of these matrices as additional parameters of the system; however, initial studies showed that this is not of advantage here.

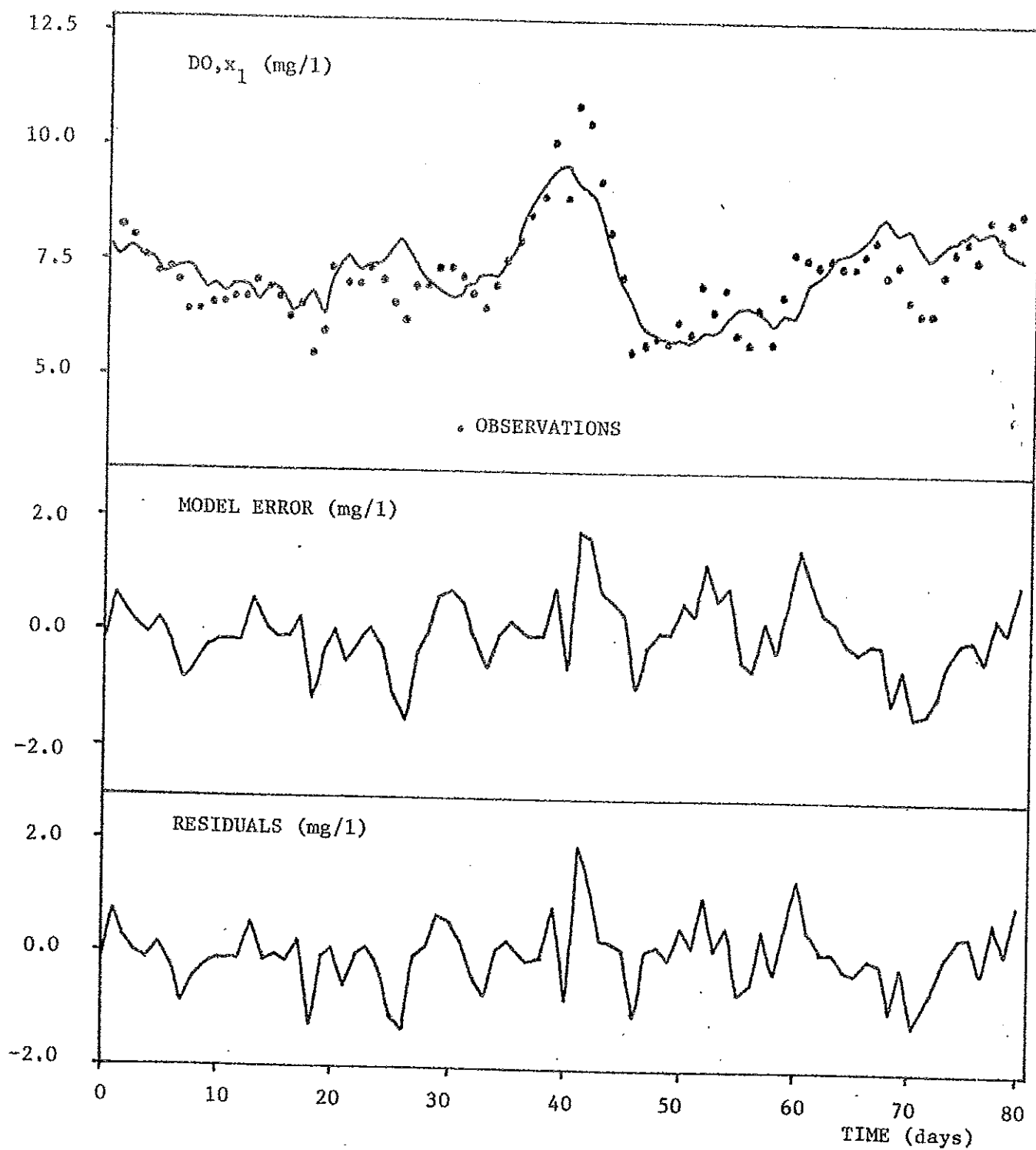


Figure 9.1 Maximum likelihood identification results for DO in model II.

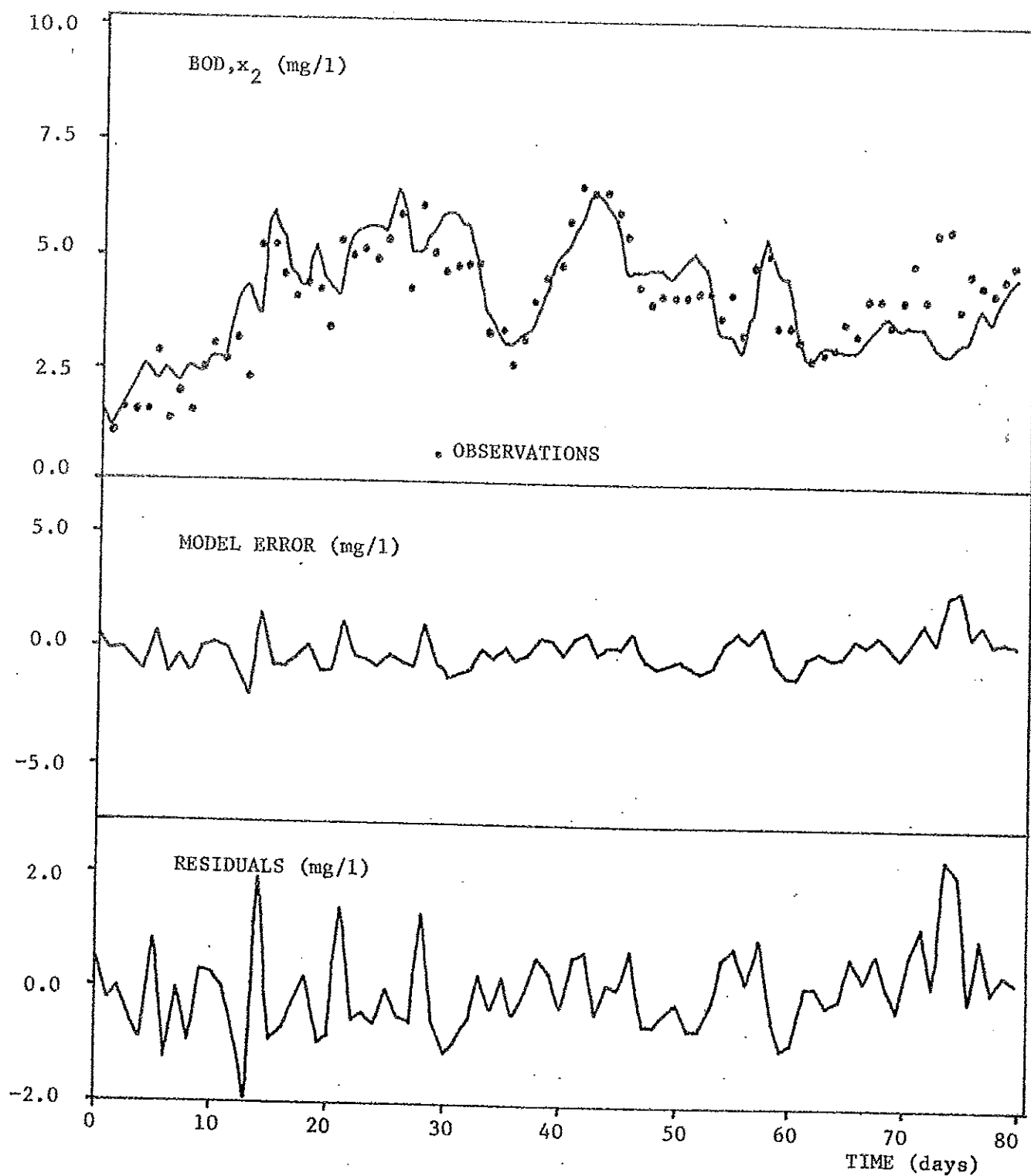


Figure 9.2 Maximum likelihood identification results for BOD in model II.

## 6. A MODEL FOR ALGAL POPULATION DYNAMICS.

The final step in the synthesis of a combined DO-BOD-algae model is the translation of the structure of eqns. II into a more meaningful physical terminology. Essentially, parts of the state-space model identified in the previous section still have a pseudo-empirical appearance; but what is important is that the dynamical features of the model agree well with the empirical observations of the real system. The task, therefore, is to retain the structure of the model and hypothesise variables and parameters which are plausible descriptions of the manner in which an algal population might interact with the DO and BOD concentrations of a reach of river.

Firstly, in general it can be stated that photosynthetic production of DO is dependent upon living algae, while dead algae in due course exert a BOD. Bearing this in mind, inspection of eqn. II shows that the pseudo-empirical variables  $x_3$  and  $x_4$  operate similarly in the description of BOD-DO interaction; indeed, we shall make the following redefinitions,

$x_3$  = concentration of living algae at the output of the reach (mg/l)

$x_4$  = concentration of dead algae at the output of the reach (mg/l)

Thus, it is required that a mathematical representation be postulated for the growth and decay of algae, together with terms accounting for photosynthetic/respiratory activity and subsequent production of BOD. Many of the ideas behind the model originate in parallel modelling and control problems in waste-water treatment (see e.g. Olsson et al [18]) and reference will be made to such material in support of the argument.

The growth of algae in a river can be likened to the continuous culture of autotrophic micro-organism species in the activated-sludge and anaerobic digestion processes (see e.g. Curds [11], Andrews [1]). In fact, the well-known Monod [17] development of the Michaelis-Menten equation for enzyme-substrate interaction has already been used for the study of algal growth kinetics [10], [20]. However, for the purposes of the present model the Monod function is employed in a slightly different manner so that, together with a specific decay rate for the algae [1], [19], the following is proposed:

Net production rate of living algae,  $x_3$  (in (mg/l)/day)

$$= \hat{\mu} \left\{ \frac{u_3(t)}{K_s + u_3(t)} \right\} - K_D x_3(t) \quad (21)$$

where  $\hat{\mu}$  is the maximum specific growth rate of algae ((mg/l)/day),  $K_s$  is a saturation coefficient for the limiting nutrient  $u_3$  (hrs. sunlight/day) and  $K_D$  is the specific decay rate of algae (days<sup>-1</sup>). The important features of the above expression are that sunlight is taken to be the rate-limiting "nutrient" or substrate\* and, since the growth rate is independent of  $x_3$ , it is implied that algae are not autotrophic organisms. It is not intended that biological theory be contradicted, but for a mathematical description the following two mitigating conditions should be considered. Firstly, preliminary simulation showed that if the specific growth rate is assumed to be dependent on  $x_3$  the dynamics of an algal population do not correspond at all with their observed

---

\* The field data would suggest that this is so, and, in any case, no observations are available for other possible rate-limiting nutrients such as nitrate, phosphate, carbon dioxide etc.

interaction with the BOD and DO. Secondly, in view of the fact that the river is a flowing system it seems unreasonable to propose that the output concentration of algae is a factor governing the growth rate of algae in the reach. Suffice it to say that with the convenient lumped-parameter structure [6] of the overall model it is not possible to transfer directly all the characteristics of micro-organism culture in waste-water treatment to the prediction of algal growth in a river.

In some studies of activated-sludge and anaerobic digestion units the significance of dead organisms is assumed to be negligible. Nevertheless, for the algal population under consideration here the omission of a separate dynamic description of dead algae seems to be inadmissible if the observed BOD variations are to be simulated correctly. Thus, making deductions from the parallel investigations of Westberg and co-workers [13], [21] the following is postulated:

Net production rate of dead algae,  $x_4$  (in (mg/l)/day)

$$= K_D x_3(t) - K_R x_4(t) - R_S \quad (22)$$

in which  $K_R$  is a rate constant for the redissolution of dead algae ( $\text{day}^{-1}$ ), and  $R_S$  is the rate of sedimentation of undissolved dead algal material ((mg/l)/day). In fact Westberg proposes more complex expressions for the decay rate of organisms in (21) and (22) and for the redissolution of dead material in (22), which strictly speaking may not be applicable here. The reasons for the inclusion of  $R_S$  are perhaps tenuous, but the model fitted the data better in this form and, furthermore, such a process could be of significance in the real system [10], [20].

Having defined eqns. (21) and (22) we are in a position

to make component mass balances across the reach of river (see figure 1) in order to derive the differential equations for  $x_3$  and  $x_4$ . Adjoining these equations to those for the DO and BOD and proposing new terms for the interaction of  $x_3$  and  $x_4$  with the state variables  $x_1$  and  $x_2$ , we obtain

$$\left. \begin{aligned} \dot{x}_1 &= - (K_1 + Q(t)/V)x_1 - K_2x_2 + \gamma_1x_3(u_3)^s - \\ &\quad - \gamma_2x_3 + (Q(t)/V)u_1 + K_1C_S(t) + D_B \quad (i) \\ \dot{x}_2 &= - (K_2 + Q(t)/V)x_2 + \gamma_3K_Rx_4 + \\ &\quad + (Q(t)/V)u_2 + L_A \quad (ii) \\ \dot{x}_3 &= - (K_D + Q(t)/V)x_3 + \hat{p} \left\{ \frac{u_3(t-T_d)}{(K_S + u_3(t-T_d))} \right\} \quad (iii) \\ \dot{x}_4 &= - (K_R + Q(t)/V)x_4 + \gamma_4K_Dx_3 - R_S \quad (iv) \end{aligned} \right\} \quad \text{III}$$

where lower case  $s$  is the exponential power for the dependence of algal photosynthetic production of oxygen on sunlight conditions,  $T_d = (t_k - t_{k-1})$  is a pure time delay of one day (i.e. the sampling interval of the data)\*, and  $\gamma_i$ ,  $i = 1, 2, 3, 4$ , are proportionality constants such that the model can be fitted to the data. All other variables are as previously defined and the argument  $t$  has been omitted from the inputs  $u$  and the state variables  $x$  for notational convenience;

---

\* This small data manipulation allows more accurate prediction of the DO conditions over the observational period  $t_{35} \rightarrow t_{46}$ ; indeed, such delayed effects could be due to the presence of a stored mass of algal population.

no living or dead algae are assumed to be present in the material influx at the upstream boundary of the experimental reach of river.

Remarkably enough, the model of eqn. III contains all the features of the system which have been hitherto identified:

- i) Eqn. III(i) shows that the photosynthetic production of oxygen at time  $t$  is dependent upon the living algae and the sunlight conditions at time  $t$  (c.f. section 4.3); also respiration will occur at a rate proportional to the concentration of living algae.
- ii) Eqn. III(ii) expresses the concept that a BOD is exerted by dead algal material which has redissolved; note that our implicit definition of BOD refers to soluble substances and is therefore not related to particulate matter which has been sedimented (e.g.  $R_g$  in eqn. III (iv)).
- iii) And, finally, if  $Q(t)/V$  is taken to be the dominant factor in the dynamics of  $x_3$  and  $x_4$ , it turns out that the associated time constants when algal effects are significant are approximately 1.5 (days) (c.f.  $\tau_1$  and  $\tau_2$  in section 5); thus, it has been possible to retain the important dynamical structure of the pseudo-empirical model given by eqn. II.

Unfortunately, with such a nonlinear model it is only possible to make verification by comparison of the deterministic simulation responses of  $x_1$  and  $x_2$  with the DO and BOD field data. The eventual quantification of the many additional parameter values in table 6 is not unique, but the model responses shown in figure 10 are very satisfactory. All other relevant parameter values are as given in table 1 with  $C_g(t)$  generated from eqn. (1) and  $D_B$  as described previously.



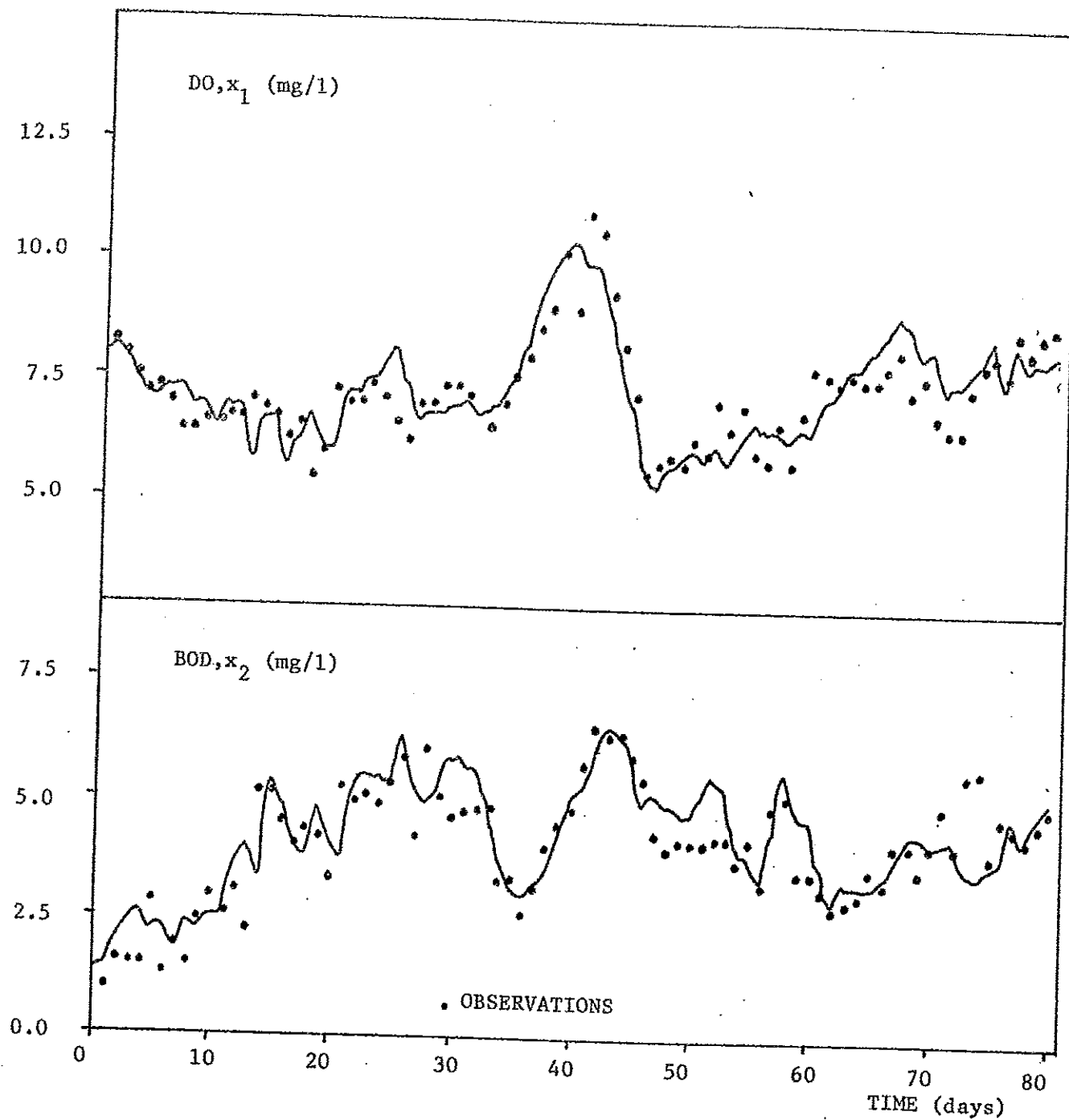


Figure 10 Simulation responses for model III.

Table 6 - Estimated parameter values for eqn. III.

Parameter	Value	Parameter	Value
$K_D$	0.35	$s$	0.55
$K_R$	0.25	$Y_1$	1.45
$\hat{p}$	2.1	$Y_2$	2.0
$K_S$	20.0	$Y_3$	16.0
$R_S$	-0.11	$Y_4$	3.0

Perhaps the significant feature of the final model is the more adequate prediction it gives over  $t_{60} \rightarrow t_{80}$ , whose effects the a priori model, eqn. I, does not describe at all well; this suggests that the disturbances caused by algae are particularly sensitive to the low flow-rate through the reach of river during this period.

Figure 11 shows the comparison between the sustained sunlight effect  $\bar{I}(t_k)$  of model I and the simulated live and dead population concentrations of algae,  $x_3$  and  $x_4$ , respectively, given by model III. Both  $x_3$  and  $x_4$  indicate the presence of a larger population of algae than would be predicted by the sustained sunlight effect over the latter period of the experiment.

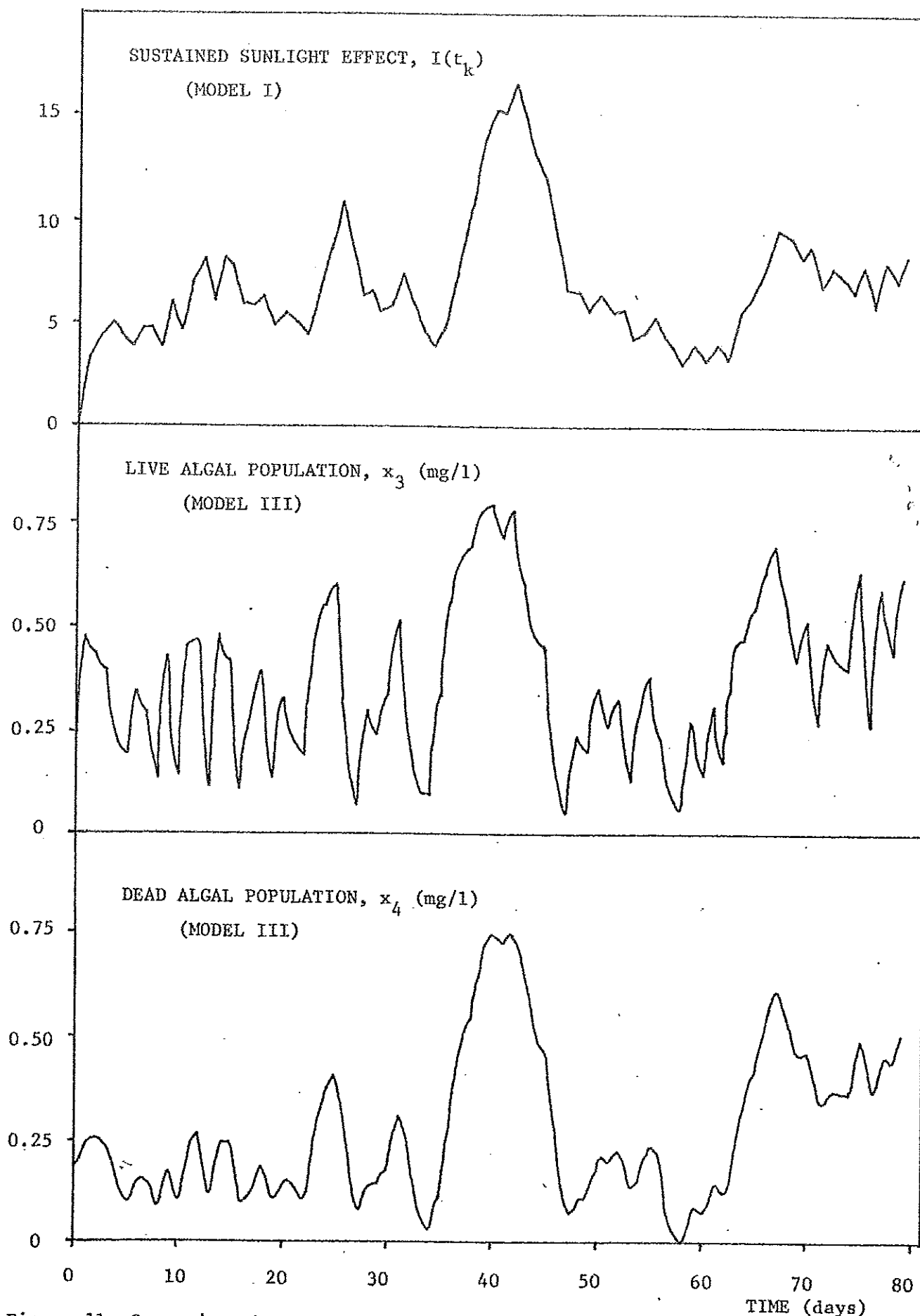


Figure 11 Comparison between the sustained sunlight effect, model I, and the predicted algal population magnitudes of model III.

## 7. CONCLUSIONS.

A dynamic model for the characterisation of DO-BOD-algae interaction in a fresh-water stream has been synthesised and verified against field data on the DO and BOD. The structure of the model depends essentially on the separation of algal dynamics into the consideration of both a live and a dead population; it is evident that the effects of the algal population are sensitive to the flow-rate and residence-time conditions in the reach.

The biochemical and biological principles governing the growth and decay of micro-organisms are complex and the proposals of this paper are doubtless a considerable simplification of the true dynamics. The model is developed from a combination of heuristic reasoning, a cross-breeding of ideas from studies on the modelling of waste-water treatment processes, and systematic identification of empirical observations. Indeed, the value of real data and the treatment of such with several different methods of identification cannot be emphasised too strongly. But the results remain tentative in the sense that we have been forced to hypothesise a model for living and dead algae in the absence of any observations relating directly to such quantities. In particular, the manner in which decomposing algal material exerts a BOD remains unclear and it is quite possible that the BOD test itself could obscure the identification of such a process.

Naturally the model presented here could be criticised as a control engineer's view of ecology and biology; nevertheless, it seems to be a step in the right direction. If nothing else it may stimulate a contraction of the lamentable gap in communication between biologists and control engineers.

## 8. REFERENCES

- [1] Andrews, J.F., "Kinetics of biological processes used for waste-water treatment", Paper presented at a Workshop on Research Problems in Air and Water Pollution, University of Colorado, Boulder, Colorado, Aug. 1970.
- [2] Åström, K.J., "Introduction to stochastic control theory", (book), Academic Press, New York, 1970.
- [3] Åström, K.J., and Bohlin, T., "Numerical identification of linear dynamic systems from normal operating records", Proc. IFAC Symposium on the Theory of Self-adaptive Control Systems, Teddington, England, 1965.
- [4] Åström, K.J., Bohlin, T., and Wensmark, S., "Automatic construction of linear stochastic dynamic models for stationary industrial processes with random disturbances using operating records", Technical paper TP 18.150, IBM Nordic Laboratory, Sweden, 1965.
- [5] Åström, K.J., and Källström, C.G., "Application of system identification techniques to the determination of ship dynamics", Paper presented at the IFAC Symposium on Identification and System Parameter Estimation, The Hague, The Netherlands, June 1973.
- [6] Beck, M.B., Ph.D. thesis, Engineering Department, University of Cambridge, December 1973.
- [7] Beck, M.B., "Water quality maintenance and control theory", Report No. 7402(C), Lund Institute of Technology, Division of Automatic Control, Sweden, February, 1974.
- [8] Beck, M.B., and Young, P.C., "A dynamic model for DO-BOD relationships in a non-tidal stream", submitted to Water Research for publication.

- [9] Beck, M.B., and Young, P.C., "Identification and parameter estimation of BOD-DO models using the extended Kalman filter", in preparation.
- [10] Chen, C.W., "Concepts and utilities of ecologic model", Proc. A.S.C.E., J. Sanit. Eng. Div., Vol. 96, No. SA5, October 1970.
- [11] Curds, C.R., "A theoretical study of factors influencing the microbial population dynamics of the activated-sludge process - I", Water Research, Vol. 7, pp. 1269-1284, 1973.
- [12] Eklund, K., and Gustavsson, I., "Identification of drum boiler dynamics", Paper presented at the IFAC Symposium on Identification and System Parameter Estimation, The Hague, The Netherlands, June 1973.
- [13] Ericsson, B., Gustafsson, B., and Westberg, N., "En studie av aktivslamprocessens kinetik - III" ("A study of the kinetics of the activated-sludge process - III"), in Swedish, Vatten, pp. 146-153, 1969.
- [14] Gustavsson, I., "Parametric identification of multiple input, single output linear dynamic systems", Report No. 6907, Division of Automatic Control, Lund Institute of Technology, Sweden, 1969.
- [15] Gustavsson, I., Selander, S., Wieslander, J., "IDPAC - User's guide", Report No. 7331, Division of Automatic Control, Lund Institute of Technology, Sweden, 1973.
- [16] Jazwinski, A.H., "Stochastic processes and filtering theory", (book), Academic Press, New York, 1970.
- [17] Monod, J., "Recherces sur la croissance des cultures bacteriennes", Hermann et Cie, Paris, 1942.

- [18] Olsson, G., Dahlgvist, K.I., Eklund, K., Ulmgren, L., "Control problems in wastewater treatment", Technical report, The Axel Johnson Institute for Industrial Research, Nynäshamn, Sweden, 1973.
- [19] Ott, C.R., and Bogan, R.H., "Theoretical analysis of activated-sludge dynamics", Proc. A.S.C.E., J. Sanit. Eng. Div., Vol. 97, No. SA1, February, 1971.
- [20] Rich, L.G., Andrews, J.F., and Keinath, T.M., "Diurnal pH patterns as predictors of carbon limitation in algal growth", Water and Sewage Works, pp. 126-130, May 1972.
- [21] Westberg, N., "A study of the activated-sludge process as a bacterial growth process", Water Research, Vol. 1, pp. 795-804, 1967.
- [22] Whitehead, P.G., and Young, P.C., "The Bedford Ouse study dynamic model; 2nd semi-annual report to the steering group of the Great Ouse Associated Committee", Control Note CN/74/1, Control Engineering Group, Engineering Department, Cambridge University, January, 1974.
- [23] Whitehead, P.G., and Young, P.C., "A dynamic stochastic model for water quality in part of the Bedford-Ouse river system", Proc. IFIP Working Conference on Modelling and Simulation of Water Resources Systems, Ghent, July/August 1974.
- [24] Young, P.C., "A recursive approach to time-series analysis", Bulletin of the Institute of Mathematics and its Applications (IMA), Vol. 10, Nos. 5/6, pp. 209-224, 1974.
- [25] Young, P.C., and Beck, M.B., "The modelling and control of water quality in a river system", Automatica (in press, to appear September 1974).

- [26] Young, P.C., and Whitehead, P.G., "A recursive approach to time-series analysis for multivariable systems", Proc. IFIP Working Conference on Modelling and Simulation of Water Resources Systems, Ghent, July/August 1974.

#### ACKNOWLEDGEMENTS.

I am grateful to Professor Karl Johan Åström and his department for the assistance given to me during the course of this study. In particular, I should like to thank Gustaf Olsson, Ivar Gustavsson and Claes Källström for their discussion and advice.