A COMPARATIVE CASE STUDY OF DYNAMIC MODELS FOR DO-BOD ALGAE INTERACTION IN A FRESHWATER RIVER

B. Beck

May 1978 WP-78-16

2361
Laxenburg
Austria

International Institute for Applied Systems Analysis

Preface

Over the next few years various Tasks of the Resources & Environment (REN) Area at IIASA will concentrate on selected problems of water resources management, ecology, and environmental quality control. Aspects of water resources management have been studied at IIASA since its inception, although only recently has the scope of these studies been extended to include more detailed analysis of the *quality* of water resources. One of the objectives of the current Task 2 of Resources and Environment, "Models for Environmental Quality Control and Management", is the *development* and *application* of models for analyzing the impact of waste discharges on the hydrophysical and ecological processes taking place in aquatic environments.

This paper, one of the first to report on the activities of Task 2 (REN), is concerned with the subject of *river water quality model development.* The paper summarizes and compares earlier extensive analyses of experimental time-series field data from a lowland river in England. In this latter sense the paper stands between publications originating from other past and prospective IIASA studies: the forthcoming McGraw-Hill publication "Modelling and Control of River Quality" discusses in detail some of the results presented here--the book is a product of a project jointly supported by the Centro Teoria dei Sistemi CNR, Milan, Italy, and IIASA; secondly, the summarizing nature of the paper overlaps with Task 2's objectives for the preparation and publication of a *survey of water quality modelling.*

Some of the reasons for Task 2's state-of-the-art survey include the desire to clarify the capabilities of water quality models and to accelerate the transfer of existing modelling technologies. It is not the intention of this paper to assist in the transfer of a packaged software for water quality models, even though a number of computational notes are included and, in principle, the models are ready for management applications. Rather, we hope that this paper will facilitate the transfer of field data for the evaluation of water quality models. And we hope that the field data will prove to be educational in the development of software and algorithms for identification and parameter estimation, since these are some of the basic tools of systems analysis in model-building.

The subject of model applications in the context of operational river basin management will be discussed in a later publication.

Summary

From recent IIASA workshops on water quality modelling a need can be identified for comparative studies of different model types against the same set of field data. Similarly, some of the motivation for a state-of-the-art survey on water quality modelling to be prepared under the auspices of IIASA stems from the desire to bring order and authenticity to a fast developing field of technology. The problem is as follows: although models can be readily applied in management and decision making, they are not always so readily subject to a prior verification against field data from the river system. One reason underlying this problem is that the relevant field data, with a sufficiently high sampling frequency and collected over a sufficiently long period, either do not exist or have not been publicized.

The primary objective of this paper is the dissemination of a set of time-series field data suitable for the identification and verification of *dynamic* models for water quality. Here water quality is interpreted as the interaction between three variables, dissolved oxygen (DO) concentration--a broad measure of the healthy state, or otherwise, of a river--biochemical oxygen demand (BOD) concentration--a macro-measure of typical municipal/domestic organic waste materials--and a population of algae. A secondary objective is the comparison, by means of response error statistics, of several models which have been derived by reference to the field data. And yet a third objective is to present a summarizing and concluding statement on a river water quality model development exercise which spans various publications over the past four or five years.

With respect to model comparison and model assessment the paper concludes with a cautionary message on the use of simple fitting error statistics; and, in any case, it is argued that judgements about the "best" model are dependent upon the intended application of the model. On the accuracy of the models as representations of the real system it is found that many questions remain unresolved, and particularly so for those aspects of the models related to the growth kinetics and death, decay properties of floating algal populations. The hope is expressed that, given the data, others will be stimulated not only to answer these questions but also to reassess the assumptions that the paper makes concerning the mixing and transport characteristics of the case study reach of river.

Abstract

A comprehensive set of field data is presented for purposes of identifying and verifying dynamic models of DO-BOD-algae interaction in a freshwater river. Several models derived on the basis of these field data are reported and their fitting error statistics are compared. A number of grounds for criticism of the models are discussed and, in particular, it is suggested that further analysis should be undertaken along the lines of more conventional advection-diffusion representations of a river's transport and dispersive properties. A summary of directly supporting studies on system identification, parameter estimation, model interpretation, and model application in operational control contexts is given, principally in the form of an appendix of abbreviated notes.

A Comparative Case Study of Dynamic Models for DO-BOD-Algae Interaction in a Freshwater River

M.B. Beck

1. Introduction

Over the past ten years many models for river water quality have been developed. A substantial proportion of this modelling effort has been concerned with understanding and quantifying the relationships between stream dissolved oxygen (DO) concentration and biochemcial oxygen demand (BOD) concentration. Recently these DO-BOD models have been extended to embrace more detailed relationships between various ecological constituents which characterize the quality of a water body, thus providing the potential for more sophisticated assessments of the impact of waste discharges on an aquatic environment. Not all of the models proposed so far, however, have had the benefit of being verified against field data; and any decision-maker or manager requiring the application of a water quality model might justifiably be sceptical and confused at the variety of models available to him.

The purpose of this paper is to offer a vehicle for such model verification and model comparison studies through the publication of a suitable set of field data. A secondary aim of the paper is to catalogue a number of models which have been derived by reference to this field data set and hence to conclude a project which has now extended quite beyond its original expectations--to a period of some five years. From the models presented it will be evident that there is a significant gap in the analysis. No form of partial differential equation, advection-diffusion model has been tested with the data, and it is hoped that others will be encouraged to complete this section of the analysis. Since this latter class of models is quite general in nature it would seem to be a straightforward matter to deduce the conditions necessary for their application to the reach of river in question.

The format of the paper is as follows. In section 2 a brief description of the characteristics of the data and river system are given. The data comprise a set of time-series for daily sampled values of dissolved oxygen concentration, BOD concentration, discharge, temperature, and sunlight conditions; they refer to a short stretch (4.5 km) of the River Cam in England for the summer period of 1972. Section 3 complements section 2 by defining the nature and notation of the classes of models to be analyzed and by formalizing a simple statistical criterion for model comparison. At this point the assumptions underlying the derivation of *ordinary* differential equation forms for dynamic models of DO-BOD interaction are restated (see also Beck and Young (1975)). These assumptions are an important distinguishing feature of the modelling approach which has been adopted. Broadly speaking there are two classes of model of interest, namely internally descriptive (mechanistic) models, examples of which are given in section 4, and black box (input/output) models, examples of which are given in section 5. A discussion of model structure identification and parameter estimation for each individual model, however, is not included; nor is there any discussion presented on the subject of model application, e.g. in operational control situations, and interpretations on the significance and forms of the models are kept to a minimum. Section 6 of the paper attempts to summarize some potentially controversial issues connected with the data and modelling studies and certain open questions: questions on the method of model assessment and on the biochemical/ecological accuracy of the models. Although "fitting error" statistics are defined and used throughout the paper no conclusion is made about which is the "best" model, since this kind of judgement depends strongly upon the objectives for the intended application of the model.

The field data are listed in Appendix 1. Other Appendixes contain data on the geometry of the river, additional estimation results and statistics, a description of the simulation of a time-variable transporation delay function, and an abbreviated directory of previously published material supporting, interpreting, and applying the results of the main body of the text.

2. Introductory Description of the River System and Field Data

The River Cam, a tributary of the Great Ouse River, flows approximately south-west to north-east across eastern England, see Figure 1. The upper reaches of its catchment area are predominantly chalky and by the time the Cam passes through Cambridge it is already a slowly moving lowland river. Upstream of Cambridge the river carries a light loading of treated industrial (pharmaceutical, fertilizer production) and municipal effluent but is still considered suitable for bathing and recreational purposes. Just downstream of Cambridge the city discharges its sewage to the river and for some distance thereafter the stream water quality is substantially degraded. The sewage receives both primary and secondary treatment prior to discharge.

Figure 2 shows the precise definition and location of the experimental reach of river with respect to the sewage outfall. Attached weed and plant growth in this section of river is significant, although the growth is frequently cropped during the summer for reasons of the considerable use made of the river by pleasure craft. The whole of the Cam's subcatchment is an intensively agricultural area. The land adjacent to the stretch of river in Figure 2 can be classified as fenland and is drained by a system of dykes whose water is from time to time pumped up into the river. One such dyke is situated about 30 m downstream of the lower weir in Figure 2. From the physical character of the system, therefore, significant local surface runoff or seepage into the river is unlikely; in additon no major tributary joins the river between the two weirs of Figure 2.

We may note that with respect to obtaining measurements which give a reasonably clear picture of DO-BOD interaction dynamics, the defined system has several advantages:

i) The input of sewage works effluent ensures that the system is suitably "excited" (i.e. variations in DO and BOD conditions can be observed which are not attributable to either measurement error or chance disturbance of the system).

ii) The critical conditions of DO sag often occur in reaches of river immediately downstream of effluent outfalls and, in this particular river, fish kills have been reported during periods of low DO levels.

iii) The weir below the effluent discharge point aids the assumption of complete mixing of the effluent with the stream as it enters the defined system.

iv) The short reach between the upper weir and upper system boundary is a precaution against obscuring the measurements of DO by entrained bubbles and other localized fluctuations resulting from the action of the weir.

The complete set of field data (see Appendix 1) consists of 81 daily sampled values for each variable; this covers the period from 6th June until

25th August, 1972. The methods of measurement used for each variable are summarized in Table 1. Here it should be noted that the upstream DO and temperature measurements were obtained from a battery-operated portable/ submersible monitor (loaned from the Water Research Centre, Stevenage), whereas the downstream temperature and DO recordings were recovered from a permanent monitoring station belonging to the Anglian Water Authority.

Table 1: Summary of Data Specifications

Variable	Location--with respect to Figure 2	Sampling Rate	Measurement Technique
DO	U,D	Continuous	Monitor
(5-day) BOD	U,D	Once per day	Single grab sample
Temperature	U,D	Continuous	Monitor
Discharge	D	Once per day	Level-discharge relationship at weir
Hours of sunlight	*	Once per day	-
Rainfall	*	Once per day	-

*Meteorological measurements were taken from a location some 8 km distant from the experimental stretch of river.

For the data of Appendix 1 sampled values at 12.00 hrs. each day were read from the strip-chart records from both types of monitor. The downstream monitor withdrew its sample from a median point in the river cross-section; the upstream monitor sampled at a point 4 m from one bank and at a depth of 1 m. The BOD measurements were taken at times varying between 09.00 and 15.00 hrs. on any given day with the sample being drawn from the centre of the respective river cross-section at a depth of approximately 0.5 m. In Appendix 1 certain simplifications have, therefore been made: (i) the sampling times for the BOD measurements are averaged at 12.00 hrs. for each day; (ii) since no significant difference could be detected in the upstream and downstream temperature measurements only the downstream record is quoted for use in the modelling exercise. Should the reader so wish, precise sampling times for the BOD measurements and three-hourly sampled values of DO and temperature, together with daily flow-rate and (five-day) BOD measurements for the effluent discharge, are available from the author for more detailed simulation purposes. However, such information is not essential to the present discussion. In Appendix 2 a set of cross-sectional area measurements are given for regularly spaced intervals along the case study reach of the river.

3. Preliminaries: Models and a Method of Model Assessment

We shall distinguish between two classes of model. The first, denoted by the term *internally descriptive model*, is a description of the system's dynamic behaviour which emodies substantial a priori knowledge of the physical, chemical, biological, and ecological phenomena governing the relationships between input, state, and output variables. The other type of model, the *black box model*, requires no such a priori information, makes no such claim to describe the internal mechanisms of the system, and is simply an empirically, or statistically, defined relationship between the observed input and output behaviour.

3.1 Internally Descriptive Model Definition

Figure 3(a) gives a schematic definition of the reach of river and some notational conventions for the measured variables. Figure 3(b) shows the transportation delay/continuously-stirred tank reactor (CSTR) *idealization* of the reach of river which permits the subsequent mathematical description of system behaviour in terms of lumped-parameter, ordinary differential equation forms. Clearly this idealization draws upon standard elements of chemical engineering reactor analysis, e.g. Himmelblau and Bischoff (1968), Buffham and Gibilaro (1970); the idealization can be shown to approximate both experimentally observed transport and dispersion mechanisms [Whitehead and Young (1974)] and the analytical properties of distributed-parameter, partial differential equation representations of advection-diffusion mass transport [Rinaldi et al (1978)].

The reasons for the tranformation of the process model from a description with time and space as the independent variables, which is intuitively more natural, to a description with just time as the single independent variable are threefold:

i) The transformation simplifies subsequent computation and analysis, since, in principle, ordinary differential equations are more easily solved than partial differential equations;

ii) Statistical procedures for model structure identification, parameter estimation, and model verification are in practice largely restricted to lumped-parameter representations--the corresponding treatment of distributed-parameter systems is considerably less well established or understood;

iii) With a view to the (originally) intended application of the model for operational control purposes [Young and Beck (1974)], the vast majority of control system synthesis methods are devoted to process dynamic characterizations in terms of time as the single independent variable.

As we shall see, even with such a potentially simplifying transformation the simulation of the transportation delay element of the idealization in Figure 3(b) presents certain difficulties.

A set of component mass balances across the two elements of Figure 3(b) yields thus the following form of continuous-time, internally descriptive model:

For the CSTR -

$$\dot{\underline{x}}(t) = (\Theta_1(t)/a_1)\underline{u}'(t) - (\Theta_1(t)/a_1)\underline{x}(t) + \underline{\mathcal{S}}(t) + \underline{\xi}(t) \quad . \tag{1}$$

For the transportation delay -

$$\underline{u}'(t) = \underline{u}(t - \tau(t)) \quad . \tag{2}$$

The general notation of equations (1) and (2) is defined in Table 2; in specific terms,

$\Theta_1(t)$ = stream discharge ($m^3 day^{-1}$)

a_1 = constant volumetric hold-up of water in defined reach of river (m^3)*

$\tau(t)$ = "length" of transportation delay element (day)

t = independent variable of time (day).

---

* In the following the omission of the argument t from any parameter definition indicates the assumption that the parameter is time-invariant.

There are three basic assumptions required to combine equations (1) and (2) in order to give the form of the internally descriptive model which is discussed subsequently:

Assumption (1): that the volumetric hold-up of water in the reach of river, a_1, is constant.

Assumption (2): that there is no interaction between variables in the transportation delay element of the process idealization.

Assumption (3): that equation (2) can be approximated *for this particular case study* by

$$\underline{u}'(t) = \underline{u}(t) \quad . \tag{3}$$

Table 2: Summary of General Notation and Variable Definitions

Variable	Definition
General	
$\underline{u}$	vector of measured input variables
$\underline{x}$	vector of state variables, or hypothetical noise-free output variables
$\underline{y}$	vector of measured output variables
Internally Descriptive Model	
$\underline{a}$	vector of parameters (coefficients)
$\underline{\theta}$	vector of variables "internal" to the model but not defined as state variables
$\underline{S}$	vector of source and sink terms related to each state variable
$\underline{\xi}$	vector of chance, random disturbances of the system
$\underline{\eta}$	vector of output measurement errors
Black Box Model	
$\underline{\alpha}$	vector of autoregressive polynomial parameters
$\underline{\beta}$	vector of input polynomial parameters
v	lumped noise process accounting for both random disturbances and measurement errors

Table 2 Cont'd.

Variable	Definition
<u>Model Assessment</u>	
$\hat{\underline{x}}$	vector of (deterministic) model output predictions
$\underline{\varepsilon}$	vector of errors between observed output and deterministic model output predictions

Assumption (1) has already been incorporated into the statement of the component mass balances of the CSTR, equation (1); and Assumption (2) is implicit in equation (2) in the sense that any physical, chemical, biochemical reactions are assumed to take place only in the CSTR.

Assumption (3) is both crucial and much more difficult to justify. The description of $\tau(t)$ merits some thought since this description needs to be time-varying according to variations in the stream discharge $\Theta_1(t)$. Methods for simulating such a time-variable transportation delay are available, see e.g. Coggan and Noton (1970) and Appendix 4, and would almost certainly be required for longer reaches of river and for time-series data in which the sampling interval is much shorter than the average detention time of the reach. On inspection Appendix 4 suggests that to include this kind of simulation for $\tau(t)$ is merely to exchange the complexity/computational effort of a distributed-parameter model for the complexity and effort of solving an increased number of ordinary differential equations. [In fact, partly for this reason transportation delays (or "dead time") are extremely awkward to handle in continuous-time control system design procedures; they are much more easily accommodated in the framework of discrete-time, or digital, control system synthesis techniques.] It should now be evident, therefore, why Assumption (3) is important in that it permits a considerable simplification. Yet at the same time some assessment should be made of the degree of inaccuracy introduced by the asumption.

Firstly, for the short study reach of the Cam with an average detention time during the experimental period of just over one day, and given the relatively slow sampling frequency (once per day), it is not possible to observe, and hence to identify or model, the response of DO-BOD interaction to *higher* frequency, input, upstream disturbances. Moreover, as Rinaldi et al (1978) point out, an idealization of the river reach as a CSTR *without* any transportation delay element provides in theory a better approximation to the advection-diffusion representation in the regime of *low* frequency disturbances. We would thus expect the models employed here to provide very poor approximations to the downstream DO and BOD concentrations as responses to impulsive (high frequency) changes in the upstream DO and BOD concentrations. This the models do, for they predict an instantaneous downstream response to variations upstream. On the other hand, with Assumption (3) the models should simulate quite well the advective transport of material downstream when conditions at the upstream boundary are changing in the manner of longer-term trends and slow periodic fluctuations, i.e. low frequency input disturbances.

Secondly, the following qualifications apply to the above kinds of argument:

i) that for the integration of equation (1) over the time interval of one day $\underline{u}(t)$ is substituted by the values measured at the *beginning* of that period (see section 4.1)--hence, the predicted downstream concentrations at 12.00 hrs. on the current day are a function of the measured upstream concentrations at 12.00 hrs. on the previous day (compare with the average detention time properties of the study reach); and

ii) that some, if not a large proportion, of the high frequency disturbances and variations in the observed process dynamics are due to stochastic effects which thus represent a kind of irreducible minimum error that can be obtained in the following modelling exercise.

Thirdly, in order to avoid confusion, let us mention that the term "transportation delay" as defined and used here is *not* equivalent to the term "time of travel". For instance, whereas the time of travel might represent the time taken to reach the peak (or centre of gravity) of the downstream response to an upstream impulse tracer disturbance, the transportation delay more closely resembles the time elapsed before any positive response to the impulse input is detected downstream. If an average value for the time of travel can be approximated by the ratio $(a_1/\Theta_1(t))$, then in general

$$\tau(t) < (a_1/\Theta_1(t)) \quad .$$

Rinaldi et al (1978) suggest one such choice for $\tau(t)$ which is based on an analysis of how the analytical properties of the transportation delay/CSTR model compare with the properties of another lumped-parameter approximation of the advection-diffusion, partial differential equation.

Bearing in mind these preceding considerations, and having noted that inclusion of a representation for $\tau(t)$ according to Appendix 4 produced apparently negligible differences, Assumption (3) was made at an early stage in the analysis and has since been preserved in all the models to be presented in section 3. Thus, by equation (3), equations (1) and (2) can be combined to give

$$\dot{\underline{x}}(t) = (\Theta_1(t)/a_1)\underline{u}(t) - (\Theta_1(t)/a_1)\underline{x}(t) + \underline{S}(t) + \underline{\xi}(t) \quad , \tag{4}$$

which, together with the output observations $y_1(t_k)$, $y_2(t_k)$ of *downstream* DO and BOD concentrations, respectively,

$$\begin{bmatrix} y_1(t_k) \\ y_2(t_k) \end{bmatrix} = \begin{bmatrix} x_1(t_k) \\ x_2(t_k) \end{bmatrix} + \begin{bmatrix} \eta_1(t_k) \\ \eta_2(t_k) \end{bmatrix} \quad . \tag{5}$$

is the *basic description of the internally descriptive model*. In equation (5),

$x_1(t_k)$, $x_2(t_k)$ are respectively the downstream concentration of DO and BOD at time $t_k (gm^{-3})$;

$\eta_1(t_k)$, $\eta_2(t_k)$ are respectively chance measurement errors associated with the output observations of DO and BOD(gm^{-3});

t_k is the kth sampling instant of time where the sampling interval $\Delta t = (t_k - t_{k-1}) = 1$ (day).

3.2 The Black Box Model

The black box model can briefly be formally stated as,

$$x(t_k) = A(q^{-1})x(t_k) + \sum_{i=1}^{m} B_i(q^{-1})u_i(t_k) \quad , \tag{6}$$

where the scalar $x(t_k)$, *either* the downstream DO concentration $(x_1(t_k))$ *or* the downstream BOD concentration $(x_2(t_k))$, is observed only in the presence of noise,

$$y(t_k) = x(t_k) + v(t_k) \quad . \tag{7}$$

Equations (6) and (7) are the *basic description of the black box model.* In equation (6) q^{-1} is defined as the backward shift operator,

$$q^{-1}\{x(t_k)\} = x(t_{k-1}) \text{ etc.} \quad , \tag{8}$$

and $A(q^{-1})$ and $B_i(q^{-1})$ are n-th order polynomials in q^{-1} defined as

$$\left.\begin{aligned} A(q^{-1}) &= \alpha_1 q^{-1} + \alpha_2 q^{-2} + \ldots + \alpha_n q^{-n} \quad , \\ B_i(q^{-1}) &= \beta_{io} + \beta_{i1} q^{-1} + \ldots + \beta_{in} q^{-n} \quad ; \quad i=1,2,\ldots,m \quad . \end{aligned}\right\} \tag{9}$$

The parameters α_i and β_{ij} are respectively elements of the vectors $\underline{\alpha}$ and $\underline{\beta}$ referred to in Table 2. $v(t_k)$ denotes that the random noise component of equation (7) is a lumped term which really covers the combined effects previously accounted for (conceptually) by $\underline{\xi}(t)$ and $\underline{\eta}(t_k)$ in the internally descriptive model.

Since the form of the black box model is restricted to the case of *single* output (state)* systems its application requires:

<u>Assumption (4)</u>: that (for black box representations) the dynamic behaviour of the downstream DO concentration can be considered independent of the dynamic behaviour of the downstream BOD concentration.

*There is a slight problem of terminology here; however, to all intents and purposes, "outputs" are equivalent to noise-corrupted observations of the "state" variables.

3.3 A Simple Method of Model Assessment

The method of model assessment is indeed simple. We must first, however, specify the exact nature of a *deterministic model prediction.*

For the *internally descriptive model* such a prediction is defined as the solution at time t_k of

$$\dot{\underline{x}}(t) = (\Theta_1(t)/a_1)\underline{u}(t) - (\Theta_1(t)/a_1)\underline{x}(t) + \underline{S}(t) \quad ; \quad \underline{x}(t_o) = \hat{\underline{x}}(t_o) \ , \quad (10)$$

given a set of (estimated) values for the initial conditions $\hat{\underline{x}}(t_o)$, the *measured* data for variables $\underline{u}(t_k)$ and $\underline{\Theta}(t_k)$, and estimated values for all parameters $\underline{a}$ implicit in the form of $\underline{S}(t)$. Precisely how the substitutions for $\underline{u}(t_k)$ and $\underline{\Theta}(t_k)$ are made will be defined in section 4.

For the *black box model* we have

$$x(t_k) = A(q^{-1})x(t_k) + \sum_{i=1}^{m} B_i(q^{-1})u_i(t_k) \quad ; \quad x(t_o) = \hat{x}(t_o) \ , \quad (11)$$

where $\hat{x}(t_o)$, $u_i(t_k)$, i=1,2,...,m, and values for the parameters $\underline{\alpha}$ and $\underline{\beta}$ are available.

From equations (10) and (11) the following vector (scalar) error quantities can be determined for the internally descriptive (black box) models,

$$\underline{\varepsilon}(t_k) = \underline{y}(t_k) - \hat{\underline{x}}(t_k) \quad , \quad (12)$$

and for each such *deterministic response error* sequence, $\varepsilon(t_k)$, we may compute corresponding sample mean, μ, and standard deviation, σ, statistics,

$$\mu = [1/(N - \delta)] \sum_{j=\delta}^{80} \varepsilon(t_j) \quad ; \quad \sigma = [1/(N - \delta - 1)] \sum_{j=\delta}^{80} (\varepsilon(t_j) - \mu)^2 \quad (13)$$

The notation of equation (13) indicates that the sampled measurements for the first day of the experiment are considered to have been taken at time t_o. Thus for all the internally descriptive models $\delta = 1$, i.e. an error can be computed for time t_1, but for the black box models δ is dependent upon n, the chosen order for the $B_i(q^{-1})$ and $A(q^{-1})$ polynomials--the reasons for this will become more evident in section 5.

We may note now that in section 3.1 and 3.2 the stochastic aspects of the models, $\underline{\xi}(t)$, $\underline{\eta}(t_k)$, are included simply for the purpose of completeness and for emphasizing the probabilistic framework of the modelling exercise. Further consideration of these terms is incidental to the main themes of the paper and only passing reference will be made to certain estimated forms of $v(t_k)$ in association with the black box modelling results, see Appendix 5.

4. Internally Descriptive Modelling Results

In this and the following section supporting remarks on model development and interpretation are restricted to a minimum. A sufficient body of literature already exists on the Cam (1972) modelling exercise, abstracts of which are given in Appendix 3.*

4.1 Model I [Beck and Young (1975)]

This is essentially a model based on the proposals of Dobbins (1964) and his assumptions are therefore reflected in the explicit form of $\underline{S}(t)$:

$$\begin{aligned}
&\text{(a)} \quad \text{DO} : \dot{x}_1(t) = (\Theta_1(t)/a_1)u_1(t) - (\Theta_1(t)/a_1)x_1(t) + a_2(\Theta_3(t) - x_1(t)) - a_3x_2(t) + a_4(t) \quad ; \\
&\text{(b)} \quad \text{BOD:} \dot{x}_2(t) = (\Theta_1(t)/a_1)u_2(t) - (\Theta_1(t)/a_1)x_2(t) - a_3x_2(t) + a_5(t) \quad .
\end{aligned} \qquad (14)$$

The additional variables are defined as

$u_1(t), u_2(t)$ = respectively the upstream (input) DO and BOD concentrations (gm^{-3});

$\Theta_3(t)$ = saturation concentration of DO (gm^{-3}).

where $\Theta_3(t)$ is computed from the following relationship with the stream water temperature $\Theta_2(t)$,

*Conversely, if there appears to be too much computational detail, this has been included to ensure that the objective of reproducability of results can be satisfied if necessary.

$$\Theta_3(t) = 14.54 - 0.39\Theta_2(t) + 0.01[\Theta_2(t)]^2 \quad . \tag{15}$$

The initial conditions, parameter values and definitions, and error statistics for this model are given in Table 3; a comparison of the deterministic model responses $\hat{\underline{x}}(t_k)$ and observations $\underline{y}(t_k)$ is given in Figure 4*. In Figure 4 the reader's attention is drawn to the performance of the model over the periods $t_{36} \rightarrow t_{48}$ (both the DO and BOD responses) and from t_{60} onwards (for the BOD response). Any significant improvement afforded by the later models will be most evident at these points in the experiment. The predicted downstream BOD concentration on day t_{58} should also be noted: it results from the effects of a thunderstorm on day t_{56}, giving rise to a peak upsteam BOD concentration on day t_{57} which probably led in turn to an actual peak downstream BOD some time between the samples of t_{57} and t_{58}. This then is precisely the kind of high frequency response characteristic that we should not expect the model to be able to reproduce accurately (see section 3.1). However, it is difficult to confirm that this is so since during high flow conditions the transportation delay in the reach approaches a minimum value and the daily sampling frequency of the data is consequently too *slow* to pick up the fast transient responses to the impulsive disturbance of the thunderstorm.

Computational note. Solutions to equation (14) are obtained iteratively by numerical integration (Runge-Kutta) over the interval $t_k \rightarrow t_{k+1}$. For this interval, therefore, the values,

$$\left.\begin{aligned} \underline{u}(t) &= \underline{u}(t_k) \\ \underline{\Theta}(t) &= \underline{\Theta}(t_k) \end{aligned}\right\} \text{ for } t_k \le t \le t_{k+1} \quad , \tag{16}$$

are substituted. Thus note that the alternative of linear interpolation may in fact yield more accurate results and especially so for the case of the storm conditions discussed above.

*See also Appendix 1 for comments on the salient features of the experimental data.

Table 3: Initial Conditions, Parameter Values, & Error Statistics for Model I

Variable (Parameter)	Definition	Value
$\hat{x}_1(t_o)$	Initial conditions for downstream DO concentration	$8.0\ gm^{-3}$
$\hat{x}_2(t_o)$	Initial conditions for downstream BOD concentration	$1.4\ gm^{-3}$
a_1	Volumetric hold-up in the reach	$1.51 \times 10^5\ m^3$
a_2	Reaeration rate constant	$0.17\ day^{-1}$
a_3	BOD decay rate constant	$0.32\ day^{-1}$
$a_4(t)$	Net rate of addition of DO to reach by combined effects of photosynthetic/respiratory activity of plants and algae and the decomposition of mud deposits	-2.7 for $0 \leq t \leq t_{19}$; -0.4 for $t > t_{19}$ (in $gm^{-3}\ day^{-1}$)
$a_5(t)$	Rate of addition of BOD to reach by local surface runoff	0 for all t (in $gm^{-3}\ day^{-1}$)
μ_1	Mean of errors in DO predictions	$0.234\ gm^{-3}$
σ_1	Standard deviation of errors in DO predictions	$0.838\ gm^{-3}$
μ_2	Mean of errors in BOD predictions	$0.820\ gm^{-3}$
σ_2	Standard deviation of errors in BOD predictions	$1.267\ gm^{-3}$

4.2 Model II [Beck and Young (1975, 1976)]

Whereas Model I does not account for the interaction of an algal population with the DO and BOD dynamics, this is incorporated into Model II by means of a new pseudo-empirical relationship for "sustained sunlight effects",

$$\text{(a) DO:}\quad \dot{x}_1(t) = (\Theta_1(t)/a_1)u_1(t) - (\Theta_1(t)/a_1)x_1(t) + a_2(\Theta_3(t) - x_1(t))$$

$$- a_3x_2(t) + a_4'(t) + a_6(\Theta_4(t) - a_8) \quad ;$$

(b) BOD: $\dot{x}_2(t) = (\Theta_1(t)/a_1)u_2(t) - (\Theta_1(t)/a_1)x_2(t) - a_3x_2(t) + a_5(t)$

$+ a_7(\Theta_4(t) - a_8)$;

where (17)

(c) $\Theta_4(t_k) = \Theta_4(t_{k-1}) + a_9[u_3(t_k)\{(\Theta_2(t_k) - a_{10})/a_{10}\} - \Theta_4(t_{k-1})]$;

with

(d) $(\Theta_4(t_k) - a_8) = 0$ for $\Theta_4(t_k) < a_8$.

The variables $u_3(t_k)$ and $\Theta_4(t_k)$ are defined as

$u_3(t_k)$ = hours of sunlight incident on the system at day t_k;

$\Theta_4(t_k)$ = "sustained sunlight effect" at day t_k (no specified units).

Figure 5 shows a significant improvement in the model responses, particularly over the period $t_{36} \rightarrow t_{48}$, given the additional initial conditions and parameter values listed in Table 4; the improved model performance is reflected in the error statistics also shown in Table 4. Model II requires,

Assumption (5): that the higher observed DO and BOD conditions for $t_{36} \rightarrow t_{48}$ are due to the growth of an algal population, which in turn is some function of the *cumulative* influence of warm, sunny periods of weather.

Computational note. Conditions similar to those of equation (14) hold for the solution of equation (17). A further condition is, in equations (17a) and (17b),

$$\Theta_4(t) = \Theta_4(t_k) \text{ for } t_k \leq t \leq t_{k+1} \quad . \tag{18}$$

4.3 Model III [Beck (1974, 1975)]

The discrete-time low-pass filter mechanism for the sustained sunlight effect in Model II, equation (17c), has an analog continuous-time form.

Table 4: Error Statistics and Additional Initial Conditions and Parameter Values for Model II

Variable (Parameter)	Definition	Value
$\Theta_4(t_o)$	Initial conditions for sustained sunlight effect	0.0*
$a_4'(t)$	Rate of addition of DO to reach by decomposition of bottom mud deposits	(as for $a_4(t)$)
a_6	Coefficient for sustained sunlight effect in DO equation	0.31*
a_7	Coefficient for sustained sunlight effect in BOD equation	0.32*
a_8	Threshold level for sustained sunlight effect	6.0*
a_9	Reciprocal time constant for discrete-time low-pass filter for the sustained sunlight effect, equation (17c)	0.25 day^{-1}
a_{10}	Arbitrary mean river water temperature	8.0 ^oC
μ_1	Mean of errors in DO predictions	-0.144 gm^{-3}
σ_1	Standard deviation of errors in DO predictions	0.675 gm^{-3}
μ_2	Mean of errors in BOD predictions	0.332 gm^{-3}
σ_2	Standard deviation of errors in BOD predictions	0.965 gm^{-3}

*No specific units are assigned to these quantities owing to the dimensional anomaly of equation (17c).

On the basis of certain observations [Beck (1975)] it is found to be more appropriate, however, to simulate the growth and interaction effects of an algal population by *two* low-pass filters in series:

$$\text{(a) DO:} \quad \dot{x}_1(t) = (\Theta_1(t)/a_1)u_1(t) - (\Theta_1(t)/a_1)x_1(t) + a_2(\Theta_3(t) - x_1(t)) - a_3x_2(t) + a_4'(t) + a_{11}(x_3(t) - a_{12}) \; ; \qquad (19)$$

(b) BOD: $\dot{x}_2(t) = (\Theta_1(t)/a_1)u_2(t) - (\Theta_2(t)/a_1)x_2(t) - a_3x_2(t) + a_5(t)$

$+ a_{13}(x_4(t) - a_{14})$;

(c) $\dot{x}_3(t) = -(1/a_{15})x_3(t) + (a_{16}/a_{15})u_3(t)$; (19) Cont'd.

(d) $\dot{x}_4(t) = -(1/a_{17})x_4(t) + (1/a_{17})x_3(t)$,

in which $x_3(t)$ = output of first low-pass filter (no specified units),
$x_4(t)$ = output of second low-pass filter (no specified units).

Notice that $x_3(t)$ interacts only with the downstream DO concentration, while $x_4(t)$ interacts only with the downstream BOD concentration; $x_3(t)$ and $x_4(t)$ therefore fulfil in equations (19a) and (19b) the equivalent roles of $\Theta_4(t)$ in equations (17a) and (17b). Table 5 summarizes the parameter values, initial conditions, and deterministic response error statistics for Model III and a comparison of the model performance with the observed behaviour is given in Figure 6. Model III can be seen to be only marginally "better" at fitting the data than Model II; however, equation (19) is useful primarily as a conceptual link between Models II and IV.

Computational note. The conditions of equation (14) together with the substitution

$$u_3(t) = u_3(t_k) \text{ for } t_k \leq t \leq t_{k+1} \quad , \tag{20}$$

in equation (19c) hold for solutions of equation (19). The inequality constraint of equation (17d) is not transferred in any equivalent form to equation (19).

4.4 Model IV [Beck(1974, 1975)]

The synthesis of ModelIV depends essentially upon interpreting $x_3(t)$ and $x_4(t)$ in Model III, equation (19), as

$x_3(t)$ = downstream concentration of a live algal population (gm^{-3})
$x_4(t)$ = downstream concentration of a dead algal population (gm^{-3})

and upon the assumption that algal population growth kinetics can be described by Monod (1949) kinetics with sunlight as the rate-limiting factor.

Table 5: Error Statistics and Additional Initial Conditions and Parameter Values for Model III

Variable (Parameter)	Definition	Value
$\hat{x}_3(t_o)$	Initial conditions for output of first low-pass filter	1.0*
$\hat{x}_4(t_o)$	Initial conditons for output of second low-pass filter	1.0*
a_{11}	Coefficient for equivalent sustained sunlight effect in DO equation	0.115*
a_{12}	"Threshold" level for equivalent sustained sunlight effect in DO equation	6.0*
a_{13}	Coefficient for equivalent sustained effect in BOD equation	0.146*
a_{14}	"Threshold" level for equivalent sustained sunlight effect in BOD equation	6.0*
a_{15}	Time-constant for first low-pass filter	1.95 day
a_{16}	Gain coefficient between $u_3(t)$ and $x_3(t)$	2.33*
a_{17}	Time constant for second low-pass filter	1.42 day
μ_1	Mean of errors in DO predictions	-0.328 gm^{-3}
σ_1	Standard deviation of errors in DO predictions	0.672 gm^{-3}
μ_2	Mean of errors in BOD predictions	-0.105 gm^{-3}
σ_2	Standard deviation of errors in BOD predictions	0.880 gm^{-3}

*No specific units are assigned to these variables and parameters.

For Model IV we have then

(a) DO: $$\dot{x}_1(t) = (\Theta_1(t)/a_1)u_1(t) - (\Theta_1(t)/a_1)x_1(t) + a_2(\Theta_3(t) - x_1(t)) - a_3x_2(t) + a_4'(t) + a_{18}x_3(t)[u_3(t)]^{a_{19}} - a_{20}x_3(t) \quad ;$$

(b) BOD: $$\dot{x}_2(t) = (\Theta_1(t)/a_1)u_2(t) - (\Theta_1(t)/a_1)x_2(t) - a_3x_2(t) + a_{21}x_4(t) \quad ;$$

(c) Live algae: (21)

$$\dot{x}_3(t) = -(\Theta_1(t)/a_1)x_3(t) + a_{22}u_3'(t)/(a_{23} + u_3'(t)) - a_{24}x_3(t) \quad ;$$

(d) Dead algae:

$$\dot{x}_4(t) = -(\Theta_1(t)/a_1)x_4(t) + a_{25}x_3(t) - a_{26}x_4(t) - a_{27} \quad ,$$

where

$$u_3'(t) = u_3(t - \Delta t) \quad , \tag{22}$$

with Δt being a pure time delay of one day, i.e. one sampling interval. The deterministic predictions of Model IV are shown in Figure 7. All other necessary information about the model is provided in Table 6. From both Figure 7 and Table 6 it is evident that Model IV is capable of a better representation of the observed system behaviour than Model II; the most significant improvement offered by Model IV concerns the simulation of the downstream BOD response from about day t_{60} onwards--Figure 7(b). Two major assumptions have been made in the derivation of equation (21):

<u>Assumption (6)</u>: that no live or dead/decaying algal matter enters the reach of river across its upstream boundary.

<u>Assumption (7)</u>: that the growth kinetics of the algal population, in equation (21c) are zero-order with respect to the concentration of live algae.

Table 6: Error Statistics and Additional Initial Conditions and Parameter Values for Model IV

Variable (Parameter)	Definition	Value
$\hat{x}_3(t_o)$	Initial conditions for concentration of live algae	0.25 gm^{-3}
$\hat{x}_4(t_o)$	Initial conditions for concentration of dead algae	0.1 gm^{-3}
a_{18}	Rate constant for photosynthetic production of DO by live algae	1.45*
a_{19}	Exponential power for dependence of algal photosynthetic DO production on sunlight conditions	0.55*
a_{20}	Rate constant for respiratory consumption of DO by live algae	2.0 (gm^{-3} DO · $[gm^{-3}$ algae$]^{-1}day^{-1}$)
a_{21}	Rate constant for BOD production by redissolved dead algal material	4.0 (gm^{-3} BOD · $[gm^{-3}$ algae$]^{-1}day^{-1}$)
a_{22}	Maximum specific growth-rate of algae	2.1 gm^{-3} day^{-1}
a_{23}	Saturation constant for growth-rate limiting factor	20 hrs sunlight day^{-1}
a_{24}	Specific decay rate constant for algae	0.35 day^{-1}
a_{25}	Rate constant for production of dead algal matter from live algal matter	1.05 day^{-1}
a_{26}	Rate constant for redissolution of deal algal material	0.25 day^{-1}
a_{27}	Rate of sedimentation of particulate dead algal material	0.11 gm^{-3} day
μ_1	Mean of errors in DO predictions	-0.073 gm^{-3}
σ_1	Standard deviations of errors in DO predictions	0.642 gm^{-3}
μ_2	Mean of errors in BOD predictions	-0.189 gm^{-3}
σ_2	Standard deviations of errors in BOD predictions	0.799 gm^{-3}

*No specific units are assigned to these parameters.

Computational note. In equation (21c) the following substitution is made, through equation (22),

$$u_3(t) = u_3'(t_k - \Delta t) = u_3(t_{k-1}) \text{ for } t_k \leq t \leq t_{k+1} \quad . \tag{23}$$

It has been suggested that such a "delaying" factor may be due in part to the presence of a stored phase of algal population mass. (Note also that the concentrations of live and dead algal populations are somewhat arbitrary; they do not, for example, have any intentional equivalence to determinations such as chlorophyll-A and dry biomass measurements.)

5. Black Box Modelling Results

5.1 Model Va

Recalling Assumption (4) from section 3.2, Model Va is a model which describes the behaviour of the downstream DO concentration independently of the behaviour of the downstream BOD concentration. In line with equation (11) Model Va is given by

$$x_1(t_k) = \alpha_1 x_1(t_{k-1}) + \beta_{11}u_1(t_{k-1}) + \beta_{30}u_3(t_k) + \beta_{31}u_3(t_{k-1}) \;, \tag{24}$$

which generates the response of Figure 8(a)--the continuous line response--with parameter values and error statistics as indicated in Table 7.

Footnote. The parameter values for Table 7 differ slightly from those quoted elsewhere in Beck (1974) and Beck (1978a). This discrepancy results from the use of two alternative parameter estimation schemes:

i) the Maximum Likelihood method of Åström and Bohlin (1966); and

ii) the recursive Instrumental Variable method of Young (1974).

The estimates of Table 7, and similarly the estimates in Tables 8 and 9 below, are derived using the latter estimator. Additional details are given in Appendix 5.

5.2 Model Vb

It can be shown [Beck (1976)] that the parameters β_{30} and β_{31} in equation (24) have a tendency to be non-stationary, i.e. they vary with

Table 7: Parameter Values, Initial Conditions and Error Statistics for Model Va

Parameter/Variable	Value	Comments
t_k	t_1	Starting time in equation (24)
δ	1	See equation (13)
$\hat{x}_1(t_o)$	8.0 gm^{-3}	Initial conditions for $\hat{x}_1(t_k)$
α_1	0.639	
β_{11}	0.229	
β_{30}	0.062	
β_{31}	0.051	
μ_1	0.016 gm^{-3}	Mean of prediction errors
σ_1	0.827 gm^{-3}	Standard deviation of prediction errors

Table 8: Parameter Values and Error Statistics for Model Vb.

Parameter/Variable	Value	Comments
μ_{Θ_1}	1.28×10^5 $m^3 day^{-1}$	Sample mean value for stream discharge
α_1	0.596	
β_{11}	0.261	
β'_{30}	0.060	
β'_{31}	0.052	
μ_1	0.020 gm^{-3}	Mean of prediction errors
σ_1	0.674 gm^{-3}	Standard deviation of prediction errors

time. Part of this time-variability of the parameters can be accounted for by restating equation (24) as:

$$x_1(t_k) = \alpha_1 x_1(t_{k-1}) + \beta_{11}u_1(t_{k-1}) + \beta_{30}'(\mu_{\Theta_1}/\Theta_1(t_k))u_3(t_k) + \beta_{31}'(\mu_{\Theta_1}/\Theta_1(t_{k-1}))u_3(t_{k-1}) \quad . \tag{25}$$

In other words we are proposing that the parameters β_{30} and β_{31} in equation (24) are not time-invariant but are better represented as functions of $\Theta_1(t_k)$, the stream discharge; μ_{Θ_1} is a sample mean value for $\Theta_1(t_k)$ introduced to normalize the associated expressions in equation (25). The results of Model Vb are summarized in Figure 8(a)--the dashed line response--and Table 8.

5.3 Model Vc

No such time-variability of the parameters as identified for the DO dynamics is discernible in the corresponding black box model for the downstream BOD behaviour, i.e.

$$x_2(t_k) = \alpha_1 x_2(t_{k-1}) + \beta_{21}u_2(t_{k-1}) + \beta_{32}u_3(t_{k-2}) + \beta_{34}u_3(t_{k-4}). \tag{26}$$

The results of Model Vc, which according to the error statistics shows a markedly better fit to the data than any of the other models, are given in Table 9 and Figure 8(b). The value of $\delta = 4$ in Table 9 arises because of the term $u_3(t_{k-4})$ in equation (26) which implies that $x_2(t_4)$ is calculated from the measured value of $u_3(t_o)$.

6. Summary of Results--Some Critical Comments

The complete set of deterministic response error statistics for Models I-V are given in Table 10; in addition a relative measure is provided of the error variance as a percentage of the variance of the original time-series data.

6.1 Model Assessment

Since Models I through IV represent a conceptual development of DO-BOD-algae interaction models [see Beck (1978b)] it is reassuring to find that parallel with this development there runs a successive reduction of model

Table 9: Parameter Values, Error Statistics and Initial Conditions for Model Vc

Parameter/Variable	Value	Comments
t_k	t_4	Starting time in equation (26)
δ	4	
$\hat{x}_2(t_3)$	1.6 gm^{-3}	Initial conditions for $\hat{x}_2(t_k)$
α_1	0.826	
β_{21}	0.054	
β_{32}	0.034	
β_{34}	0.057	
μ_2	0.030 gm^{-3}	Mean of prediction errors
σ_2	0.668 gm^{-3}	Standard deviation of prediction errors

Table 10: Survey of Error Statistics for All Models

	DO			BOD		
	$\mu_1 (gm^{-3})$	$\sigma_1 (gm^{-3})$	σ_1^2/σ_D^2 (%)	$\mu_2 (gm^{-3})$	$\sigma_2 (gm^{-3})$	σ_2^2/σ_B^2 (%)
Original Data	7.282(μ_D)	1.067(σ_D)	-	4.112(μ_B)	1.265(σ_B)	-
Model Errors						
I	0.234	0.838	61.7	0.820	1.267	100.0
II	-0.144	0.675	40.0	0.332	0.965	58.2
III	-0.328	0.672	39.7	-0.105	0.880	48.4
IV	-0.073	0.642	36.2	-0.189	0.799	39.9
Va	0.016	0.827	60.1	-	-	-
Vb	0.020	0.674	39.9	-	-	-
Vc	-	-	-	0.030	0.668	27.9

fitting error variances. For the model representation of the DO dyanmics it is apparent that Model II offers, for this particular data set, a distinct improvement over the a priori model (Model I) but that thereafter Models III and IV provide only marginal increments in accuracy. It can be concluded, nevertheless, that the more significant improvements and alterations in the description of BOD dynamics in Models III and IV do not degrade the performance of the models with respect to the DO dynamics. Model IV (the a posteriori model), in particular, reflects a concentration of effort on improving the simulated BOD responses over the final period of the data; a reward for this effort in terms of a relatively large drop in the response error variance for BOD is thus, perhaps, only to be expected. The black box modelling results reveal two important features:

i) that the introduction of time-varying coefficients in Model Vb substantially improves upon the accuracy of Model Va; and

ii) that the rather simply structured black box model for BOD, Model Vc, gives a better performance than all of the more complex internally descriptive models.

This latter point prompts the questions of how, in model assessment, does one determine which model is "best" *in some sense*, upon what criteria should this judgement be based, and can we measure whether a "significant" addition of model complexity is matched by a correspondingly "significant" addition in model accuracy. Although certain aspects of these questions may be answered by the argument that the choice of the correct model depends upon the intended model application, it is still useful to consider other aspects of the questions in a fairly general, abstract context.

Most systems analysts are aware of the intuitive notion that the quality of a model is judged by some balance between model accuracy and model complexity. So to assess the models presented here on the basis of fitting error statistics alone assumes a somewhat narrow view of model assessment, especially when the sample number of observations is probably too small to lend significant meaning to such an analysis of variance. The crucial problem, of course, is the development of some more representative measure which can be applied with ease and which allows the comparison of quite differently structured models, e.g. partial differential equations, ordinary differential equations, difference equations. In this respect recent results of Maciejowski (1977) are potentially of considerable interest.

By using the theory of algorithms and by borrowing ideas from algorithmic information theory Maciejowski is able to construct a measure of model "goodness" derived from a comparison of the lengths of two specially defined computer programs. The first program, or base program, simply generates a look-up table for the original data sequence. The second program embodies the algorithms that compute the set of model predictions *and* it also generates a look-up table of the associated model fitting errors, i.e. the length of this second program is a function of model complexity and model accuracy. Thus the shorter the length of the candidate model's program the better is said to be the capability of that model to represent the observed process behaviour. It is worth noting, then, that for the restricted case of Models Va and Vc as a joint model of DO and BOD dynamics in the Cam, Maciejowski (1977) arrives at the following conclusion: that (depending upon certain technical details of program coding) a model with equally good "predictive power" would be one which merely draws a straight line, the respective sample average values, through the downstream DO and BOD data points!

6.2 Accuracy of the Models as Representations of the Physcial System*

Apart from such portentous statements as the above on this particular modelling exercise, the major grounds for critical comment and appraisal concern the biological/ecological content of Model IV, this model being the end-product of the analysis.

Firstly, the ecology, such as it is, is clearly naive and macroscopic in its approximation to reality. The biological processes of death, decay, and redissolution of dead algal material are, in particular, the weakest hypotheses in the model. If the dead algal material does indeed lead to the production of an additional BOD load in the river, then Model IV is better at predicting this effect over the latter period of the experiment than any of the other internally descriptive models. It is suspected that the primary factor in providing the better prediction is the inclusion of stream discharge in the mass balance for the dead algal population, although it is not evident how the effect might be related to the low flows dominant at that time. In any event, the issues of why and whether it is dead algal matter *in the river* that causes the apparently high downstream BOD's cannot be resolved on the basis of the field data for two reasons:

*See also Appendix 3 for further comments on diurnal variations and sedimentation processes in the River Cam.

- no measurements of phytoplankton in the river are available;
- algal respiration *in the five-day BOD bottle* test will equally give rise to a higher BOD measurement.

Secondly, the possibility of nitrification in the river cannot be discounted and this too might give rise to erroneously high carbonaceous BOD observations at the downstream system boundary. Evidence obtained during the experiment, however, indicates that patterns of oxygen uptake rates in the BOD test are essentially similar for samples taken from both the upstream and downstream locations. From this it would be difficult to establish whether nitrification was or was not significant; but neither should it be concluded that nitrification is really responsible for the effects described here by the introduction of live and dead algal population balances. Later evidence from a similar experiment in 1975, an exceptionally hot summer, suggests both that the sewage works obtains a high level of nitrification and that factors relating to the aqueous nitrogen cycle are of considerable importance in this stretch of river.

Thirdly, the form of the Monod growth-rate function, and the justification for its introduction, require careful consideration. For example, the relative magnitudes of the saturation constant, a_{23}, and the typical values for sunlight conditions imply that growth-rates are in practice approximately *linearly dependent* upon $u_3'(t)$. Is there, therefore, any valid reason for retaining the additional complexity of the Monod function in the model? The description of algal growth kinetics is not strictly speaking that of Monod growth kinetics since it is independent of the concentration of live algae. In addition, is it feasible that the growth cycle of algae might be better approximated by the "conceptual analog" of *three*, as opposed to two, low-pass filters in series? We might hypothesize that the outputs of the three filters are equivalent to "stored", "active" and "dead" phases of the algal population where,

- the stored algae do not interact with the DO and BOD dynamics but have a growth-rate which is a function of sunlight conditions and the concentration of the active population;
- otherwise the active and dead algal masses fulfil the roles of live and dead algae, respectively, as in Model IV, with the rate of production of the active state being a function of the concentration of the stored algal matter and not a function of sunlight conditions.

Such hypotheses would, in principle, give some justification for the data manipulation of using $u_3'(t)$ instead of $u_3(t)$ in the algal growth-rate function of Model IV.

Fourthly, on points of somewhat finer detail, the field data do not permit any resolution of whether the photosynthetic DO production is due to attached or floating algae; the assumption here has been that it is the latter. The evidence available would certainly suggest that stream flow-rate is important in determining the amount of DO produced by photosynthesis. For instance, implicit in Model Vb--notably a black box model--is the relationship that as flow-rate decreases, the sunlight incident on the river produces a proportionately higher amount of dissolved oxygen. But beyond this kind of macroscopic cause-effect relationship it is not easy to distinguish between the relative significances of attached or floating algae, even though a corollary of the proposed relationship would be a dependence of photosynthesis rates on turbidity.

Next, in connection with more familiar aspects of DO-BOD models it can be argued that a_2 and a_3, the reaeration rate and BOD decay rate constants, should properly be accounted for as functions of flow-rate and temperature. The estimated evaluation of the parameter $a_4'(t)$ in Table 3 should also be questioned. The most probable reasons for the apparently higher initial estimated rates of oxygen consumption by bottom mud deposits are as follows:

- that the BOD measurements for $t_0 \rightarrow t_{13}$ are systematically biased, being lower than the true values of in-stream BOD concentrations (see Appendix 1);
- that the downstream DO sensor had been drifting prior to day t_{20} when it was recalibrated--there are, however, no records now available with which to check this supposition.

Finally, as mentioned in the introductory section of the paper, the approach adopted for modelling transport and mixing properties of the river reach is not the approach commonly encountered in the literature. Further to the discussion of section 3.1 it is possible that alternative approximations to the hydrodynamical regime of the river, incorporating techniques such as that outlined in Appendix 4, may give both better characterizations of the experimental data and different insights to the observed ecological/biochemical behaviour. Since the fundamental philosophy underlying the development of Model IV from Model I is one which embodies a large measure of confidence in the assumptions of the a priori model, a re-examination of these assumptions would lead to a re-examination of all the subsequent models. Expectations of substantially different results, nevertheless, should perhaps not be too high. The a priori model, Model I, can be said to simulate observed behaviour adquately except for certain quite specific intervals of the experiment. Thus when expressions for the sources and sink terms of Model I are cast within a different set of assumptions about transport and mixing properties of the river, the net result might only be a change in the estimated values for the associated parameters a_2, a_3, $a_4(t)$. Thereafter, our interpretations of the desired structural modifications of Model I, although not necessarily the additional parameter values, might remain essentially similar to those made here.

7. Conclusions

The objectives of this paper have been:

i) the dissemination of field data which can be used for the verification of DO-BOD interaction models; and

ii) the comparison of a number of such (dynamic) models which have been derived by reference to those field data.

The opportunity has also been taken to present a summarizing and concluding statement on modelling studies with respect to the Cam (1972) experiment. Many questions remain unresolved and it is hoped that the interest of others will be sufficiently stimulated to provide alternative answers. Some of these questions concern the following:

- the development of terms for expressing the decay, redissolution and exertion of a BOD by dead algal material;
- the relationship for growth of an algal population with sunlight conditions as a rate limiting factor;
- the possibilities for different interpretations of the observed behaviour of DO-BOD interaction when different assumptions are made about the transport and dispersive properties of the reach of river.

In this context, one of the problems of working with the same set of field data over an extended period of time is that the analyst becomes blind to certain new avenues of thought.

On the other hand, any experimental data, if they are carefully collected, are worthy of a broadly based analysis which explores differing ways of identifying the basic cause/effect mechanisms governing the system's observed behaviour. The problem in this context, however, is that the modelling exercise eventually approaches the limits in the accuracy and scope (i.e. the number of state variables measured) of the field data. When this limiting point has been reached what is really required is another, better designed, and more comprehensive experiment. For the Cam such an experiment was conducted in the summer of 1975 and the associated data are currently receiving a preliminary analysis.

It is hoped that the Cam (1972) data will have some usefulness beyond the requirements of model verification studies. Perhaps this usefulness will be rather modest for the purposes of investigating operational control schemes, which, with regard to water quality management in river basins, await a number of technical developments before they can have a proper focus on reality. The area of system identification and parameter estimation is probably where the data can be used to the greatest advantage. Apart from the possibilities for parameter estimation in partial differential equation model forms, real field data from a familiar system provide the basis for an excellent tutorial on the use of the various available algorithms of analysis, e.g. Extended Kalman Filtering, Maximum Likelihood, and Instrumental Variable methods.

Acknowledgements and Author's Footnote

During the course of carrying out the experiment and throughout all the subsequent analytical effort I have had good reason to become indebted to a number of people and organizations for their help. I should therefore like to thank members of the Anglian Water Authority (then the Great Ouse River Authority) and the Water Research Centre, Stevenage (then the Water Pollution Research Laboratory) for their assistance in loaning equipment and laboratory facilities for the experiment. From the point of view of guidance and discussion in the interpretation of results and in the application of identification and estimation algorithms I am especially grateful to Professor Peter Young and Dr. Paul Whitehead, both now with the Centre for Resource and Environmental Studies at the Australian National University, and to Gustaf Olsson of the Division of Automatic Control, Lund Institute of Technology, Sweden. The financial support of the Royal Society, London, during one year in Sweden and for a further three years in Cambridge is gratefully acknowledged.

From time to time I have been asked to provide colleagues with a copy of the Cam (1972) field data. This has encouraged me in the hope that still others too will find these data useful for their own purposes. If this hope is realized then I should be pleased to receive details of any associated results.

REFERENCES

Åström, K.J. and Bohlin, T. (1966) - *Numerical Identification of Linear Dynamic Systems from Normal Operating Records* - in P.H. Hammond, Ed., "Theory of Self-Adaptive Control Systems", Plenum, New York, 96-111.

Beck, M.B. (1974) - *Maximum Likelihood Identification Applied to DO-BOD-Algae Models for a Freshwater Stream* - Report 7431(C), Division of Automatic Control, Lund Institute of Technology, Sweden.

Beck, M.B. (1975) - *The Identification of Algal Population Dynamics in a Non-Tidal Stream* - in G.C. Vansteenkiste, Ed., "Computer Simulation of Water Resources Systems", North-Holland, Amsterdam, 483-494.

Beck, M.B. (1976) - *Identification and Parameter Estimation of Biological Process Models* - in G.C. Vansteenkiste, Ed., "System Simulation in Water Resources", North-Holland, Amsterdam, 19-43.

Beck, M.B. (1978a) - *Modelling of Dissolved Oxygen in a Non-Tidal Stream* - in A. James, Ed., "Mathematical Models in Water Pollution", John Wiley, London, 137-166.

Beck, M.B. (1978b) - *Random Signal Analysis in an Environmental Sciences Problem* - Applied Mathematical Modelling, 2, 1, 23-29.

Beck, M.B. (1978c) - *Model Structure Identification from Experimental Data* - in E. Halfon, Ed., "Theoretical Systems Ecology", Academic, New York (in press).

Beck, M.B. and Young, P.C. (1975) - *A Dynamic Model for DO-BOD Relationships in a Non-Tidal Stream* - Water Research, 9, 769-776.

Beck, M.B. and Young, P.C. (1976) - *Systematic Identification of DO-BOD Model Structure* - Proc. Am. Soc. Civil Engrs., J. Env. Eng. Div., 102, EE5, 909-927.

Buffham, B.A. and Gibilaro, L.G. (1970) - *A Unified Time Delay Model for Dispersion in Flowing Media* - Chemical Engineering Journal, 1, 31-36.

Coggan, G.C. and Noton, A.R.M. (1970) - *Discrete-Time Sequential State and Parameter Estimation in Chemical Engineering* - Trans.Instn. Chem. Engrs., 48, T255-T264.

Himmelblau, D.M. and Bischoff, K.B. (1968), Process Analysis and Simulation, John Wiley, New York.

Maciejowski, J.M. (1977) - *The Assessment of Rival Models of the Same Finite Set of Observations* - in H. Akashi, Ed., "Preprints IFAC Symposium on Environmental Systems Planning, Design and Control", Pergamon, Oxford, 227-231.

Monod, J. (1949) - *The Growth of Bacterial Cultures* - Annual Review of Microbiology, 3, 371-394.

Rinaldi, S., Soncini-Sessa, R., Stehfest, H. and Tamura, H. (1978) - *Modelling and Control of River Quality* - McGraw-Hill, Düsseldorf (forthcoming).

Streeter, H.W. and Phelps, E.B. (1925) - *A Study of the Pollution and Natural Purification of the Ohio River* - Bulletin No. 146, U.S. Public Health Service.

Whitehead, P.G. (1976) - *Dynamic Models in the Planning Role* - in G.C. Vansteenkiste, Ed., "System Simulation in Water Resources", North-Holland, Amsterdam, 83-98.

Whitehead, P.G. (1977) - *A Dynamic Stochastic Model for Non-Tidal Rivers* - in "The Bedford Ouse Study: Symposium Proceedings", Anglian Water Authority, Huntingdon, 49-75.

Whitehead, P.G. and Young, P.C. (1975) - *A Dynamic-Stochastic Model for Water Quality in Part of the Bedford Ouse River System* - in G.C. Vansteenkiste, Ed., "Computer Simulation of Water Resources Systems", North-Holland, Amsterdam, 417-438.

Young, P.C. (1974) - *A Recursive Approach to Time-Series Analysis* - Bulletin Institute of Mathematics and its Applications, 10, 209-224.

Young, P.C. and Beck, M.B. (1974) - *The Modelling and Control of Water Quality in a River System* - Automatica, 10, 5, 455-468.

Young, P.C. and Whitehead, P.G. (1977) - *A Recursive Approach to Time-Series Analysis for Multivariable Systems* - International Journal of Control, 25, 457-482.

APPENDIX 1

The Cam (1972) Experimental Field Data

Table A1 lists the Cam (1972) experimental field data. The columns of data are denoted respectively as follows:

Column 1: Sample data time, for modelling purposes (day)

Column 2: Date, with respect to 1972

Column 3: Upstream DO concentration (gm^{-3})

Column 4: Upstream BOD concentration (gm^{-3})

Column 5: Downstream DO concentration (gm^{-3})

Column 6: Downstream BOD concentration (gm^{-3})

Column 7: Stream discharge ($10^5 m^3 day^{-1}$)

Column 8: Stream temperature (^oC)

Column 9: Sunlight incident upon local area (hrs. per day)

Column 10: Rainfall in local area (mm)

N.B. (i) The underlined value at t_{34} in column 5 denotes a value interpolated for a missing downstream DO concentration observation.

(ii) The underlined value at t_{57} in column 4 denotes a value of upstream BOD concentration measured after a thunderstorm on day t_{56} (see rainfall - column 10); in some analyses, see Appendix 5, the given value was substituted by a value of 6.50 (gm^{-3}).

(iii) The measured values for BOD concentrations, columns 4 and 6, during $t_0 \rightarrow t_{13}$ (inclusive) are suspected to be underestimates of the true stream BOD conditions. These measurements are derived on the basis of carrying out the five-day BOD bottle test on diluted samples of river water; for the initial period of the experiment it had been anticipated that stream BOD levels might be quite high. In the event this precautionary measure was unnecessary and subsequent comparisons of BOD's obtained from diluted and undiluted samples indicated that analyses of the diluted samples gave consistently low BOD readings. This observation is partially confirmed when BOD measurements of the sewage works effluent are substituted for other modelling purposes (see Appendix 3).

TABLE A1: The Cam (1972) Experimental Field Data

1	2	3	4	5	6	7	8	9	10
0	June 6	9.67	2.00	8.00	2.30	1.71	16.50	12.90	0.0
1	7	9.56	2.25	8.20	1.05	1.77	16.00	8.10	1.2
2	8	9.25	2.40	8.00	1.65	1.64	15.50	7.30	1.5
3	9	9.36	2.75	7.50	1.55	1.70	15.50	2.80	3.5
4	10	9.57	1.90	7.20	1.60	1.55	15.75	3.10	0.0
5	11	9.43	2.75	7.30	2.90	1.80	15.50	7.70	0.0
6	12	9.52	1.95	7.00	1.35	1.66	16.00	4.90	0.1
7	13	9.32	2.80	6.40	2.00	1.71	16.75	1.20	0.0
8	14	9.04	2.45	6.40	1.55	1.61	17.00	11.30	0.0
9	15	9.09	3.05	6.60	2.55	1.63	17.00	0.80	4.5
10	16	8.99	2.75	6.60	3.05	1.71	17.50	11.90	0.1
11	17	8.94	4.70	6.70	2.70	1.55	18.00	9.20	0.0
12	18	9.08	4.70	6.70	3.20	1.58	17.25	0.00	1.0
13	19	9.23	3.40	7.10	2.30	1.65	16.50	13.50	1.6
14	20	9.32	7.45	6.90	5.25	1.57	16.25	7.40	0.0
15	21	9.29	5.35	6.70	5.25	1.61	15.75	0.00	0.0
16	22	9.39	4.40	6.30	4.60	1.57	15.25	6.40	0.0
17	23	9.47	4.50	6.60	4.15	1.64	15.50	8.30	0.0
18	24	9.09	6.30	5.50	4.45	1.64	16.00	0.70	0.0
19	25	8.81	4.70	6.00	4.30	1.55	17.00	6.60	0.0
20	26	8.81	4.50	7.30	3.45	1.44	17.75	2.80	0.0
21	27	8.56	7.30	7.00	5.35	1.41	18.75	2.50	4.5
22	28	8.71	7.10	7.00	5.00	1.44	18.50	10.50	0.0
23	29	8.27	6.80	7.40	5.15	1.33	19.00	10.30	0.0
24	30	8.31	6.05	7.10	4.95	1.31	19.75	11.70	1.4
25	July 1	8.89	7.55	6.60	5.40	1.36	19.00	1.30	2.0
26	2	8.66	4.60	6.20	5.95	1.40	17.25	0.00	0.8
27	3	8.20	5.95	7.00	4.30	1.31	17.50	6.10	1.6
28	4	8.29	7.10	7.00	6.15	1.29	17.00	2.70	0.0
29	5	8.44	7.95	7.40	5.10	1.22	17.25	5.50	0.0
30	6	8.00	7.45	7.40	4.70	1.24	18.00	9.90	0.0
31	7	8.62	6.70	7.20	4.80	1.19	18.00	1.40	4.2
32	8	9.07	5.35	6.80	4.85	1.31	18.00	0.50	5.6
33	9	8.62	4.00	6.50	4.90	1.38	17.25	1.40	0.0
34	10	8.80	4.00	7.05	3.35	1.32	17.50	6.60	0.0
35	11	8.85	3.75	7.60	3.40	1.17	17.75	11.20	0.0
36	12	9.21	3.70	8.00	2.65	1.14	18.50	11.60	0.0
37	13	8.89	3.15	8.60	3.25	1.12	19.50	12.50	0.0
38	14	9.01	3.70	9.00	4.10	1.19	20.50	13.60	0.0
39	15	9.05	4.45	10.20	4.60	1.03	20.50	12.60	0.0
40	16	8.55	3.80	8.90	4.90	1.01	20.50	9.60	0.0
41	17	8.54	4.70	11.00	5.85	1.03	20.75	12.90	0.0
42	18	8.62	6.05	10.60	6.60	1.00	20.50	6.50	0.0
43	19	7.85	5.20	9.30	6.40	1.01	20.50	5.00	0.0
44	20	7.37	5.15	8.20	6.45	1.03	20.25	5.60	0.0
45	21	7.67	3.40	7.20	6.00	1.03	20.25	0.00	3.9
46	22	7.48	5.45	5.60	5.50	1.03	19.75	0.00	0.9
47	23	7.47	6.10	5.80	4.35	1.15	19.75	4.40	0.5
48	24	7.61	6.55	5.90	4.05	1.12	20.25	2.00	0.0
49	25	7.38	6.25	5.80	4.25	1.05	20.50	5.60	0.0
50	26	7.37	7.35	6.30	4.20	1.03	20.00	2.20	0.0

TABLE A1 (contd.)

1	2	3	4	5	6	7	8	9	10
51	27	8.00	7.55	6.00	4.20	1.01	18.75	4.50	0.0
52	28	8.17	6.40	7.10	4.30	1.01	18.00	0.10	0.5
53	29	8.12	3.65	6.50	4.30	1.03	17.75	4.40	0.0
54	30	8.00	4.95	7.00	3.80	0.94	18.75	5.70	2.7
55	31	7.49	3.55	6.00	4.30	1.07	18.50	1.60	4.1
56	Aug. 1	7.85	6.05	5.80	3.35	1.09	18.00	1.50	28.5
57	2	7.24	9.80	6.60	4.95	2.28	17.50	0.90	0.0
58	3	7.52	5.55	5.80	5.15	2.07	17.00	5.80	0.0
59	4	7.69	6.20	6.90	3.60	1.46	17.00	1.00	13.0
60	5	8.62	3.85	7.80	3.60	1.23	17.00	5.50	0.0
61	6	8.57	3.40	7.70	3.25	1.17	17.25	1.40	6.3
62	7	8.53	4.40	7.50	2.85	1.26	18.25	9.60	0.2
63	8	8.22	3.85	7.70	3.00	1.26	18.25	6.30	0.0
64	9	8.26	3.20	7.50	3.10	1.03	17.50	8.00	0.0
65	10	8.08	2.85	7.50	3.70	1.01	18.00	9.70	0.9
66	11	8.22	3.00	7.80	3.45	0.98	18.00	10.80	0.0
67	12	8.62	3.40	8.10	4.20	1.01	18.75	5.90	0.0
68	13	8.40	3.20	7.30	4.20	1.05	19.00	4.00	0.0
69	14	8.98	2.95	7.60	3.65	1.00	19.00	7.50	0.3
70	15	8.54	3.40	6.80	4.25	0.96	19.00	1.10	0.0
71	16	8.18	3.95	6.50	5.00	0.93	18.50	8.00	0.1
72	17	8.80	2.70	6.50	4.20	1.03	18.00	4.70	0.0
73	18	8.99	2.85	7.40	5.70	1.01	17.00	4.50	0.0
74	19	8.76	3.60	7.90	5.75	0.96	16.25	10.70	0.0
75	20	9.03	3.55	8.10	4.00	0.87	17.25	0.50	0.0
76	21	9.18	5.15	7.70	4.80	0.98	17.25	12.20	0.0
77	22	8.62	3.50	8.60	4.60	1.00	17.75	3.90	0.0
78	23	7.79	4.90	8.20	4.40	0.98	17.50	10.20	0.0
79	24	7.70	5.20	8.60	4.75	0.93	18.00	10.60	0.0
80	25	8.76	5.20	8.70	5.00	0.91	18.25	8.40	0.0

Additional field data

These data are available on request and consist of the following:

- Three-hourly sampled measurements of upstream and downstream DO concentration, and stream temperature, starting at 12.00 hours on June 6th and finishing at 18.00 hours on August 25th.
- Actual sampling times for upstream and downstream BOD concentrations, corrected to the nearest three-hourly sampling instant.
- Once-daily averaged values for the volumetric flow rate and BOD concentration of the sewage works discharge over the experimental period.

APPENDIX 2

Cross-Sectional Dimensions of the River Cam

Table A2 gives (approximately) rectangular cross-sectional dimensions for the channel of the River Cam at roughly 200m intervals downstream from the upstream reach boundary. Two sets of dimensions at the downstream boundary are provided since at this point the channel divides to allow the main discharge to pass over the weir, while a second channel is used for navigation through a lock. The figures in parentheses denote the channel leading to the weir; the other figures represent a cross-section approximating the dimensions of the two channels combined.

Note that with respect to the reach length the figure of 4.7km has become enshrined in earlier publications; the correct figure is the one given in Table A2. Note also that for modelling purposes, i.e. in Table 3, the value of $1.51 \times 10^5 m^3$ has been substituted for the volume of water held in the reach; on the basis of the figures of Table A2, the volumetric hold-up of water is calculated as $1.48 \times 10^5 m^3$. This discrepancy arises because the more accurate details of Table A2 were not computed until after a more comprehensive experiment on the Cam in 1975. All dimensions in Table A2 are, of course, subject to the assumption of a nominal head of water in the reach.

TABLE A2

Distance from upstream boundary (km)	Rectangular cross-section, breadth × depth (m)
0.000	22.9 × 1.33
0.101	19.7 × 1.33
0.302	22.4 × 1.05
0.503	18.0 × 1.62
0.704	17.4 × 1.43
0.905	18.9 × 1.52
1.107	20.4 × 1.52
1.308	18.0 × 1.33
1.509	19.2 × 1.52
1.710	18.1 × 1.43
1.912	17.5 × 1.62
2.113	18.8 × 1.62
2.314	19.5 × 1.52
2.515	20.7 × 1.33
2.716	22.1 × 1.43
2.918	21.3 × 1.43
3.119	27.6 × 1.52
3.320	19.2 × 1.62
3.521	24.5 × 1.43
3.723	27.7 × 1.62
3.924	29.6 × 1.33
4.125	28.0 × 1.43
4.326	32.0 × 1.43
4.527	50.4 × 1.29
(4.527	24.2 × 1.05)

APPENDIX 3

Notes on Previously Published Works and Some Unpublished Work

The purpose of this Appendix is to supplement the brief analysis of the models and modelling results of the paper by summarizing, in abstract form, the details of previously publsihed works and some unpublished studies on the Cam (1972) experiment. The order of the articles listed follows the development of the subject in preference to the chronological development of the models.

1. "The Modelling of Dissolved Oxygen in a Non-Tidal Stream"
   (Beck, 1978a)

This article (written in 1975) gives a review of DO-BOD interaction models as they have evolved from the classical studies of Streeter and Phelps (1925). The article is somewhat restricted in terms of a literature review since it focuses attention on a sanitary engineering approach to water quality modelling; it is thus lacking in its treatment of similar lines of investigation originating from the point of view of ecology. Such ecological models as are available, however, have been generally applied to large estuarine systems rather than to smaller freshwater rivers.

2. "A Dynamic Model for DO-BOD Relationships in a Non-Tidal Stream"
   (Beck and Young, 1975)

The purpose of the paper is to present the arguments leading to the formulation of Model II (section 4.2) and to show how this model gives a better fit to the experimental field data than the a priori model, Model I (section 4.1). An interpretation of the sustained sunlight effect might be as follows. Sewage effluent entering the river just upstream of the experimental system creates a nutrient-rich environment in which populations of algae may expand rapidly to significant proportions under the stimulus of longer periods of warm, sunny weather. Model verification and parameter estimation are treated in a purely deterministic framework as a matter of repeated "trial and error" simulation comparisons with the field data.

3. "Systematic Identification of DO-BOD Model Structure"
   (Beck and Young, 1976)

This paper is the statistical counterpart of paper 2 above. It concentrates on the technical problem of model structure identification as defined within the overall context of system identification and parameter estimation. Model structure identification is the process of establishing, by reference to the field data, that the model includes all the significant physical, chemical and biological relationships between variables; and further, that these relationships have the correct form, for example, the form of first-order linear growth kinetics or the form of Monod growth kinetics. (The next stage of analysis, parameter estimation, would then attempt, for example, to derive accurate estimates of either the linear growth rate constant or the saturation constant and maximum specific growth rate constant of the Monod function.) The particular method employed to solve the model structure identification problem in this paper is the Extended Kalman Filter (EKF). In fact there is no easy way of identifying a correct structure for internally descriptive dynamic models and success in the application of the EFK depends strongly on a reasonable a priori knowledge of the model parameter values, such as those estimates obtained in paper 2.

4. "The Identification of Algal Population Dynamics in a Freshwater Stream" (Beck, 1975)

This paper describes the results of applying Maximum Likelihood (ML) parameter estimation to Model III, an internally descriptive model, and to the black box models Va and Vc. Using arguments parallel to those for the description of micro-organism cultures in the wastewater treatment processes of activated sludge and anaerobic digestion, the paper brings together both empirical evidence and the identification/estimation results for the synthesis of Model IV. Model IV itself is verified, and its parameter values are estimated, by simple trial and error deterministic simulation methods. The paper links together Model II (from papers 2 and 3) and Model IV and shows how this can be achieved through the analysis of Models III, Va and Vc.

5. "Maximum Likelihood Identification Applied to DO-BOD-Algae Models for a Freshwater Stream" (Beck, 1974)

The report amplifies the ML estimation results of paper 2; a summary of these results is given in Appendix 5 for comparison with the equivalent Instrumental Variable (IV) - Approximate Maximum Likelihood (AML) parameter estimates.

6. "Random Signal Analysis in an Environmental Sciences Problem" (Beck, 1978b)

This and the following paper are, to some extent, more concerned with the subjects of modelling, system identification, and parameter estimation, than with the subject of DO-BOD-algae interaction. They thus exploit the Cam (1972) modelling exercise as a means for making statements on these broader issues. Paper 6 emphasises the interpretation of modelling - or more strictly speaking, model structure identification - as a procedure of repeated hypothesis testing and decision making. In other words, any given model is a working hypothesis, the validity of which should ideally be tested against experimental field data.

Having carried out the test, the systems analyst is then required to decide whether the hypothesis is adequate; and if the current hypothesis is inadequate, according to the criteria of the analyst, a subsequent hypothesis must be generated and also evaluated by reference to the data. In the light of this interpretation of modelling, the paper reviews the complete conceptual process of deriving Model IV from the starting point of Model I. The paper does not enter into any detailed discussion of parameter estimation methods, nor does it attempt to formalize the notions of hypothesis testing and decision making.

7. "Model Structure Identification from Experimental Data"
(Beck, 1978c)

A theme clearly emerging from the above papers is that model structure identification is a problem central to success or failure in the modelling of complex, or poorly defined, systems. This paper defines the context of model structure identification within the subject of system identification and parameter estimation--a subject which also includes the topics of experimental design, model verification, and model validation. The paper places considerable importance on the role of the EKF algorithms (see also paper 3) in model structure identification; hence the paper presents a fairly detailed statement of how to apply the algorithms and how to interpret the results thereby obtained. The paper complements the work reported in papers 3 and 6.

Notes on Diurnal Variations, Sedimentation, and Model Applications

A consideration of these items has been kept separate because throughout the paper attention has been directed towards field data and models which do not deal with either diurnal variations or the sedimentation of particulate material from the sewage works effluent.

Diurnal Variations

Some brief remarks on the inclusion of these effects in the models are given in paper 2. The observed features in the data can be summarized as follows:

- distinct patterns of diurnal variations in the downstream DO concentration become established after day t_{30} and continue uninterrupted until the end of the experiment, except for the two days succeeding the thunderstorm;
- prior to day t_{30} diurnal variations in the (downstream) DO are indistinct with an amplitude of probably little more than $\pm$ 0.25 gm^{-3};
- after day t_{30} the amplitude of the diurnal variations rises on occasion to a maximum value of $\pm$ 2.00 gm^{-3};
- at all times the diurnal variation in the upstream DO concentration where discernible, is significantly less than the variations observed downstream.

The phase of the downstream diurnal variations, or alternatively the timing of their peak values, shows curious changes over the experimental period and perhaps therefore deserves special mention. Figure A3.1 shows a plot of the approximate intervals of the experiment during which the peak of the diurnal oscillations occurred at more or less the same time in each successive day. The only explanation offered for Figure A3.1 is that the downstream dissolved oxygen concentration reflects a complex balance between the phase of the algal photosynthetic/respiratory cycle and the phase of diurnal variations in BOD loadings imposed on the stream by the sewage works discharge. For the first half of the experiment, when the river flow-rate is steadily decreasing, the phase of the (transported) BOD loadings at the downstream boundary can be expected to change proportionately. Indeed, over this initial period the effects of algae are not dominant, whereas the mean detention time of the reach, calculated on the basis of mean volumetric hold-up and the stream discharges of Appendix 1, is seen to vary from 0.9 days to 1.5 days. This is a total change of 0.6 days (14 hours) in the detention time of the reach, which, assuming a constant phase for the sewage discharge, should effect an equivalent change in the phase of the downstream DO diurnal variations, as demonstrated by Figure A3.1. For the second half of the experiment stream discharge is approximately constant, apart from the thunderstorm, and in any case the effects of algae are expected to dominate, which implies that the phase of diurnal oscillations should be roughly constant.

Sedimentation

In order to evaluate a model which predicts downstream variations on the basis of the sewage works effluent quality, certain strong assumptions have to be made concerning the upstream DO and BOD concentrations at the effluent outfall. These assumptions may not be so stringent in practice, however, because the effluent BOD tends to dominate upstream BOD conditions, while the upstream weir dominates the DO conditions sufficiently for the values of u_1 to be substituted for the pattern of stream DO variations at the effluent outfall. Given such assumptions, the analysis reveals the following two salient features:

- that the predicted downstream BOD concentrations considerably over-estimate the observed concentrations for the period $t_0 \rightarrow t_{13}$;
- that in general the model gives higher downstream BOD concentrations than expected.

We have already alluded to the first point both in section 6 of the paper and in Appendix 1; it is a consequence not so much of error in the model, but of error in the BOD measurements. If a correction for the second point is hypothesized as an increased rate constant for BOD decay, a_3, such a correction substantially degrades the performance of the model in its predictions of the downstream DO levels. On the other hand, if the constant a_3, is divided into two parts conceptually, that is a_{31} and a_{32} say, where a_{31} (effectively the decay rate constant) fulfils the role of a_3 in the DO equations of Models I through IV, and where a_{32} enters the corresponding BOD equations as a term representing sedimentation, it is possible to improve the model's BOD performance without degrading its DO performance. Suitable values for a_{31} and a_{32} are found to be 0.32 day^{-1} and 0.16 day^{-1} respectively, so that to all intents and purposes the degree of DO-BOD interaction is preserved as for the models of the paper, but that a portion of the sewage works effluent BOD, presumably that portion attached to particulate matter, settles on to the river bed. Certainly the proposal that sedimentation is significant in the short stretch of river between the effluent and the upstream weir, but is not significant below the weir, seems plausible.

Model Applications

In section 3.1 it is stated that the originally intended application of the DO-BOD interaction models was to be in the synthesis of automatic, on-line control schemes for the day-to-day maintenance of stream DO levels. Thus, apart from the more realistic nature of a dynamic model as a description of a system which is rarely at a true steady state, the character of the models presented in the paper is aimed primarily at operational, and not design/planning, water quality management and control. (Even so, this does not necessarily preclude the use of dynamic models in the planning phases of river basin management as demonstrated by Whitehead (1976).)

There are at least three ways in which one can attempt to control the DO concentration at some point in the river system downstream of an effluent outfall. The first two of these three ways both view the problem of DO control as a problem of manipulating the BOD loading placed on the receiving river by the sewage discharge: (i) either one regulates the degree of BOD removal from the raw sewage, or (ii) one regulates the rate of treated sewage discharge to the stream by employing a post-treatment detention lagoon. Simulation results with Model I for case (i) and with Model II for case (ii) applied to the Cam (1972) data are reported in Young and Beck (1974). Clearly there are a number of assumptions implicit in these studies which are not valid in practice. Among the most important technical constraints on this kind of operational control are that the degree of BOD removed from sewage cannot be varied at will from one day to the next; that the required instrumentation, telemetry/communications networks are costly, or do not exist, and that there may not be sufficient land available for the construction of a large post-treatment lagoon.

The third form of DO control, namely artificial instream aeration, is attractive for the very reason that it seems more immediately practicable. Whitehead (1977), for example, discusses such an operational control scheme using the Cam (1972) data to demonstrate his results. His model, however, while being similar in some senses to Model II, is yet substantially different from all the models presented in the text; a full report of Whitehead's dynamic model for the Cam can be found in Young and Whitehead (1977).

APPENDIX 4

A Method for Time-Variable Transportation Delay Simulation

This Appendix describes a method of time-variable transportation delay simulation proposed by Coggan and Noton (1970); in fact it is worth noting that Coggan and Noton incorporate this form of simulation in an application of the same Extended Kalman Filtering algorithms that are used for analysis of the Cam (1972) data (Beck and Young, 1976).

The essential concepts behind the simulation are that the transportation delay element of Figure 3(b) can itself be imagined as a combination of n_a, say, fixed length (time-invariant) transportation delays and n_b, say, CSTR's in series (see Figure A4.1). The purpose of the time-invariant transportation delay section is to simulate the minimum expected transportation delay through the reach of river. (Recall that the term "transportation delay" denotes the time taken before any response is detected downstream as a consequence of any change in the upstream substance concentration.) The purpose of the multiple CSTR's is to simulate "flexibility" in the total transportation delay, $\tau(t)$, as it varies between the minimum, τ_{min}, and maximum, τ_{max}, expected values for the given stretch of river. Precisely how the numbers of elements n_a and n_b are chosen will be discussed below.

Suppose that we have as input to the first discrete-time delay element a concentration of (conservative) substance, $z(t)$, and that as output from the last CSTR element a concentration of that same substance, $z(t - \tau(t))$, where $\tau(t)$ is the time-variable transportation delay referred to in the main body of the paper. (Recall also that by Assumption 2 in section 3.1 it has been assumed that materials flowing through the transportation delay behave as conservative substances.) The simulation of the total transportation delay may then be represented by,

$$\text{(a)} \quad \begin{cases} z_1(t_j) = z(t_{j-1}) \\ z_2(t_j) = z_1(t_{j-1}) \\ \quad \vdots \\ z_{n_a}(t_j) = z_{n_a-1}(t_{j-1}) \end{cases} \qquad \text{(A4.1)}$$

$$
\text{(b)} \quad \begin{cases} \dot{z}_{n_a+1}(t) = - z_{n_a+1}(t) \,/\, T(t) + z_{n_a}(t) \,/\, T(t) \\ \dot{z}_{n_a+2}(t) = - z_{n_a+2}(t) \,/\, T(t) + z_{n_a+1}(t) \,/\, T(t) \\ \quad \vdots \\ \dot{z}_{n_a+n_b}(t) = - z_{n_a+n_b}(t) \,/\, T(t) + z_{n_a+n_b-1}(t) \,/\, T(t) \end{cases} \qquad \text{(A4.1) contd.}
$$

with

$$z(t - \tau(t)) = z_{n_a+n_b}(t)$$

where $z_i(t_j)$, $i = 1, 2, \ldots, n_a$, is the output of the ith discrete-time delay element and $z_i(t_j)$, $i = n_a + 1, n_a + 2, \ldots, n_a + n_b$, is the output of the $(i - n_a)$th CSTR element. The notation of t_j to represent discrete-time instants draws a distinction between t_j and the discrete-time notation t_k of the paper. This is because the length of each time-invariant transportation delay element in the above simulation of equations (A4.1(a)), that is $\delta t = (t_j - t_{j-1})$, may, or may not, be equivalent to the sampling interval $\Delta t = (t_k - t_{k-1})$ of the measured field data. In order to match the solutions of equations (A4.1(a)) at the instants t_j, equations (A4.1(b)) are integrated over the intervals $t_{j-1} \rightarrow t_j$. Thus note how equations (A4.1(a)) are connected through z_{n_a} to equations (A4.1(b)); one would therefore expect the substitution of,

$$z_{n_a}(t) = z_{n_a}(t_j) \text{ for } t_j \leq t \leq t_{j+1} \qquad \text{(A4.2)}$$

Notice further that the time-constant (or mean residence time) for each CSTR element, $T(t)$, is time-varying; with respect to models II and III of the text, a low-pass filter is the same concept as a CSTR. The variability of $T(t)$ is where the necessary flexibility appears in the simulation, since for a long transportation delay $T(t)$ should be large, i.e. giving a slow response, and for a short transportation delay $T(t)$ should be very small, i.e. giving a fast response. As with $z_{n_a}(t)$, for computational purposes,

$$T(t) = T(t_j) \text{ for } t_j \leq t \leq t_{j+1} \quad (A4.3)$$

Looking at equation (A4.1), there are several choices to be made in order to implement the simulation, and these concern

- the integration time-steps $(t_j - t_{j-1})$;
- the number of elements n_a and n_b;
- the specification of $T(t)$ for the CSTR elements;
- the computation of $\tau(t)$.

Bearing in mind the use of the overall DO-BOD interaction model to compare model predictions with observations at time t_k, it is sensible to choose the integration time-step such that the sampling interval of the data is some integer multiple, d, of this time-step,

$$\Delta t = d.\delta t \quad (A4.4)$$

Thus having defined δt by the choice of d, n_a can be chosen as

$$n_a = \text{int pt } [\tau_{min}/\delta t] \quad (A4.5)$$

in which int pt [·] means the integer part of the ratio between the minimum expected transportation delay, τ_{min}, and the integration time-step. Similarly, n_b can be chosen according to (Coggan and Noton, 1970),

$$n_b = \text{int pt } [0.5(\tau_{max}/\delta t - n_a)] + 1 - \text{int pt } [n_a \delta t/\tau_{max}] \quad (A4.6)$$

where τ_{max} is the expected maximum transportation delay in the reach of river. The time-constant of the CSTR elements is calculated on the basis of

$$T(t) = (\tau(t) - n_a \delta t)/n_b \quad (A4.7)$$

subject to the condition that

$$n_a \delta t < \tau_{min} \quad (A4.8)$$

which ensures stability of the simulation, i.e. $T(t) > 0$ for all t. Finally, a simple, but heuristic means of computing $\tau(t)$ is

$$\tau(t) = \tau_{min} + \left(\frac{\Theta_{1min} - \Theta_{1max}}{\Theta_1(t) - \Theta_{1max}}\right)(\tau_{max} - \tau_{min}) \tag{A4.9}$$

in which Θ_{1min} and Θ_{1max} are the stream discharges corresponding to minimum and maximum values of the transportation delay, τ_{min} and τ_{max} respectively, and where $\tau(t)$ is to be always smaller than the mean residence time of the reach, i.e.

$$\tau(t) < (a_1/\Theta_1(t)) \tag{A4.10}$$

An Example Simulation for the Case Study

From the data of Appendix 1 and for a value of 1.51 x $10^5 m^3$ for the volumetric hold-up of water in the reach, we have:

Minimum mean residence time = 0.66 day for Θ_{1min} = 2.28 x $10^5 m^3 day^{-1}$

Maximum mean residence time = 1.74 day for Θ_{1max} = 0.87 x $10^5 m^3 day^{-1}$

If we choose τ_{min} = 0.55 day and τ_{max} = 1.5 day, and d = 2 in equation (A4.4), then

$$\delta t = 0.5 \text{ day}$$

and by equation (A4.5), $n_a = 1$. Substituting for these figures in equation (A4.6) gives $n_b = 2$.

Since both the upstream DO and BOD concentrations must be modified by a transportation delay simulation, we have for Model I of the paper the following combination of equation (14) and equations (A4.1),

(a)
$$\dot{x}_1(t) = (\Theta_1(t)/a_1)z_3(t) - (\Theta_1(t)/a_1)x_1(t) + a_2(\Theta_3(t) - x_1(t)) - a_3x_2(t) + a_4(t)$$

(b)
$$\dot{x}_2(t) = (\Theta_1(t)/a_1)z_6(t) - (\Theta_1(t)/a_1)x_2(t) - a_3x_2(t) + a_5(t)$$

(c)
$$\begin{cases} z_1(t_j) = u_1(t_{j-1}) \\ \dot{z}_2(t) = -z_2(t)/T(t) + z_1(t)/T(t) \\ \dot{z}_3(t) = -z_3(t)/T(t) + z_2(t)/T(t) \end{cases} \tag{A4.11}$$

(d)
$$\begin{cases} z_4(t_j) = u_2(t_{j-1}) \\ \dot{z}_5(t) = -z_5(t)/T(t) + z_4(t)/T(t) \\ \dot{z}_6(t) = -z_6(t)/T(t) + z_5(t)/T(t) \end{cases} \qquad \text{(A4.11) contd.}$$

The differential-difference equations (A4.11) are solved by integration over the interval $t_j \rightarrow t_{j+1}$ with the substitutions

$$\begin{cases} \underline{u}(t_j) = \underline{u}(t_k) \text{ for } t_k \leq t_j \leq t_{k+1} \\ \underline{\Theta}(t) = \underline{\Theta}(t_k) \text{ for } t_k \leq t \leq t_{k+1} \end{cases} \qquad \text{(A4.12)}$$

in line with the interpolation scheme used elsewhere in the paper. It might, in this instance, however, be more appropriate to make a linear interpolation for $\underline{u}(t_j)$ and $\underline{\Theta}(t)$. In equation (A4.11), equation (A4.11(c)) denotes the transportation delay simulation for upstream DO concentrations, and equation (A4.11(d)) the same for upstream BOD concentrations.

APPENDIX 5

Some Additional Error Statistics and Parameter Estimation Results

A different form of model error sequence to that given in equation (12) can be computed. For the internally descriptive model this involves imbedding the model, equations (4) and (5), within a Kalman filter formulation from which can be generated the (innovations process) residual errors,

$$\underline{\varepsilon}'(t_k|t_{k-1}) = \underline{y}(t_k) - \underline{\hat{x}}'(t_k|t_{k-1}) \tag{A5.1}$$

Here the one-step-ahead predictions $\underline{\hat{x}}'(t_k|t_{k-1}) = [\hat{x}_1'(t_k|t_{k-1}),\hat{x}_2'(t_k|t_{k-1})]$ of downstream DO and BOD concentrations are distinctly different from the predictions defined by equation (10). Whereas the deterministic model predictions $\underline{\hat{x}}(t_k)$ utilize the measured information on $\underline{u}(t_k)$ and $\underline{\Theta}(t_k)$, the one-step-ahead predictions $\underline{\hat{x}}'(t_k|t_{k-1})$ at time t_k utilize <u>in addition</u> the measured output data up to and including $\underline{y}(t_{k-1})$. The error sequences $\underline{\varepsilon}'(t_k|t_{k-1})$ are alternatively termed the one-step-ahead prediction errors. In practice, the statistical properties of these errors are dependent upon certain assumptions about the statistical properties and variance-covariances of the random processes $\underline{\xi}(t)$ and $\underline{\eta}(t_k)$ in equations (4) and (5). For this reason a comparison of different model performances on the basis of such an error criterion is not necessarily as straightforward as it might seem at first sight.

Similarly, one-step-ahead prediction error sequences can be calculated for the black box model, equations (6) and (7). Upon substituting for,

$$x(t_k) = y(t_k) - v(t_k) \tag{A5.2}$$

from equation (7) in the right-hand side of equation (6) we obtain

$$x(t_k) = A(q^{-1})y(t_k) + \sum_{i=1}^{m} B_i(q^{-1})u_i(t_k) - A(q^{-1})v(t_k) \tag{A5.3}$$

Now suppose that in general the lumped, coloured noise sequence $v(t_k)$ can be modelled as the following transformation of a white noise sequence $e(t_k)$, say,

$$v(t_k) = \frac{D(q^{-1})}{C(q^{-1})} e(t_k) \tag{A5.4}$$

where the additional polynomials are defined as

$$\begin{cases} C(q^{-1}) = 1 + c_1 q^{-1} + \dots + c_n q^{-n} \\ D(q^{-1}) = 1 + d_1 q^{-1} + \dots + d_n q^{-n} \end{cases} \tag{A5.5}$$

and where

$$E\{e(t_k)\} = 0 \text{ and } E\{e(t_k)e(t_j)\} = 0 \text{ for } k \neq j \tag{A5.6}$$

In equation (A5.6) $E\{\cdot\}$ is the expectation operator. The one-step-ahead prediction $\hat{x}'(t_k|t_{k-1})$ is then defined as

$$\hat{x}'(t_k|t_{k-1}) = A(q^{-1})y(t_k) + \sum_{i=1}^{m} B_i(q^{-1})u_i(t_k) - A(q^{-1})\hat{v}(t_k|t_{k-1}) \tag{A5.7}$$

or

$$\hat{x}'(t_k|t_{k-1}) = A(q^{-1})y(t_k) + \sum_{i=1}^{m} B_i(q^{-1})u_i(t_k) - \frac{A(q^{-1})D(q^{-1})}{C(q^{-1})}\,\varepsilon'(t_k|t_{k-1}) \tag{A5.8}$$

with the one-step-ahead prediction errors,

$$\varepsilon'(t_k|t_{k-1}) = y(t_k) - \hat{x}'(t_k|t_{k-1}) \tag{A5.9}$$

A comparison of equation (A5.7) with equation (11) shows clearly how the determination model predictions and the one-step-ahead predictions differ in their utilization of the measured output information $y(t_k)$. The one-step-ahead prediction, equation (A5.8), also includes a term which incorporates a function of the one-step-ahead prediction errors. In fact, recalling the definition of $A(q^{-1})$ in equation (9), equation (A5.8) defines the one-step-ahead prediction $\hat{x}(t_k|t_{k-1})$ at time t_k to be a function of the measured output data and previous prediction errors up to and including $y(t_{k-1})$ and $\varepsilon'(t_k|t_{k-1})$. (Application of the backward shift operator, equation (8), to the one-step-ahead prediction errors gives $q^{-1}\{\varepsilon'(t_k|t_{k-1})\} = \varepsilon'(t_{k-1}|t_{k-2})$. Implicit in equation (A5.8) is the assumption that the best estimate of the noise sequence, $\hat{v}(t_k|t_{k-1})$, can be derived on the basis of the noise process model, equation (A5.4), with $\varepsilon'(t_k|t_{k-1})$ substituted as an approximation of $e(t_k)$. Thus the one-step-ahead prediction error sequences for the black box models are dependent upon the way in which the noise processes are characterized in any given model (see below).

Table A5.1 presents a summary of the one-step-ahead prediction error statistics for Models I through V. The salient features of Table A5.1, with respect to Table 10 of the text, is that all models give smaller one-step-ahead prediction error variances than the corresponding deterministic response error variances, except notably for BOD in Model IV; and that apart from the a priori model the statistics of the DO prediction errors are all rather similar.

Table A5.2 gives, for completeness, a comparison of the identified noise model structures and parameter estimates and estimation errors in the black box models when Maximum Likelihood (ML) and Instrumental Variable-Approximate Maximum Likelihood (IV-AML) estimators are used. For ML estimation the noise process model of equation (A5.4) is necessarily constrained as,

$$v(t_k) = \frac{D(q^{-1})}{[1 - A(q^{-1})]} e(t_k) \qquad \text{(A5.10)}$$

that is,

$$C(q^{-1}) \equiv [1 - A(q^{-1})] \qquad \text{(A5.11)}$$

and hence, for example, the parameter value $c_1 \equiv -\alpha_1$ is inserted where appropriate in Table A5.2. For reference purposes note also that the statistics of Table A5.1 for Models Va, Vb, Vc are those derived with an IV-AML estimator.

Since the ML estimation results of Table A5.2 are based on the modified data point $u_2(t_{57}) = 6.5(\text{gm}^{-3})$, it is probably for this reason alone that the ML and IV-AML estimates of β_{21} in Model Vc differ significantly. Indeed, given the relatively large estimation error for the IV-AML estimate of β_{21}, it is debatable whether the associated term $u_2(t_{k-1})$ should be included in the model structure. It must be admitted, however, that the method of computing parameter estimation errors for the IV-AML estimator is only approximate. Table A5.2

indicates one further significant difference between the IV-AML and ML results, as follows. For the ML Models Va and Vc the equivalent continuous-time first-order time constants for the DO and BOD dynamics are respectively 2.98 days and 3.49 days, i.e. closely similar. For the IV-AML Models Vb and Vc the two time constants are 1.93 days and 5.23 days respectively for the DO and BOD. On the assumption that the sunlight conditions, $u_3(t_k)$, are providing the primary input disturbances and that it is the responses to this input, as opposed to the upstream DO or BOD conditions, that the models are preferentially estimating, then one can conclude that the IV-AML estimated models confirm findings reported elsewhere (Beck, 1975, 1978b). This observation, namely that the downstream DO concentration responds more quickly than does the downstream BOD concentration to a change in sunlight conditions, is analagous to interpretations of the role of $u_3(t)$ in Models III and IV. However, one should perhaps not place too much emphasis on this sort of appraisal of black box model results since they may be no more meaningful than some spurious statistical property of the field data.

TABLE A5.1: Survey of one-step-ahead Prediction Error Statistics for all Models

	DO			BOD		
	μ_1' (gm^{-3})	σ_1' (gm^{-3})	$(\sigma_1')^2/\sigma_D^2$(%)	μ_2' (gm^{-3})	σ_2' (gm^{-3})	$(\sigma_2')^2/\sigma_B^2$(%)
Original data	7.282 (μ_D)	1.067 (σ_D)	-	4.112 (μ_B)	1.265 (σ_B)	-
One-step-ahead Prediction Errors						
I	0.236	0.702	43.3	0.638	1.085	73.6
II	0.076	0.558	27.3	0.243	0.822	42.2
III*	0.009	0.583	29.9	-0.032	0.768	36.9
IV	-0.072	0.577	29.2	-0.170	0.854	45.6
Va	0.008	0.542	25.8	-	-	-
Vb	0.001	0.524	24.1	-	-	-
Vc	-	-	-	0.003	0.628	24.6

* Results obtained with the substitution of $u_2(t_{57}) = 6.5$ (gm^{-3}) instead of the measured value of 9.8 (gm^{-3}).

TABLE A5.2: A Comparison of ML and IV-AML Estimation Results for Models Va, Vb, and Vc

Model	Parameter	IV-AML	ML
Va	α_1	0.639 ± 0.155	0.715 ± 0.064
	β_{11}	0.229 ± 0.120	0.174 ± 0.050
	β_{30}	0.062 ± 0.024	0.057 ± 0.016
	β_{31}	0.051 ± 0.027	0.044 ± 0.017
	c_1	-0.644 ± 0.036	(-0.715 ± 0.064)
	c_2	-0.139 ± 0.036	-
Residuals std. dev.	σ_1'	0.542	0.554
Vb	α_1	0.596 ± 0.123	
	β_{11}	0.261 ± 0.091	
	β_{30}'	0.060 ± 0.018	
	β_{31}'	0.052 ± 0.022	
	d_1	-0.641 ± 0.027	
Residuals std. dev.	σ_1'	0.524	
Vc	α_1	0.826 ± 0.078	0.751 ± 0.062
	β_{21}	0.054 ± 0.056	0.102 ± 0.042
	β_{32}	0.034 ± 0.019	0.048 ± 0.015
	β_{34}	0.057 ± 0.021	0.060 ± 0.020
	c_1	-0.311 ± 0.035	(-0.751 ± 0.062)
	c_2	-0.160 ± 0.035	-
	d_1	-	-0.520 ± 0.128
Residuals std. dev.	σ_2'	0.628	0.627

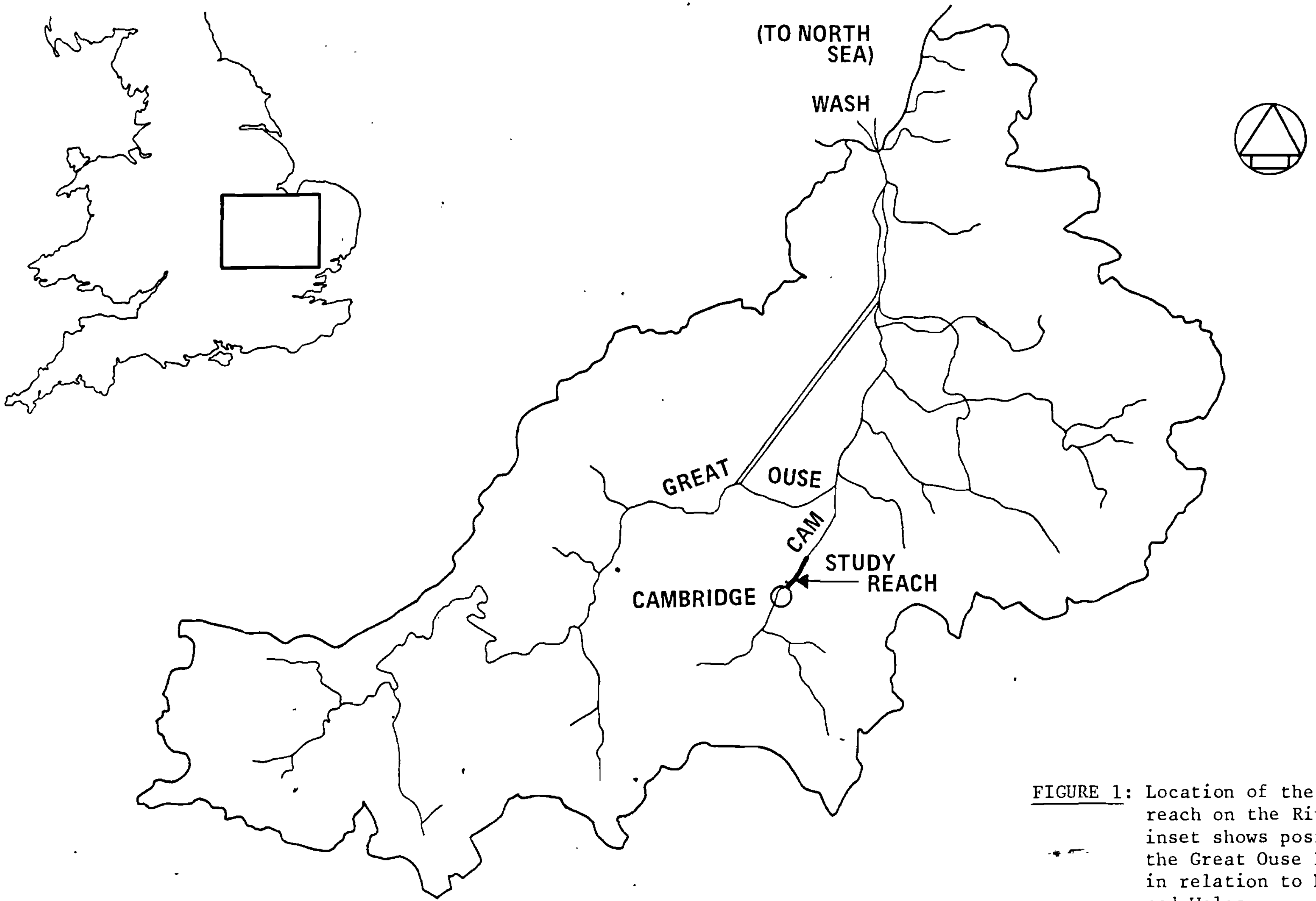

FIGURE 1: Location of the study reach on the River Cam; inset shows position of the Great Ouse Basin in relation to England and Wales.

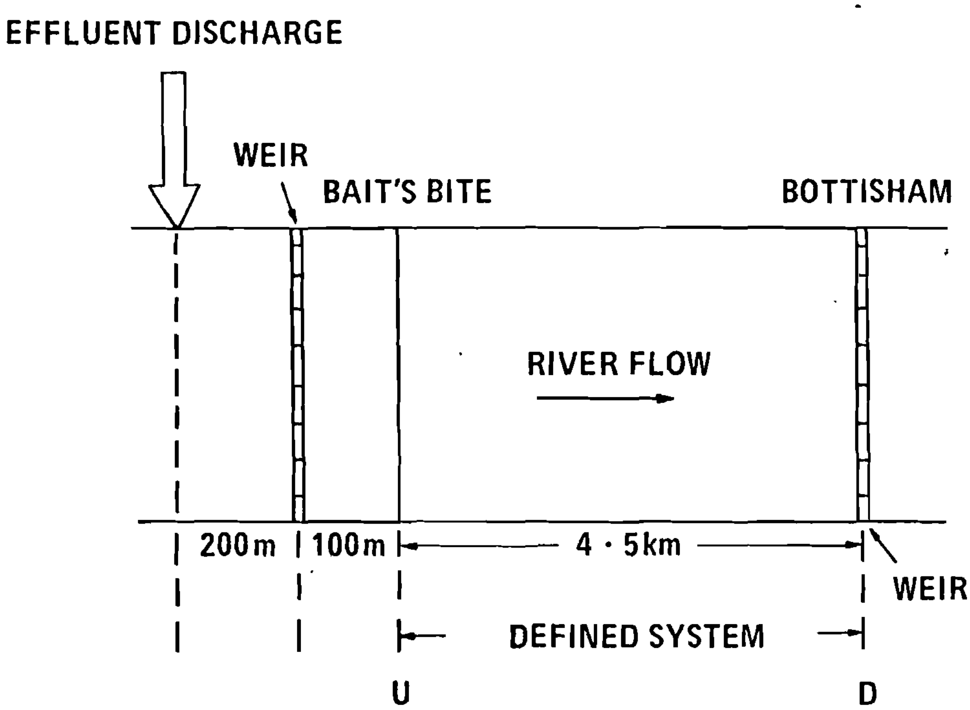

FIGURE 2: Schematic definition of the study reach showing the location of the effluent discharge from Cambridge Sewage Works.

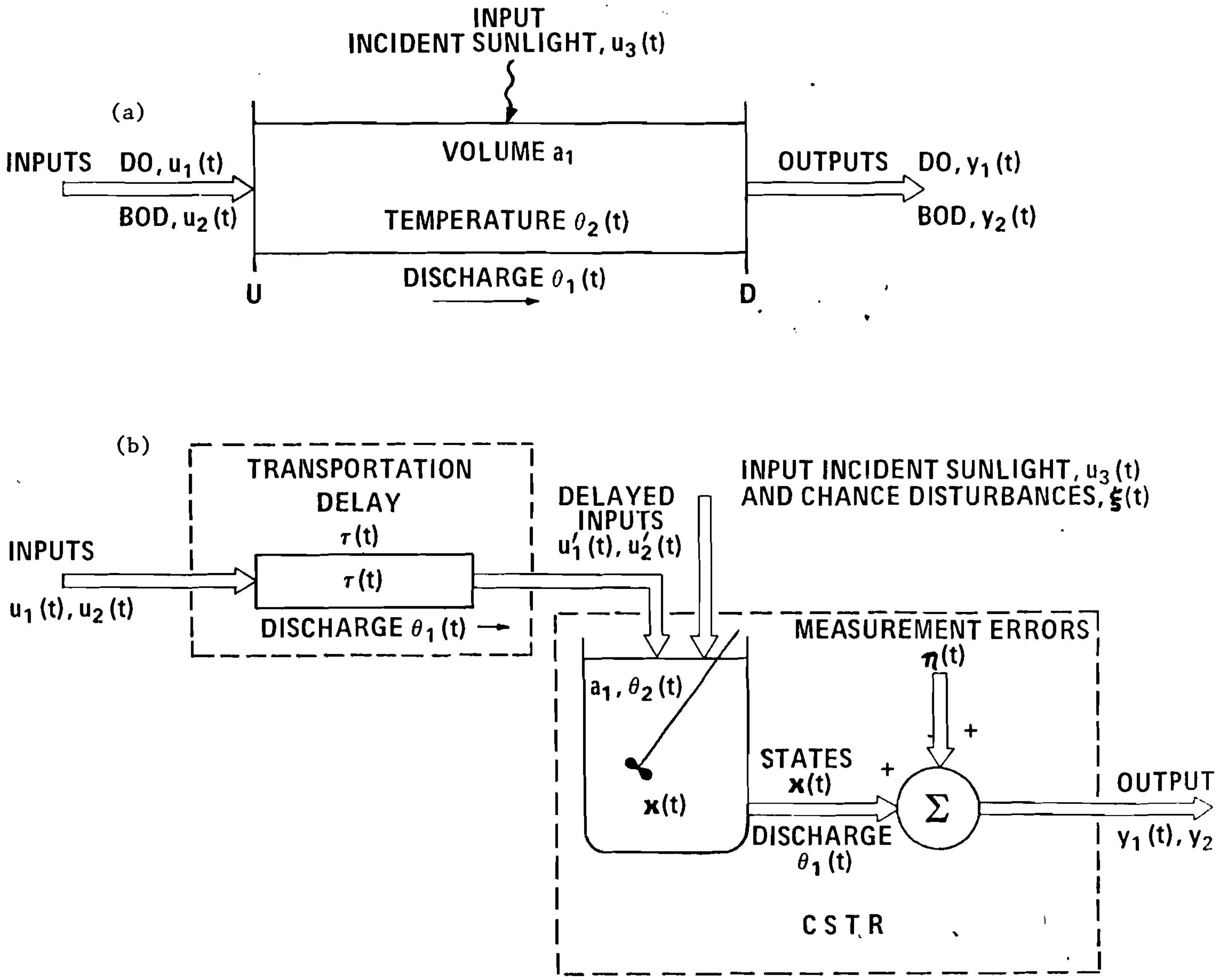

FIGURE 3: (a) Study reach of river with some notational conventions for the measured variables;
(b) transportation delay and continuously stirred tank reactor (CSTR) idealization of the reach of river.

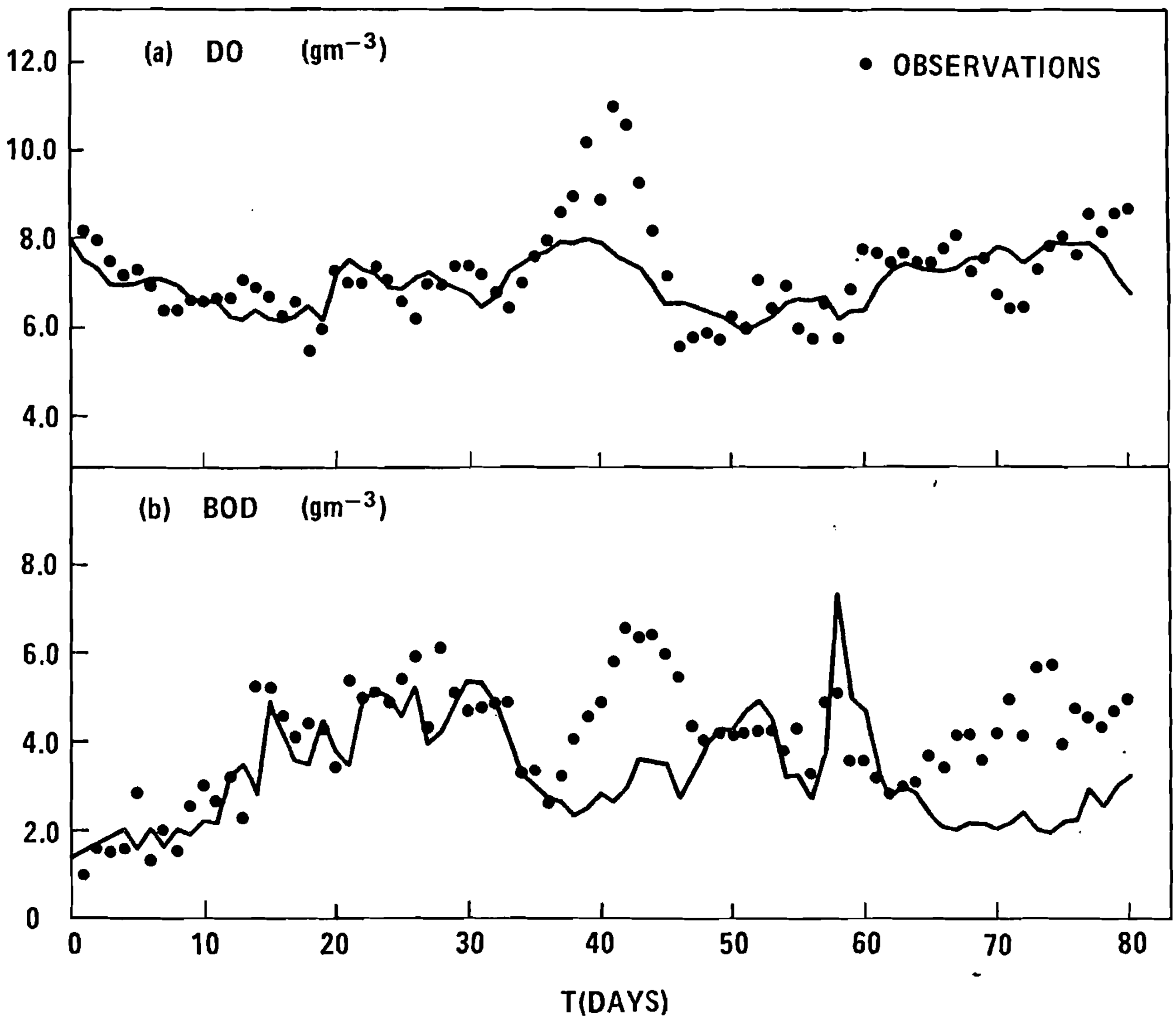

FIGURE 4: Deterministic model responses, $\hat{x}(t_k)$, and observations, $y(t_k)$, for model I:

(a) downstream DO concentration;
(b) downstream BOD concentration.

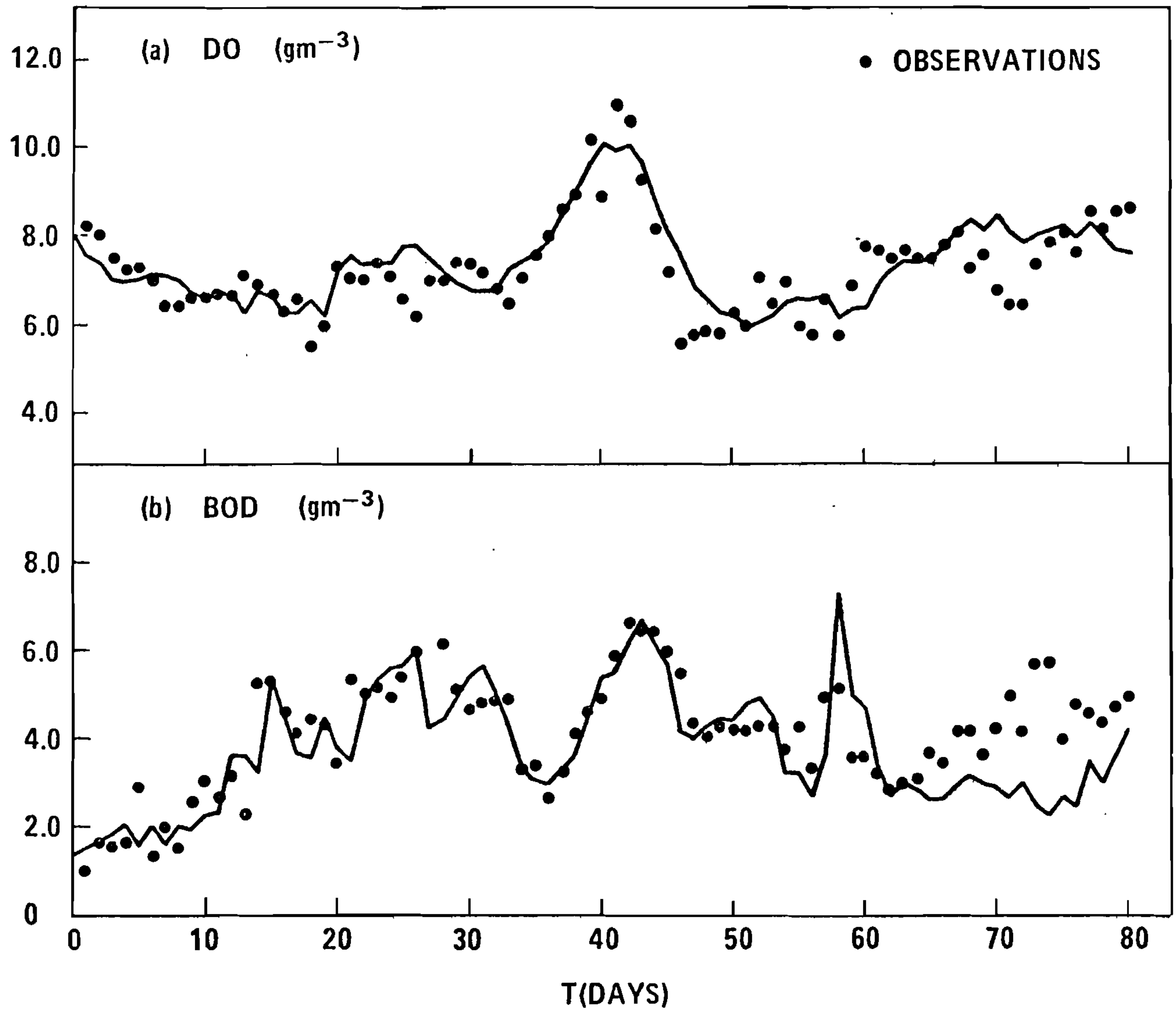

FIGURE 5: Deterministic model responses, $\hat{x}(t_k)$, and observations, $y(t_k)$, for Model II:

(a) downstream DO concentration;
(b) downstream BOD concentration.

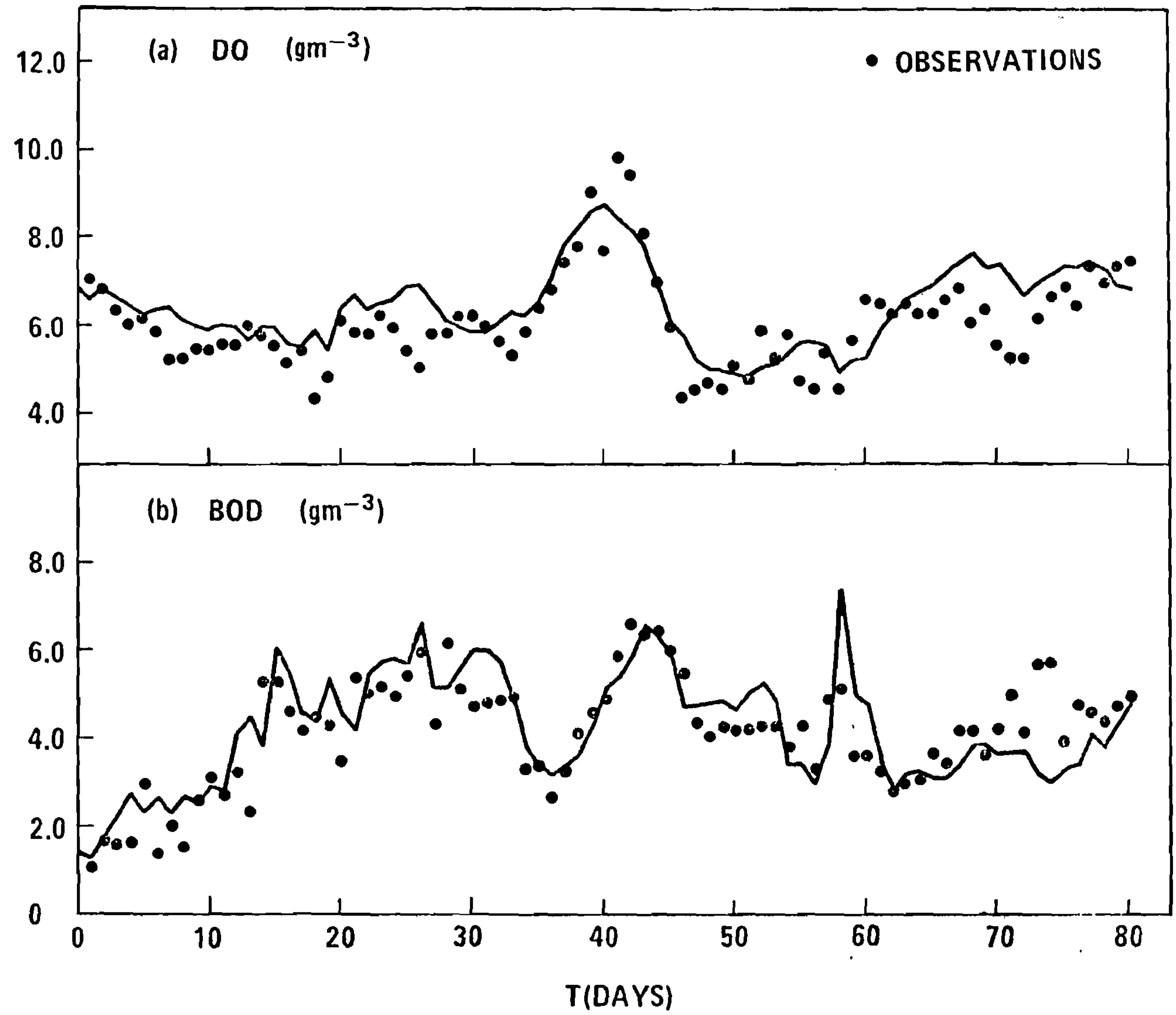

FIGURE 6: Deterministic model responses, $\hat{x}(t_k)$, and observations, $y(t_k)$, for Model III:

(a) downstream DO concentration;
(b) downstream BOD concentration.

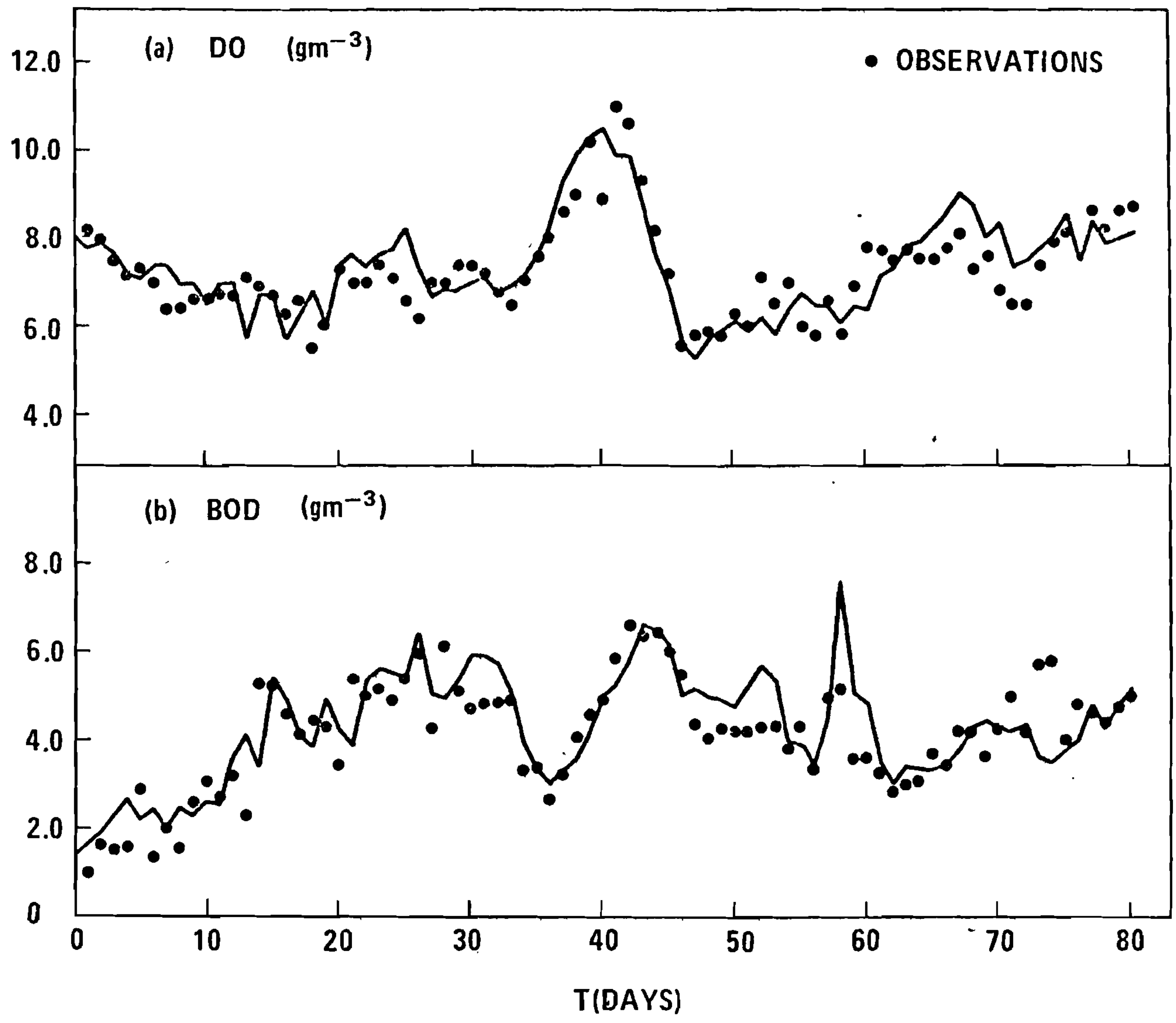

FIGURE 7: Deterministic model responses, $\hat{x}(t_k)$, and observations, $y(t_k)$, for Model IV:

(a) downstream DO concentration;
(b) downstream BOD concentration.

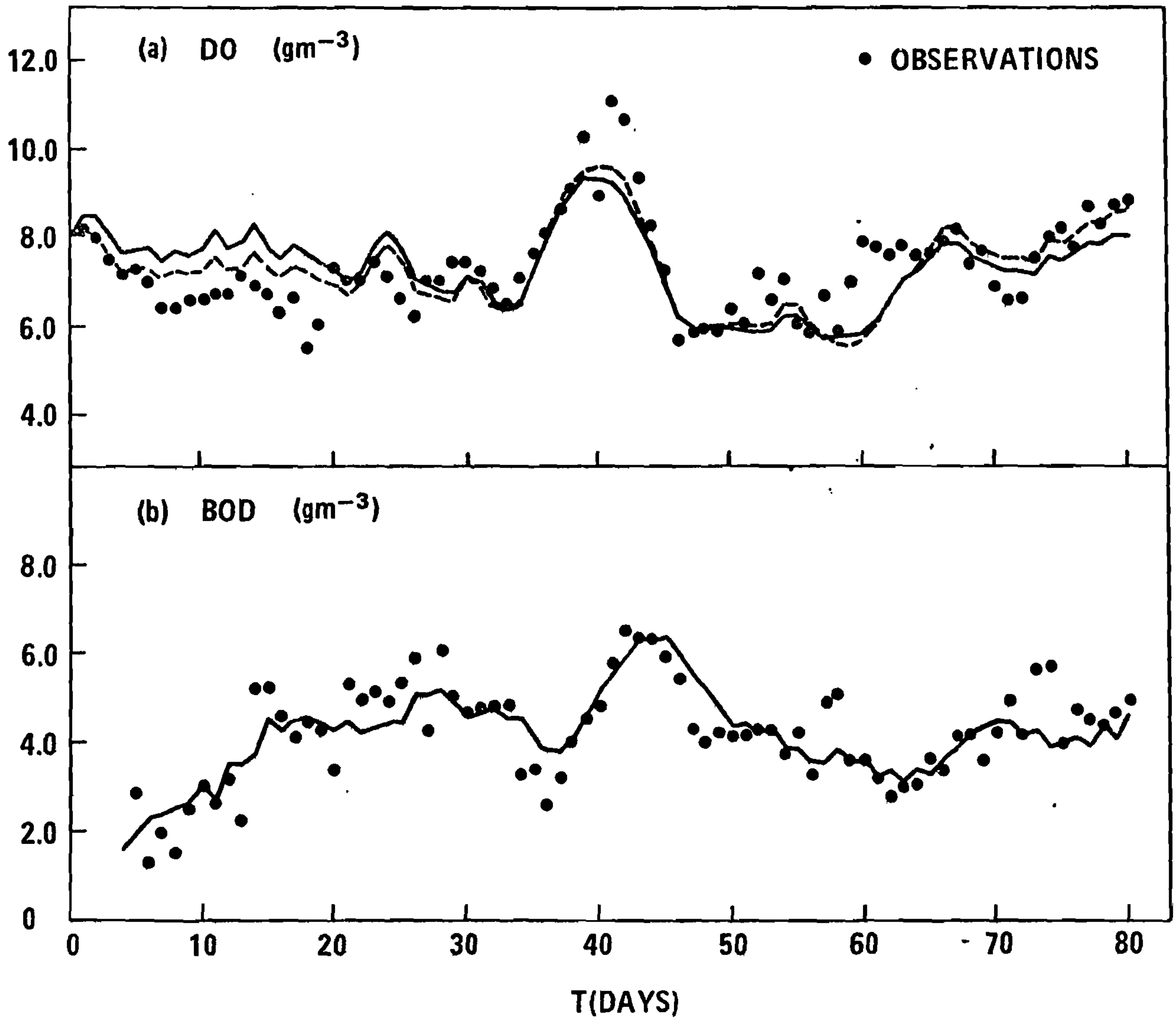

FIGURE 8: Deterministic model responses, $\hat{x}(t_k)$, and observations, $y(t_k)$, for:

(a) downstream DO concentration, Model Va - continuous line, Model Vb - dashed line;
(b) downstream BOD concentration, Model Vc.

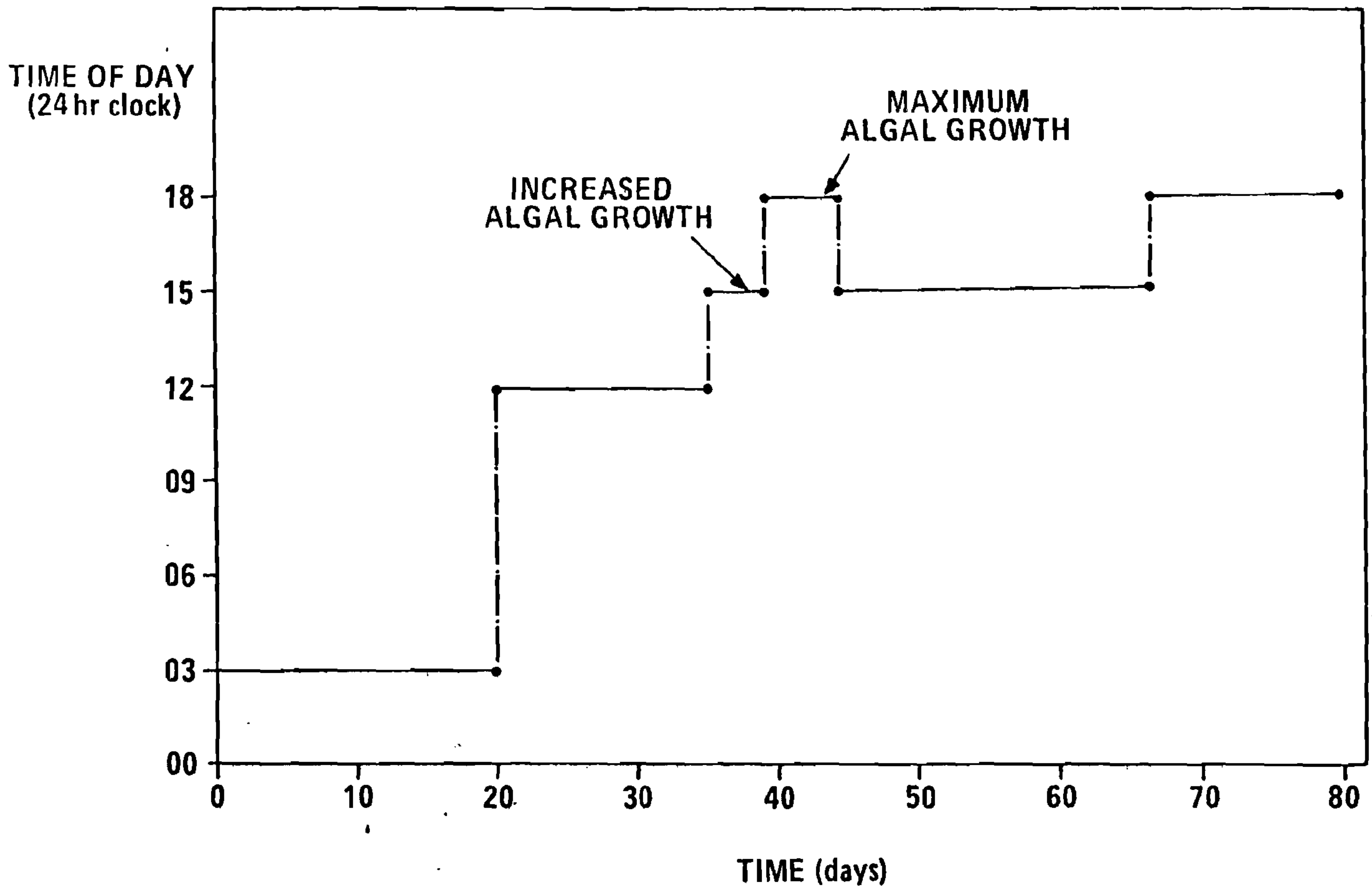

FIGURE A3.1: Timing of the peak diurnal DO concentration, on a 24 hr. clock basis, for various periods of the experiment.

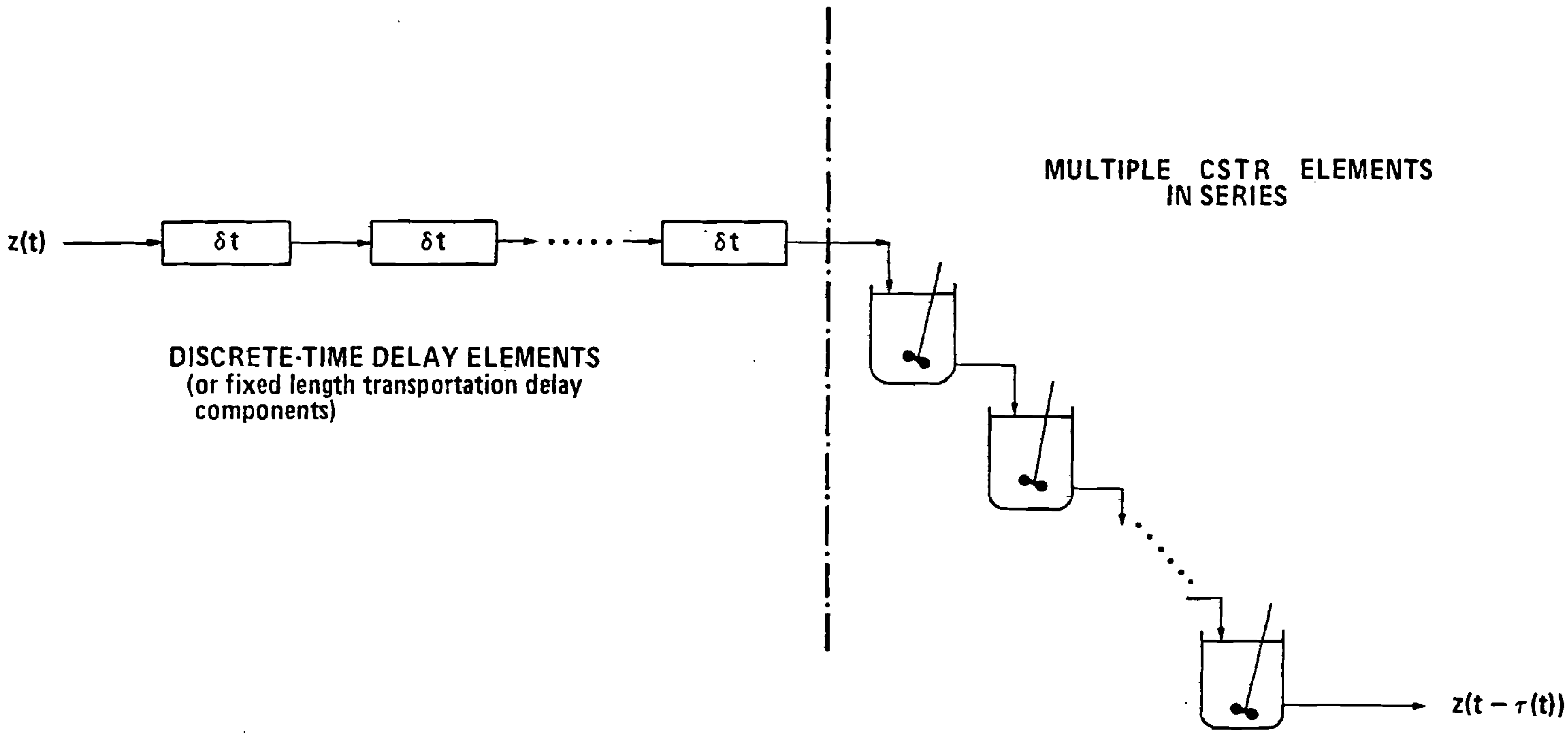

FIGURE A4.1: Conceptual representation of a simulation for a time-varying transportation delay.