TFSD MODELLING GROUP WORKSHOP

10-11 May 1989

Summary Report by

P G Whitehead
Institute of Hydrology
Wallingford

12 June 1989

SUMMARY REPORT

1. Introduction

At the request of TFS Directors a modelling workshop was held at the Institute of Hydrology on 10/11 May 1989. The aims of the workshop were

(a) to provide a forum for scientists within TFSD to present their modelling research;

(b) to provide information on modelling techniques available within TFSD;

(c) to exchange ideas on new techniques or new areas of model application;

(d) to identify areas of collaboration across institutes;

(e) to identify factors limiting the development and application of models.

2. Meeting Report

The response to the first letter of invitation to the meeting was extremely encouraging with over 25 modelling papers/presentations offered and interest from 50 TFSD scientists. The final programme (shown in Appendix 1) consisted of 19 presentations with 44 participants (see Appendix 2 for list of attendees). The 19 papers ranged from descriptions of detailed mechanistic models of processes in rivers, lakes, catchments and ecological systems through to system models that could be used for management purposes.

A common problem with any modelling meeting is to agree a common definition of modelling terms that can be used by all participants. Appendix 3 was provided for the meeting and contains a glossary of terms commonly used in mathematical modelling. This list should also assist non-modellers to become familiar with the jargon.

The meeting was considered very successful by the participants and fulfilled most of the objectives specified above. It was particularly interesting to hear of the modelling research being undertaken or contemplated at institutes and to have an opportunity to discuss ideas and problems both during the meeting and at an informal function in the evening of the 10 May. What was particularly striking was the different approaches in the institutes, the different level of skills and future scope for modelling.

As might be expected the Institute of Hydrology makes most use of models with models providing both a process investigation role to assess mechanisms and interactions and a management role for environmental impact assessment and for planning, design and operational purposes. The techniques and software developed are used to keep IH at the forefront of the science as well as generate significant contract income. There have been similar moves in this direction by IFE in recent years particularly in the development of lake flow and chemical models and aluminium speciation models. Also within ITE there is increasing use of models for studying forest growth, atmospheric deposition processes and ecological interactions. In many ways the IH lead in modelling is to be expected given the more quantitative nature of the research and the availability of physical laws and chemical equations that can be subjected relatively easily to modelling techniques. Biological systems are inevitably more difficult to model. Indeed, there has been a traditional view by qualitative biologists that the inherent instability and diversity of biological systems precludes a systems modelling approach. This encouragingly did not seem to be the view of TFSD biologists attending the meeting. There seemed to be a concensus that biological models linked to physico-chemical models were of value in determining controlling mechanisms and for predictive purposes. Some interesting comments on modelling biological systems and the response to the meeting are given in letters in Appendix 4.

3. Resources and Future of Modelling in TFSD

The final session of the workshop was devoted to resource problems and the future of modelling in TFSD. It was clear from the meeting that whilst at IH modelling was well established this was not the case at IFE or ITE. At both these institutes there are excellent modelling initiatives but there are simply not the resources or the skills to provide a broad modelling strategy for the future. This problem cannot be overcome without employing a different type of scientist to these institutes. It is difficult, if not impossible, to teach a scientist trained in a traditional biology/ecology discipline the techniques of modelling. Unfortunately there are no university courses that meet the TFSD requirements of biological/ecological modelling. Perhaps this needs to be raised with NERC higher education sections. Learning the techniques of modelling is far more than just running a computer based model because this is often easy - there are plenty of user-friendly models available. To really make use of models, however, it is necessary to understand the mathematical basis of them, the inherent dynamical behaviour, the interactions between model components and the limitations of the models. Thus it is necessary to recruit into ITE and IFE modellers with a mathematical, physics or engineering background who are capable of understanding the technical modelling aspects but who also can work closely with biologists and ecologists. The concept of multidisciplinary teams working on modelling is probably the best way forward for TFSD. To some extent this is already being pursued with inter-disciplinary groups being established at ITE Bush on deposition modelling and IFE have joined forces with University Groups such as Birmingham on lake modelling and IH on modelling Caesium transport. Within IH nearly all projects have built into them modellers with mathematical knowledge at the very outset to ensure that relevant data is collected for model calibration and validation and that dominant processes and mechanisms are identified and properly researched.

There are also new areas of research that make a modelling response by TFSD essential. For example, climate change studies will inevitably have a major modelling component since models are probably the only realistic method of assessing impacts. Many experiments designed to assess climate change process should also be viewed as a means of calibrating and validating models. There are other areas such as the programme on River Ecology in which models could be particularly useful (see Appendix 5 on models of Algal Growth in the River Thames).

4. Conclusions

The modelling workshop proved to be an informative, interesting and successful meeting with many new links established between institutes. The concensus of the meeting was that an annual meeting was not required but that a meeting every other year would be very beneficial and allow sufficient time for new modelling initiatives to have been developed and progressed.

APPENDIX 1

TFSD MODELLING WORKSHOP
to be held at the Institute of Hydrology, Wallingford
10-11 May 1989

Wednesday, 10 May

12.30	Arrival and Lunch at IH.	
13.30	Welcome by Mr Frank M Law, Head of Engineering Hydrology, IH	
13.45	Modelling Environmental Systems - An Overview - Paul Whitehead (IH)	
14.15	MAGIC - A model to assess long term trends in acidification	- Alan Jenkins (IH)
14.35	Problems of Modelling Wet and Dry Deposition in Upland Catchments	- Neil Cape (ITE)
14.55	FORTNITE - A model to assess forest growth and Forest Nutrient Cycling	- Phil Ineson (ITE)
15.15	Dynamic Modelling of Radionuclides in Upland Ecosystems	- Neil Crout (ITE)
15.35	Tea	
15.45	River Chemical Process Modelling	- Alan House (FBA)
16.05	QUASAR - A model for simulating flow and quality in river systems	- Richard Williams (IH)
16.25	Modelling biological behaviour in Rivers	- Steve Ormerod
16.45	Modelling Physical Processes in Lakes	- Glen George (FBA)
17.05	Modelling Chemical Processes in Lakes	- Bill Davidson (FBA)
17.30	Disperse to Hotels	

Thursday, 11 May

09.00	Flood Event Modelling	- David Boorman (IH)
09.15	A Simple Framework for Assessing Performance of Dynamic Models	- Nick Bonvoisin (IH)
09.30	Dynamic Hydrological Models - lumped conceptual approach	- Charles Eeles (IH)
09.45	Dynamic Hydrological Models - distributed mechanistic approach	- Ann lver (IH)
10.00	Coffee	
10.15	Interactive Ant and Butterfly Population Models	- Graham Elmes (ITE)
10.35	Modelling Cyclic Behaviour in Grouse Populations	- Malcolm Mountford (ITE)
10.55	Modelling Interactions between the Cinnabar Moth and its Food Supply	- Ken Lakhani (ITE)
11.10	PHABSIM - A model linking hydrology and stream ecology	- Alan Gustard (IH)
11.20	General Circulation Models - the IH/Met Office Modelling Programme	- Hans Dolman
11.45	Other Short Presentations and Brain Storming Session a) Current Gaps in Modelling b) Collaboration between Institutes and Universities c) Modelling Strategy for TFSD d) Resources and Funding - A Community Programme?	
12.30	Lunch	
13.00	Demonstrations of QUASAR, MAGIC, Flood Model, Framework Modelling Package, HYROM, etc	

APPENDIX 2

PARTICIPANTS - MODELLERS GROUP WORKSHOP

Participant	Paper
ITE Bush	
Dr Neil Cape	/
Dr M Cannell	
Mr Stephen Robinson (student)	
ITE Merlewood	
Dr Phil Ineson	/
Dr David Lindley	
Brita Svensson	
Bengta Carlsson	
Dr J M Sykes	
Dr T V Callaghan	
Dr M Hornung	
ITE Banchory	
Dr Malcolm Mountford	/
ITE Monks Wood	
Dr Ken Lakhani	/
Dr Mike Roberts	
Dr Keith Bull	
Dr T M Roberts (11 May only)	
Dr Clive Pinder	
Dr Arthur Marker	
Dr Mark Hill	
ITE Furzebrook	
Dr Graham Elmes	/
Dr Steve Chapman (11 May only)	
Dr Ralph Clarke	
ITE Nottingham	
Dr Neil Crout	/
FBA Windermere	
Dr Glen George	/
Dr Bill Davison	/
Ms Margaret Hurley	
FBA Wareham	
Dr Alan House	/
Dr Hugh Dawson	
UWCC	
Dr Steve Ormerod	/
NERC Unit Newcastle	
Dr Trevor Cooper	

MODELLERS GROUP WORKSHOP - IH PARTICIPANTS

Participant	Paper
Frank Law	
Paul Whitehead	/
Alan Jenkins	/
Richard Williams	/
David Boorman	/
Nick Bonvoisin	/
Charles Eeles	/
Ann Calver	/
Alan Gustard	/
James Bathurst	
Alice Robson	
Hans Dolman	
David M Cooper	
David Waters	
Trevor Staples	

APPENDIX 3

A Glossary of Terms used in Mathematical Modelling

by

P.G. Whitehead
Institute of Hydrology
Wallingford

May 1989

Modeling Terminology

This glossary has been provided to clarify the terms used in mathematical modeling. The terms have been grouped logically, and these groups of definitions are intended to be read together, so that contrasting definitions may be easily compared.

For reference purposes, an index is provided. The number appearing against each term is its sequence number in the glossary.

INDEX TO APPENDIX 1

Autoregression	34
Black Box model (see input-output)	41
Coefficient (see parameter)	19
Compartment (see compartmental model)	15
Constituent (see state)	17
Correlation	33
Deterministic	3
Deterministic simulation	24
Design model	13
Difference Equation	10
Differential Equation	8
Discrete time equation	9
Distributed parameter model	20
Dynamic	7
Dynamic time series analysis	40
Flux	11
Empirical	5
Estimation	27
Exogenous	18
Forecasting	22
Generated data	26
Holistic modeling	42
Input-output model	41
Least squares regression	28
Lumped parameter model	21
Mass balance (or conservation)	16
Mathematical model	1

Mechanistic (conceptual) 4
Monte Carlo (see stochastic simulation) 25
Moving average 35
Noise 36
Operational model 14
Parameter 19
Periodic (see seasonal) 32
Planning model 12
Prediction 23
Recursive 39
Reductionist modeling 43
Seasonal 32
State 17
Statistical properties of time series 30
Steady state 6
Stochastic 2
Stochastic simulation 25
Time series 29
Time series analysis 38
Trend 31
Turnover (see flux) 11
White noise 37

APPENDIX 1. MODELING TERMINOLOGY

1. Mathematical Model

A mathematical model is a set of equations which, together with logical statements expressing relations between the variables of interest, represents (or approximates to within the required accuracy) the behaviour of the system. Mathematical models are often categorised as stochastic or deterministic; steady state or dynamic; mechanistic or empirical. For the present purposes, a further categorisation, on the basis of the model's objective, is useful. All of these categories are defined in this glossary.

2. Stochastic (probabilistic)

If any of the variables in the model are regarded as random then these are termed stochastic variables. For example, the amount of rain falling on any one day in the future cannot be reliably predicted. However, from a knowledge of previous rainfall statistics, we can ascribe a probability of, for example, no rain, 5mm, 10mm, etc., falling on that day. The amount of rainfall can therefore be regarded as having a known probability distribution, and is capable of being any one of a number of values, each with a known certain probability. A further example is the treatment of run-off from agricultural land as a stochastic input.

3. Deterministic

If any of the variables in the model are regarded as free from random variations then they are called deterministic terms. A deterministic model of a river is one which assumes that all phenomena that characterise the river can be explained precisely;

in other words changes in environmental conditions, biochemical reactions etc., are all known perfectly and the model reproduces the response to these phenomena exactly. A deterministic term can be mechanistic or empirical.

4. Mechanistic (conceptual)

These terms are those which are, for example, based upon the physics, chemistry and biology of the processes involved. The equation relating the mass transport of a conservative chemical compound present in the river in terms of its initial concentration, the river velocity and a dispersion co-efficient is a mechanistic equation.

5. Empirical

Empirical terms have little physical basis, but are included in a model because they have been found to fit observations made in the real system. The following equation relating the reaeration of a river in terms of depth, velocity, percentage saturation and temperature is an empirical equation.

$$d = 0.508\ U^{.067}\ H^{-0.85}(C_s - C)$$

6. Steady State

The values of the variables which would obtain if all inputs remained constant, and if the various mechanisms involved reached an equilibrium. This is called the steady state condition.

7. Dynamic

A dynamic model is concerned with the variation of the variables in time (often over relatively short periods) when the system is subjected to time varying conditions, such as the transient input of effluent into a river system. Dynamic models can be in the form of differential equations which represent continuous time systems or difference equations which represent discrete time systems.

8. Differential Equation

An equation relating the instantaneous change in a variable (the differential coefficient) to the instantaneous value of that variable (and possibly other variables). Such an equation is used to describe the continuous time variation of a physical/ biological system and is usually obtained by recourse to physical/ biological laws, such as Newton's Laws of motion or mass conservation. For a specified sampling interval, the solution of a continuous-time differential equation can be expressed in terms of a discrete-time or difference equation.

9. Discrete time equation

An equation relating the value of a system variable at a given time instant to the value of that variable (and possibly other variables) at previous instants of time. Here the 'sampling interval' between time instants is usually constant and defined.

10. Difference equation

An equation relating the changes in a system variable over discrete intervals of time to the changes that have occurred in that variable (and possibly other variables) during previous time intervals. A discrete-time equation can always be converted into a difference equation if desired.

11. Flux or turnover

The rate of change of some system variable, such as mass.

12. Planning Model

A model which can be used to investigate the overall conditions of a river system and the effect upon it of generalised alternative effluent disposal strategies. In the past this type of model has usually been steady state and deterministic, but we feel it should at least take probability into account.

13. Design Model

A model which can be used to determine the 'average' effect upon a given section of river of a particular discharge (possibly in terms of a probability distribution or the 'moments' of the distribution).

14. Operational Model

A model which can be used to calculate the effect of variations in inputs on the river and hence to develop short-term control strategies (e.g. for flood alleviation, control of algal growth). Such models will normally be dynamic and may include stochastic, mechanistic and empirical terms.

15. Compartmental Model

A model consisting of a number of linked compartments. Each compartment is a conceptual area (or volume) within a system in which all of the boundaries are defined. Inputs or outputs of information may occur across these boundaries. Variables internal to the compartment are termed state or endogenous variables, and variables affecting the compartment from outside are termed exogenous or input variables. The dynamic modeling techniques used here estimate the dynamic mass balances of various variables within the compartment. The dynamic behaviour of adjacent compartments may also be linked, via input variables, so that an overall estimate of the behaviour of the system may be gained.

16. Mass balance (or conservation)

The concept that all mass in a physical system must be accounted for in some manner. For example the mass flow at the output of a compartment must be accounted for in terms of the input mass flow, and the sources and sinks of mass within the compartment. A similar conservation concept can be applied to energy, momentum etc.,

17. State (constituent)

Any variable whose distribution in distance and time is central to and is to be calculated by the model. Dissolved oxygen, BOD, ammonia, nitrate, chloride are examples of states. (The concept of state has a precise mathematical meaning but this is unimportant in the present context).

18. Exogenous

Any variable which acts as an input to the system and whose distribution in distance and time is regarded as known. Air temperature and rainfall are often exogenous variables as their values, or distribution, are supplied to the model as known inputs.

19. Parameter or coefficient

An unknown appearing in an equation whose value is to be estimated by comparing (fitting) the quality observed in the river to that calculated by the model using the same conditions. The following equation is usually used to express the rate of decay of BOD:

$$L = L_0E - kt.$$

In this equation k is the parameter and it is estimated by using BOD values at points in the river system.

20. Distributed parameter model

A distributed parameter model is a spatio-temporal model usually in the form of partial differential equations.

21. Lumped parameter model

A lumped parameter model is a purely temporal (ordinary differential equation) description of a system at a specific spatial location.

22. Forecasting

The production of values for a variable which are expected to occur at specified times in the future, from a knowledge of the values of that variable (and possibly other variables) up to the present time. Forecasting is usually performed for short time periods ahead in relation to the total time span of the available data. Forecasting is usually based on time-series analysis (see later).

23. Prediction

The prediction of values for a variable which might reasonably be expected to occur at a given time in the future if assumptions made about other factors (e.g. size and location of sewage discharges) apply at that future time. Prediction is usually performed for long time periods ahead compared to the time span of available data. Prediction is, of course, related to forecasting and may use time-series methods or mechanistic models.

24. Deterministic simulation

Solution of a deterministic model, usually with a computer, under specified input and parameter conditions in order to assess the behaviour of the model under these conditions.

25. Stochastic simulation or Monte Carlo Analysis

If the assumptions made in prediction or forecasting include stochastic terms then the prediction takes the form of a probability distribution. The calculation is performed a large number of times, each time with randomly chosen values for the stochastic terms, and each time yielding a different result for the variable of interest. When all the results are taken together the required probability distribution can be found. This operation is called stochastic simulation.

26. Generated data

The values of the stochastic terms used in a stochastic simulation are called generated (or synthetic) data and must be chosen to preserve the structure (defined later) exhibited by those variables in the available data. This is usually carried out on a computer with the random elements generated from 'white noise' (see later) sources.

27. Estimation

The determination of the 'best' values for the parameters of the model, for example, by the comparison of observed values in the river with values calculated using the model under conditions identical to those which actually occurred. Many estimation procedures are available from simple deterministic fitting to sophisticated statistical estimation methodology.

28. Least squares regression

A method of estimation in which the 'best set' of parameters is defined as those which minimise the sum of the square of the difference between the observed and calculated quantities over all observation points.

29. Time series

A time series is a set of values for a variable observed at regular intervals of time (e.g. daily river flows at a particular point). Time-series analysis is the analysis of a time-series in order to characterise its behaviour in model terms (see later).

30. Statistical properties of time series

These may include any or all of the following components:

31. Trend

A persistent change in the variables. In the case of river data it may arise from, for example, slow climatic changes, increases in population, urbanisation.

32. Seasonal (periodic)

A regular pattern of a fixed length (the period) which is repeated throughout the data (for example temperature is seasonal with periods of one day and one year).

33. Correlation

A component which reflects the dependence of a variable on values of one or more other variables.

34. Autoregression

An expression which reflects the dependence of a variable at one instant of time on its values at earlier times.

35. Moving average

An expression which reflects the dependence of a variable at one instance of time on the values of another variable(s) at the present and previous instant of time.

36. Noise

The measurements of variables associated with a system may be partially obscured by random variations in the data. These random variations are often called 'noise' and the time series formed by subtracting the model generated estimates of the system variables from the measured original time series is called a noise or residual sequence. This series may also have general statistical properties.

37. White noise

A series of serially uncorrelated random numbers taken from a normal distribution with zero mean and a constant specified variance (by analogy with white light).

38. Time series analysis

A general term for statistical methods specifically used for the analysis of a time-series to determine its statistical properties in modeling terms, either within the series itself or in terms of a similar series for other exogenous variables. Usually the parameters of such time-series models are constant

However, it is appreciated that when analysing time-series it is possible that the underlying processes are "non-stationary" and, therefore, the model parameters may be changing with time.

39. Recursive analysis

Serial processing of data, in which data is entered and processed a sample at a time whilst working serially through the entire data set.

40. Dynamic time series analysis

This is an extension of simpler techniques and allows the estimates of each parameter to vary as further points are added to the series. This allows for the analysis of non-stationary series and ensures that the 'currently best' model is available for forecasting.

41. Input-output (or Black Box) model

One which is formulated by relating observed variations of a variable to observed variations in the same or other variables without regard to the physical (mechanistic) processes occurring within the system.

42. Holistic modeling

The conception and modeling of a multi-compartment system as a complete entity. Identification and estimation are carried out on the complete system.

43. Reductionist modeling

Conception and modeling of the system as a collection of sub-systems in which each sub-system is modelled independently and the sub-systems are assembled together to provide a model of the complete system. Here identification and estimation are carried out on the separate sub-systems.

APPENDIX 4

To: P.G.Whitehead
From: D.M.Cooper - Institute of Hydrology
Date: 11/5/89

I think the modelling workshop was useful particularly for clarifying ideas. I came away with one over-riding impression: the quite major difference between modelling biological and non-biological processes. Hydrology is basically a branch of geophysics/chemistry, like meteorology or oceanography. The divergence between "holistic" and "reductionist" approaches is fairly minor, the latter being clearly seen as integrated versions, usually in space, of the former. There are therefore common strands to water flow models, based on:

1. Mass conservation of water
2. Dynamics of water, eg. Darcy's Law
3. Mass conservation of solutes
4. Dynamics of solute movement, eg. Fick's Law
5. Reaction chemistry

All the hydrological and hydrochemical models discussed contain these elements, which are well anchored in physical theory. Where approximations are made, for example in the water dynamics in MAGIC, it is still clear how the model incorporates points 1 to 5. I think it would be useful to examine all such models to see how they fit into this framework, and that this forms the basis for a coherent research programme.

In contrast, it seems to me as a layman, that ecological modelling cannot easily be reduced to basic principles; there are no obvious fundamental laws, and population dynamics is far more complex than any part of hydrology. While some linkage may be possible, opportunities for integrated hydrological/ecological modelling seem rather limited. Examples might be the use of hydrological information as input to ecological models, in the same way as meteorological information, and feedback from biological communities to hydrology through changes in evapotranspiration or rates of chemical reactions governed by microbiological activity.

As far as integrated hydrological and hydrochemical modelling is concerned a good place to start might be with the cycling of particular elements. The carbon, nitrogen and phosphorus cycles are already well understood because of their agricultural importance. It might be useful to study the cycling of other elements within particularly simple catchments. This falls into the framework of points 1 to 5.

UNIVERSITY OF NOTTINGHAM
DEPARTMENT OF PHYSIOLOGY AND ENVIRONMENTAL SCIENCE

SUTTON BONINGTON
LOUGHBOROUGH LE12 5RD
Telex: 37346 UNINOTG
Facsimile: (05097) 3917

NOTTINGHAM (0602) 484848
Ext:

Ref NMJC/AK

12/5/89

Dear Paul,

thankyou for allowing me to participate, as a non-NERC person, in this week's modelling workshop. You asked for any comments on the meeting, and I may as well pass mine on while they're still fresh in my mind.

I think the meeting compared favourably with the equivalent AFRC workshop which took place last month, indeed the fact that it took place over 2 half days was a definite advantage as this enabled more wideranging conversations in the pub. However, if one of the objectives is to encourage collaboration between the various institutes and universities then perhaps the meeting should be made open to university groups, such as ourselves, who have related interests. This is certainly the approach the AFRC have taken. Indeed, this may overcome some of the objections to an annual meeting, voiced by Bill Davidson, namely that there wouldn't be anything new to talk about in such a short time!

I hope these these thoughts are helpful,

Kindest Regards,

Neil

Dr Neil Crout

SCHOOL OF PURE AND APPLIED BIOLOGY

Acid Waters Unit

Head of Department: Professor R W Edwards BSc DSc FIBiol FIWPC FIFM

UWCC Acid Waters Unit
c/o Welsh Water Authority
Penyfai House
Furnace
Llanelli - Dyfed
SA15 4EL

☎ (0554) 757031

Dr. Paul Whitehead,
Institute of Hydrology,
Crowmarsh Gifford,
Wallingford.
Oxon.

12th May, 1989.

Dear Paul,

I enjoyed the TFS modelling workshop. Thanks for giving me the opportunity to present some work.

I'm sure I came away thinking of fruitful areas of collaboration, and have now written to Alan Gustard with this in mind. It strikes me that use of the Brianne Data on flow, habitat characteristics and biology could provide for an interesting application of PHABSIM to upland streams. Indeed, it could give us a much needed tool for assessing the biological effect of climate-induced changes on flow pattern. We would, of course, need to find some funding.

If you do write to Bernard Tinker, therefore, it might be worth putting a few thoughts to him:-

1) That ecological models CAN be married to produce useful results. This has been shown in the MAGIC studies.

2) That the collaboration between ourselves at UWCC and IH has already led to useful information used by government in both policy formation and heuristic studies.

3) That something like PHABSIM might provide scope for further collaboration in a similar vein.

All the best.

Yours sincerely,

Steve

S.J. Ormerod.

P.S.
Is there any chance of help from NERC with my expense for the workshop?

APPENDIX 5

Water Res. Vol. 18, No. 8, pp. 945–953, 1984
Printed in Great Britain. All rights reserved

0043-1354/84 $3.00 + 0.00
Copyright © 1984 Pergamon Press Ltd

MODELLING ALGAL BEHAVIOUR IN THE RIVER THAMES

P. G. WHITEHEAD[1] and G. M. HORNBERGER[2]

[1]Institute of Hydrology, Wallingford, England and [2]Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22903, U.S.A.

(*Received August* 1982)

Abstract—Forecasting the movement and growth of algae in river systems is particularly important for operational managers responsible for the distribution and supply of potable water. Algae affect the taste and smell of water and pose considerable filtration problems at water treatment plants. In a collaborative study with the Thames Water Authority, algal models have been developed for the River Thames. The non-linear processes controlling algal growth are examined using a generalized sensitivity analysis technique and the dominant parameters controlling system behaviour are identified. The extended Kalman filter (EKF) is then used to estimate these important parameters. The technique of using generalized sensitivity analysis prior to EKF estimation is suggested as a pragmatic approach to the problem of identifying the subset of physically, chemically or biologically meaningful parameters controlling system behaviour in mechanistic models.

Key words—algal models, River Thames, model identification, parameter estimation, sensitivity analysis, Kalman filter, water quality modelling

INTRODUCTION

With the increasing demands on the Thames as the principal source of water for London, it is not surprising that in recent years concern over present and future water quality has been expressed. Water quality problems of immediate interest to the Thames Water Authority reflect the multiple use of the river as the principal disposal pathway of industrial and domestic effluent in addition to being a major source of water for agricultural, industrial and domestic purposes. In particular, progressively increasing levels of nitrates in surface and groundwater systems have exceeded WHO and EEC standards and the Thames Water Authority have restricted abstractions during periods when concentrations of nitrate in the river are high (Onstad and Blake, 1980). Furthermore, major algal blooms occur on the river and these present operational management problems for the Water Authority. Abstracted water is pumped into reservoir storage prior to distribution to water treatment plants and algal growth affects water taste and smell and causes filtration problems. The prediction of algal growth, transport and decay is, therefore, of considerable importance in water supply management. In the paper an algal model is described and a sensitivity analysis technique utilized to identify key parameters controlling algal behaviour. Finally, the extended Kalman filter is employed to estimate these parameters using data from the River Thames.

MODELLING APPROACH

Mass-balance model

Algal distribution and growth processes in the River Thames have been the subject of research by a number of biologists this century (Fritsch, 1902, 1903, 1905; Rice, 1938; Kowalczewski and Lack, 1971; Lack, 1971; Bowles, 1978), but modelling techniques have not been used heretofore to obtain an adequate description of the system. There have been few modelling studies of algal growth and transport processes in rivers in general, although the analysis of flow and quality data using modelling techniques have developed considerably in recent years (Thomann, 1972; Beck and Young, 1976; Whitehead *et al.*, 1979, 1981). In this paper the development of mechanistic models for algal transport and growth is stressed. By mechanistic we mean a model containing mathematical expressions for the various physical, chemical and biological phenomena controlling system behaviour. We examined initially a mass-balance model as applied to three years of weekly algal data over 1974, 1975 and 1976 for six reaches of the river shown in Fig. 1.

The aim of this study was to determine to what extent transport alone could explain the observed variations in chlorphyll-*a* data. The results of this analysis are given in Whitehead and Hornberger (1984). It was found, not surprisingly, that transport alone could not account for the observed variations in algal levels but that complex processes of algal growth and death or sedimentation were occurring. In order to model such behaviour it is necessary to hypothesize mechanisms for these processes.

Rather than take a standard model developed for a particular system the approach herein has been to evaluate the most likely factors controlling algal growth and losses and to represent these mathe-

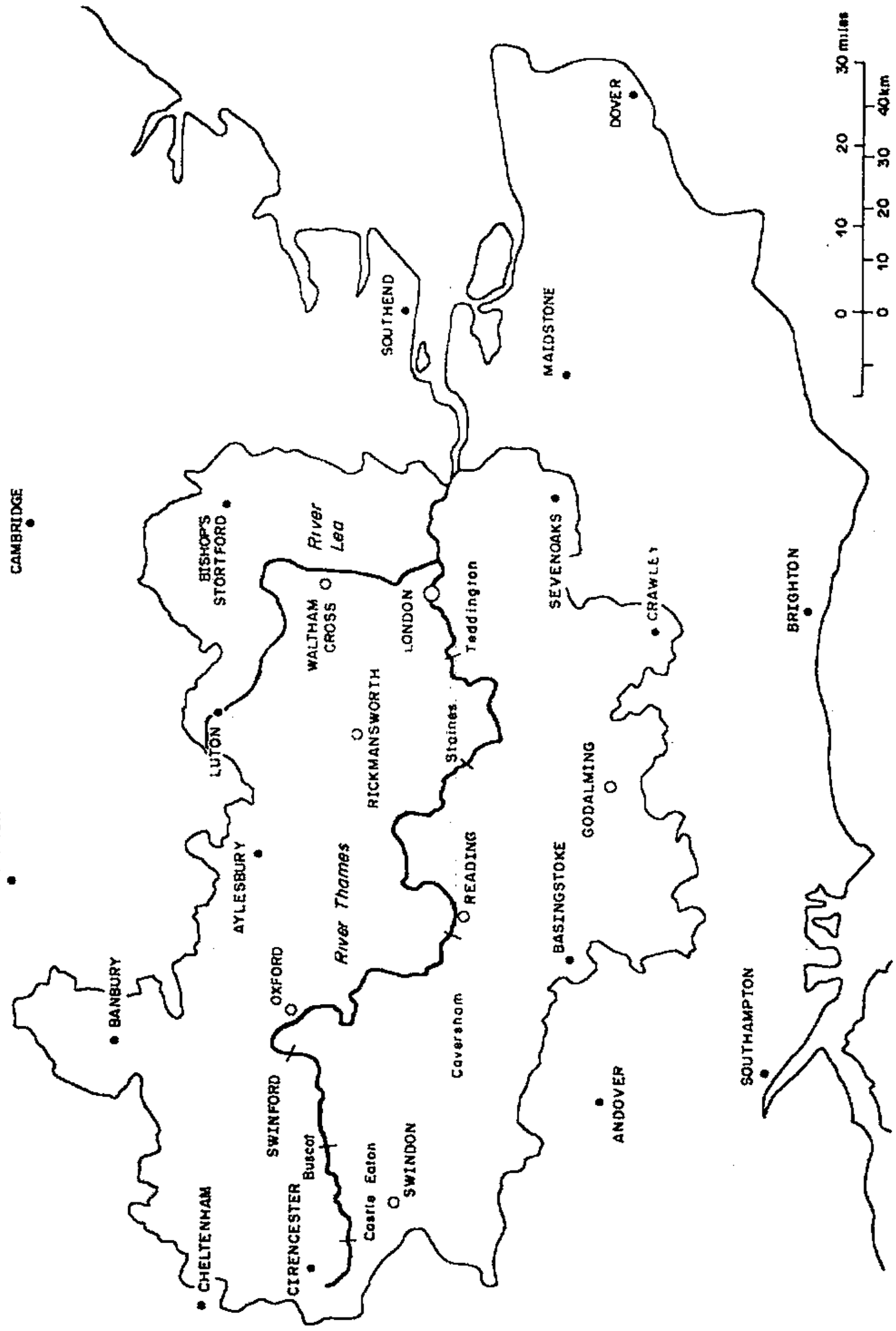

Fig. 1. The Thames catchment area (reach boundaries at Castle Eaton, Buscot, Swinford, Caversham, Staines and Teddington).

matically. The four factors considered particularly important for algal growth are:

(i) the growth coefficient;

(ii) the effect of solar radiation which under conditions of unlimited nutrient availability provides the main driving force for algal growth;

(iii) the effect of turbulence which tends to increase with increasing flow causing resuspension of sedimented material and reducing light penetration;

(iv) the self-shading factor in which algal populations grow to the point where light penetration is reduced by algae themselves.

The loss processes were assumed to be related to the concentration of algae via a first order decay term but nonlinear forms were used for the light limitation terms. The basic mass balance equations for describing concentrations of live and dead algae are similar to those developed by Beck (1978) but include mathematical terms to describe the four factors discussed above. The equations are as follows:

Live algae

$$\frac{dx_1(t)}{dt} = k_1\,Q_i(t)\,u(t) - k_1\,Q_o(t)\,x(t) - k_2\,x_1(t) + k_3\,\frac{I(t)}{Q_o(t)}\left(\frac{k_4}{k_4 + (x_1(t))^{k_5}}\right)\left(\frac{I(t)}{k_6}\right)^{k_7} \times \exp\left[1 - \left(\frac{I(t)}{k_6}\right)^{k_7}\right]. \quad (1)$$

Dead algae

$$\frac{dx_2(t)}{dt} = -k_1\,Q_o(t)\,x_2(t) + k_2\,x_1(t) - k_8\,x_2(t) \quad (2)$$

where $x_1(t)$ and $x_2(t)$ represent the live and dead algae respectively at the output (downstream) boundary of the reach, measured as chlorophyll-*a* (μg l^{-1}); $u_i(t)$ represents the input (upstream) algae concentration (μg l^{-1}); $Q_i(t)$ and $Q_o(t)$ represent the upstream and downstream flow rates; $I(t)$ is the solar radiation level (W cm^{-2}); k_1 determines the residence time characteristics of the model such that $k_1Q_o(t) = 1/\tau$ where τ is the residence time; k_2 is the algal death rate; k_3 is the growth coefficient; k_4 is a half-saturation level for the self-shading function, $\{k_4/[k_4 + (x_1(t))^{k_5}]\}$ and k_5 is included as a power term on $x_1(t)$ to enhance the self-shading factor at high algal concentrations; k_6 represents the optimal solar radiation level in the term $[I(t)/k_6]^{k_7} \exp\{1 - [I(t)/k_6]^{k_7}\}$ which accounts for the decrease in algal growth under low light intensity and the apparent decrease in growth under extremely high light intensity conditions in the Thames (Steele, 1978); k_7 enhances the effect of this solar radiation term; k_8 is included in the dead algae equation to account for the loss of algae by sedimentation. An additional parameter k_9 is included in the model as a temperature threshold below which algal growth is zero, i.e. $k_3 = 0$ for $T < k_9$ where T is water temperature °C.

Estimation of model parameters

Many researchers have developed phytoplankton growth models for simulation purposes. In general, the approach to parameter estimation has been to select parameters quoted in the literature and assume that these values pertain to the system under investigation. Formal methods of parameter estimation have been used in few studies; e.g. Lederman *et al.* (1976) applied non-linear parameter estimation techniques to data from batch cultures of phytoplankton to directly estimate model parameters and Whitehead (1980) used an instrumental variable algorithm applied to differential equation models of water quality to estimate parameters. In this paper the extended Kalman filter (EKF) technique has been used to estimate model parameters.

The EKF is a recursive algorithm in which an estimate of the unknown parameter vector $\hat{\alpha}$ is updated while working serially through the data. The estimate $\hat{\alpha}$ of α at the kth instant in time is given by an algorithm of the following form:

$$\hat{\alpha}_k = \hat{\alpha}_{k-1} + G_{k|k-1}\{y_k - \hat{y}_{k-1}\} \quad (3)$$

where the second term on the right hand side is a correction factor based on the difference between the latest measurement y_k and the estimate $\hat{y}_{k|k-1}$ of that determinand derived from the model using estimated model coefficients obtained at the previous time point. $G_{k|k-1}$ is a weighting matrix whose elements are calculated essentially as a function of the levels of uncertainty (or error) specified for the model in the output response and the unmeasured input disturbances. A full description of the technique can be found in Jazwinski (1970) or Young (1974) and applications of the EKF for modelling nitrate, chloride, dissolved oxygen and BOD are given by Whitehead *et al.* (1981) for the Bedford Ouse River system and by Beck and Young (1976) for the River Cam.

Estimating parameters in mechanistic models is often difficult because of the non-linear nature of the process equations and the interdependence of parameters. In the case of the algal model there are nine interrelated parameters to determine and estimation is particularly difficult. This is an important aspect of the modelling study since it is generally not possible to obtain reliable estimates of the large number of parameters in most simulation models. It is preferable to eliminate parameters that cannot be identified with a given data set or to set those parameters which are thought to be well known and to then estimate the remaining parameters. Up to now there has been no systematic method of selecting the subset of parameters for optimization. A trial and error procedure is normally used to select these parameters but given the non-linear nature of most simulation models such an approach can present problems of interpretation and is certainly not rigorous. A generalized sensitivity analysis can aid in this parameter selection to ensure that the optimal set of parameters are obtained. Such a technique has been developed by Spear and Hornberger (1980) and applied to aquatic ecosystem problems by Hornberger and Spear (1981).

Generalized sensitivity analysis

The generalized sensitivity analysis technique is based on the utilization of a simulation model together with a classification algorithm. The classification algorithm allows the model outputs to be identified as either representative or as not representative of the observed behaviour. The idea is to inject uncertainty into the simulation model by selecting the parameters from specified probability distributions rather than from experimentally derived values. The simulation is repeated using different parameter sets and the parameter set classified as either producing or not producing a behaviour. Subsequent to these Monte Carlo trials, statistical analysis of the parameter sets is used to identify the key parameters causing the model to reproduce the observed behaviour. The theory behind this statistical analysis is based on the separation between the cumulative probability distributions of two parameter sets, and a Kolmogorov–Smirnov two sample test is utilized to test the separation. The test is described by Spear and Hornberger (1980) and the statistic $d_{m,n}$ is determined as the maximum vertical distance between the cumulative probability distribution curves for n behaviours and m non-behaviours. Thus, large values of $d_{m,n}$ indicate that the parameter is important for simulating the behaviour. The value of $d_{m,n}$ can be compared with a 90% confidence bound value to check that it is statistically significant. Further refinements of the technique are presented by Spear and Hornberger (1980) and Hornberger and Spear (1980, 1981).

APPLICATION TO THE THAMES ALGAL MODEL

In the algal modelling study on the Thames it is first necessary to define the system behaviour. The two important features of algal growth within the river are the presence of a spring bloom and the subsequent fall to relatively low levels in early summer. On this basis simulations are classified as a behaviour if the algal concentration, x_1, is at any time, above 100 $\mu g\ l^{-1}$ and below 400 $\mu g\ l^{-1}$ during a 5 week period in spring and if, in addition x_1 falls below 100 $\mu g\ l^{-1}$ and remains below this level for at least 2 weeks during the 5 weeks after the spring bloom.

The model parameters were selected initially on the basis of published information such as travel times for the Thames determined by the Water Authority or growth rates for algae in the Thames. As previously discussed there is considerable uncertainty associated with many of the parameters in the model. In the case of growth rates, for example, Swale (1962), measured a growth rate for *Stephanodiscus hantzschii* of 0.46 day^{-1} and Bowles (1978) determined a growth rate for *Asterionella* of 1.28 day^{-1} from loading studies on the Thames. Lund (1949) also determined a growth rate for *Asterionella formosa* of 1.73 day^{-1} under ideal growth conditions although this reduced to 0.138 day^{-1} under field conditions. The situation in the Thames is complicated by the changing nature of the river with relatively slow flow in the lower reaches compared with the flow in upper reaches between Buscot and Swinford.

A complete list of the mean parameter values for the Monte Carlo simulation runs is given in Table 1. In the Monte Carlo runs the parameter values are

Table 1. Monte Carlo simulation results for reach 5 of the River Thames

Monte Carlo simulation runs	1		2		3		4
Critical $d_{m,n}$ at 90% confidence level	0.326		0.430		0.470		*
Parameter value (P) Distribution separation (S)	P	S	P	S	P	S	P
k_1 related to travel time τ. $k_1 Q_a = 1/\tau$	0.5	0.2	0.16	0.14	0.16	0.34	0.16
k_2 algal death rate (weeks^{-1})	0.3	0.2	0.3	0.16	0.3	0.19	0.6
k_3 algal growth rate (weeks^{-1})	10	0.15	8.0	0.33	10	<u>0.48</u>	12
k_4 algal saturation level ($\mu g\ l^{-1}$)	100	0.17	100	0.1	100	0.17	100
k_5 power in saturation term	2	0.73	2.5	0.68	3	0.87	4
k_6 optimal solar radiation (W hours cm^{-2} per week)	20,000	0.51	13,000	0.5	15,000	<u>0.51</u>	10,000
k_7 power in light attenuation term	2	0.30	2	0.42	2	0.31	2
k_8 sedimentation rate (weeks^{-1})	0.3	0.13	0.3	0.26	0.3	0.25	0.3
k_9 temperature threshold effect (°C)	8	0.11	8	0.16	8	0.17	8
% behaviour (based on 100 simulations)	48%		79%		86%		98%

*Statistics invalid since distribution of non-behaviours indeterminable from 2% of simulations.

Table 2. Statistics for 9 parameters in simulation run 1

Parameter	Normalized mean under behaviour	Normalized mean under non-behaviour
k_1	0.21	0.10
k_2	-0.24	0.57
k_3	0.10	0.18
k_4	0.62	0.21
k_5	0.54	-0.70
k_6	-0.56	0.42
k_7	0.34	-0.24
k_8	-0.84	0.53
k_9	-0.11	-0.18

selected randomly assuming a rectangular distribution with a range of ±50% of the mean of the parameter. This ensures that a wide spread of parameter values is selected and that behavioural patterns are fully explored.

Table 1 shows the parameter values used in four sets of Monte Carlo simulations together with the maximum separation between the parameter distributions and the critical separation $d_{m,n}$ at the 90% confidence level. It is particularly interesting to note that relatively few parameters appear to be significant in determining behaviour. Over the four simulations only three parameters are clearly identified as critical, these being the growth rate k_3, the power term in the saturation factor k_5 and the optimal solar radiation level k_6. In the first simulation only 48% of the runs satisfy the behaviour criteria. From analysing the behaviour-producing parameters it is possible to determine whether to increase or decrease the mean values of parameters in order to increase the percentage of behaviour. For example, in the case of the power term parameter k_5 in the saturation function, the normalized mean under the behaviour is 0.54 as shown in Table 2 suggesting that this parameter should be increased. By increasing this parameter the shape of the saturation function is altered thus enhancing the effect of the saturation level. Similarly in the case of k_6, the optimal solar radiation level, the normalized mean is −0.56 suggesting a reduction in this parameter. The Monte Carlo simulations therefore can be used as a crude estimation procedure and the percentage of behaviours increased from 48 to 98% over the four runs using this approach.

From a systems point of view what is particularly significant is that only three of the nine parameters control system behaviour. In most modelling studies of ecological or hydrological systems it is conventional to assume that each parameter is equally important. As previously mentioned, in many simulation studies a trial and error procedure of model calibration occurs in which a subset of the parameters is adjusted until a reasonable model fit is obtained. With large complex models this process can be particularly difficult because of interactions between parameters and mechanisms. The generalized sensitivity analysis approach can therefore be used in this situation to determine the dominant parameters controlling behaviour in a systematic manner.

APPLICATION OF THE EKF

In the Thames algal model is proved impossible to apply a technique such as the extended Kalman filter to estimate all nine parameters. The EKF technique applied in this situation gives parameter values which are either clearly incorrect or show colinearity in which one parameter increases as another decreases to cancel out its effect. Thus in order to obtain reasonable parameter estimates the EKF is applied to the three critical parameters indicated by the sensitivity analysis with the remaining parameters set to values estimated from independent laboratory or field measurements.

The estimation results obtained by the EKF for the fourth and fifth reaches are typical of those for the

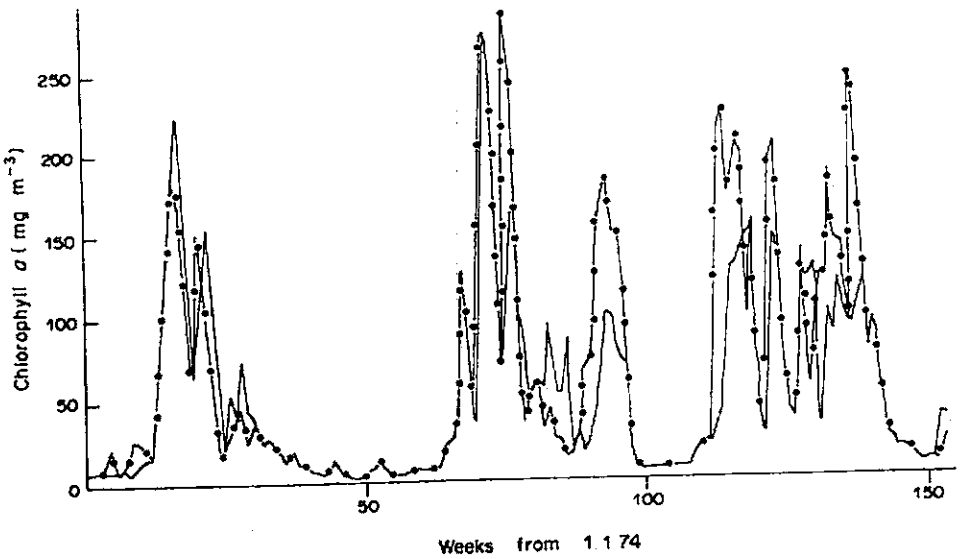

Fig. 2. Estimated (——) and observed (●) chlorophyll-a for 5th reach.