HARD OR SOFT ENVIRONMENTAL SYSTEMS?

M.B. Beck
International Institute for Applied Systems Analysis, Austria
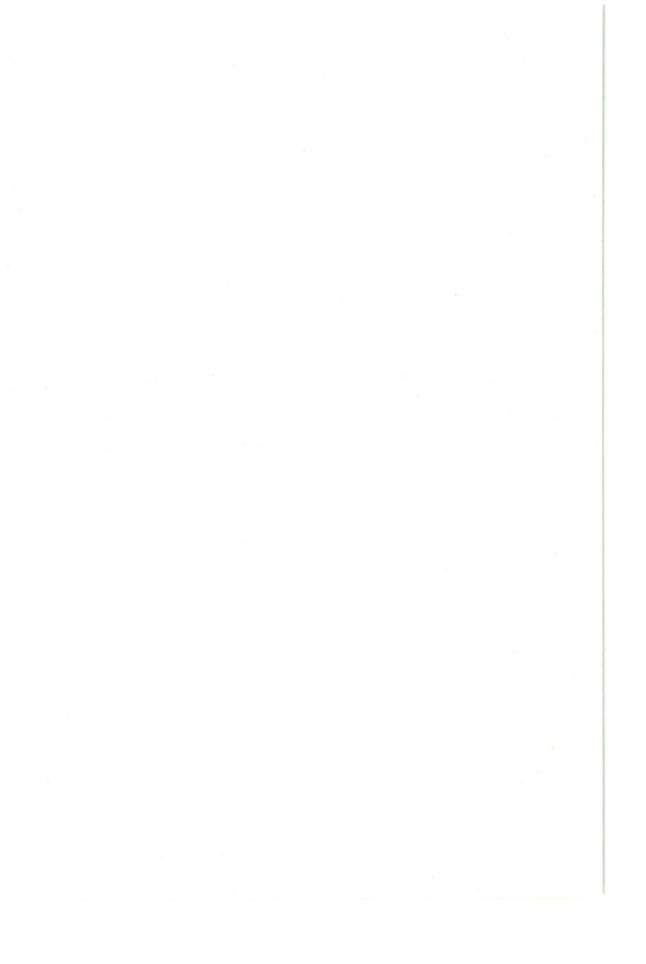
FOREWORD

In recent years there has been considerable interest in developing models for river and lake ecological systems, much of it directed toward large and complex simulation models. However, this trend gives rise to concern on several important counts. In particular, relatively little attention has been given to the problems of uncertainty and errors in field data, of inadequate amounts of field data, and of uncertainty about parameter estimates and the relations between important variables.

The work of the International Institute for Applied Systems Analysis (IIASA) on environmental quality control and management is addressing problems such as these, and one of the principal themes of the work is to develop a framework for modeling poorly defined environmental systems.

This paper discusses, in qualitative terms, the preliminary outlines of such a framework. It argues that modeling poorly defined environmental systems presents some special methodological problems (also discussed previously in "Model Structure Identification from Experimental Data," reproduced from E. Halfon, editor, *Theoretical Systems Ecology: Advances and Case Studies*, Academic Press, New York, 1979, as IIASA Research Report RR-80-4, February 1980), and that there is a need to reconcile the growing incompatibility between what can be simulated in principle and what can be observed in practice.

Thus, uncertainty and the reliability of models and forecasts are key concerns of this paper.

JANUSZ KINDLER
Chairman
Resources and Environment Area

HARD OR SOFT ENVIRONMENTAL SYSTEMS?

M.B. BECK

International Institute for Applied Systems Analysis, A-2361, Laxenburg (Austria)

(Accepted for publication 22 April 1980)

ABSTRACT

Beck, M.B., 1981. Hard or soft environmental systems? *Ecol. Modelling*, 11: 233—251.

Recent trends in lake and stream water quality modeling indicate a conflict between the search for improved accuracy through increasing model size and complexity, and the search for applicability through simplification of already existing models. Much of this conflict turns on the fact that that which can be simulated in principle is simply not matched by that which can be observed and verified in practice. This paper is concerned with that conflict. Its aim is to introduce and clarify some of the arguments surrounding two issues of key importance in resolving the conflict: uncertainty in the mathematical relationships hypothesized for a particular model (calibration and model structure identification); and uncertainty associated with the predictions obtained from the model (prediction error analysis). These are issues concerning the reliability of models and model-based forecasts. The paper argues, in particular, that there is an intimate relationship between prediction and model calibration. This relationship is especially important in accounting for uncertainty in the development and use of models. Using this argument it is possible to state a dilemma which captures some limiting features of both large and small models.

INTRODUCTION

On occasion, it is important for the systems analyst to step back from the detail of his work and thence, from the suitably distanced point of view of the informed layman, to consider what has been achieved and what problems remain to be solved. The benefit to be derived from this would be a hoped-for improvement in the ease of establishing discussion among systems analysts themselves. The dangers are clearly that many important details will be omitted and that one will crudely oversimplify extremely subtle arguments. Yet, avoidance of these dangers is presumably part of the craft of applied systems analysis, a daunting thought indeed with which to start this paper.

The objective here is to introduce two key issues of lake and stream water quality modeling, and to describe these issues in essentially familiar terms. The two key issues of interest center upon problems of uncertainty, that is: first, uncertainty in the mathematical relationships hypothesised for a particular model; and second, uncertainty associated with the predictions obtained from the model. These are issues concerning the reliability of models

and model-based forecasts. Let us begin, however, by providing an explanation for the title of the paper, a title which may appear curious to some and obscure to others.

It has become the custom, in certain circles, to talk about 'hard' and 'soft' systems. On the one hand, electrical circuit systems are presumably 'hard': experiments can easily be conducted to identify the behavior of such systems and, indeed, a priori theory is capable of predicting accurately what the nature of that behavior should be. By contrast, most of us would agree that social systems can be called 'soft' systems for the purposes of analysis: a priori theory for such soft systems is strongly colored by the opinions of the analyst; existing theory is unlikely to lead to accurate prediction of future behavior; and planned experiments with the system are particularly difficult, if not impossible, to implement. Karplus (1976) has accordingly introduced the notion of a spectrum of models where the models range from 'white box' models (of electrical circuit systems) to 'black box' models (of socio-economic systems). Models for water quality-ecological systems are indicated on this spectrum as being dark grey. One supposes that this classification reflects the level of insight into system behavior that is embodied in the model. And this in turn suggests that dark-grey models might, with sufficient research effort, be progressively whitened. This last point, however, is debatable. The significance of Karplus' spectrum is rather the situation of water quality-ecological systems midway between the poles of hard and soft systems.

At the International Institute for Applied Systems Analysis (IIASA), research on applied systems analysis spans applications to many different kinds of systems, most of them extremely complex, and nearly all of the Institute's projects involve some form of modeling activity. For example, these modeling activities include:

(1) macro-economic model development;

(2) models for studying future energy demand and supply;

(3) regional agricultural production models;

(4) a health-care system model for use in health service planning;

(5) the development of a model for urban dynamics, with special emphasis on migration, employment change, and labor-force participation;

(6) cybernetic models related to the management of technological innovation;

(7) models for environmental quality control and management, which are of primary importance for this discussion.

Given the high degree of dependence on models as a part of the problem-solving procedure, it is natural to ask questions about the reliability of those models. And given the 'relatively hard' character of water quality-ecological systems within IIASA's research portfolio (for most of the systems associated with the preceding list might be said to be soft systems) it is natural to believe that 'objective' standards (empirical evidence) should be available for resolving those questions of reliability. This is not so. The availability of field

data, scarce though they may be, and the relatively hard nature of water quality-ecological systems seem to have misled some analysts into mistaken perceptions about the ease of analyzing such systems. It might be thought, for instance, that techniques of analysis that have proven so successful in their application to much harder systems, e.g. sytems of the type associated with aircraft and industrial process control, can be extended naturally into our present field of interest. Hard systems analysis has generated an impressive array of techniques that perform powerfully on well-posed problems. But, as this paper attempts to demonstrate, such well-posed problems are problems that can only be addressed once the primary, and much less tractable problems of relatively soft systems analysis have been solved. For instance, Maciejowski (1979) argues that the use of many standard (statistical) techniques for determining sufficient model complexity is questionable in the case of 'badly defined' systems (here 'badly defined' is used in the same sense as in Young, 1978). It is against this background of the limited applicability of existing methodology, then, that we find the significance of the relatively soft nature of water quality-ecological systems. Because it is thus especially difficult to demonstrate a rigorous analysis of model calibration and reliability, it is tempting to fill the resulting vacuum in the analysis with evidence that is not much more than mere opinion. But because of the relatively hard nature of water quality-ecological systems it is still more tempting to appeal to that hard character in order to disguise 'opinion' under the cloak of 'objective evidence'.

WHY CALIBRATION AND PREDICTION?

Having justified the title and setting of the paper, and let us emphasize the point that terms like 'hard' and 'soft' are only to be understood qualitatively, why should calibration and prediction be important? In order to answer this question we must try and identify recent trends in water quality-ecological modeling.

The immense possibility for complex systems simulation created by the advent of electronic computers has fostered the growth of large water quality models. 'Large' is, of course, one of those terms related to subjective perceptions, although a glance at much of the recent literature on water quality-ecological modeling will give an impression of the intended meaning (for example, Russell, 1975; Patten, 1975, 1976; Jørgensen and Harleman, 1978; Scavia and Robertson, 1979). In contrast, it is possible to state (quite subjectively) that at present only 'small' models have been rigorously calibrated against in situ field data, from which we may infer that small models contain no more than, for instance, 10 difference or ordinary differential equations (Beck, 1980a). By 'rigorous' calibration is meant the application of a formal algorithm for parameter estimation, where estimates of the error covariances (i.e. confidence bounds) for the parameter values so derived can also be ob-

tained. This definition, therefore, excludes trial-and-error tuning of the model parameter values in order to fit the model to the data. Let us thus assert, as an opening statement for the following discussion, that the ability to construct large simulation models does not necessarily either increase one's understanding of a system's actual behavior, or strengthen the validity of the models as approximations of reality.

This trend towards largeness is not the only trend, however. Going in the opposite direction we see a 'need' emerging for smaller models. This need is occasionally expressly stated (for example, Thomann and Winfield, 1976; Thomann, 1978), but is more often the most engaging debating point reflected in workshop proceedings (for example, Russell, 1975; Vansteen-kiste, 1975, 1978; Beck, 1978a). There are several reasons for wanting a small model: because it is not possible to verify a larger model against the available field data; because the responses generated by larger models are not readily intelligible; and, because the overlying techniques for optimal management and policy design cannot accommodate large models.

One may suggest that these opposite trends indicate a conflict: a conflict between the search for (supposedly) improved accuracy through increasing model size and complexity and the search for applicability through simplification of already existing models. Much of this conflict turns on the fact that that which can be simulated in principle is simply not matched by that which can be observed and verified in practice. This paper certainly intends to involve itself in that conflict, and not least in order to clarify some of the issues. Our usage of the terms hard and soft, large and small, and their juxtaposition, is an oversimplification of the issues and conflicts. But the objective is to encourage dialogue, not a set of non-intersecting monologues. The discussion of calibration will expose limitations of the available methodology; that is to say, a limitation brought about by inadequate perception of the relatively hard/soft nature of water quality-ecological systems. The discussion of prediction concludes with a dilemma that captures some limiting features of both large and small models. That dilemma is intended to stimulate the hoped-for dialogue.

CALIBRATION

Most of us would wish to be reassured that the patterns of behavior simulated by our models do in fact resemble actual patterns of behavior. Thus there is a need for model calibration (or system identification), an exercise with which one typically associates curve-fitting and parameter (coefficient) estimation. But the word 'calibration' is misleading. It suggests an instrument (the model, and in this case an instrument for prediction) whose design is complete and whose structure is beyond further argument. All that remains to be done is to make minor adjustments to some of the fittings, i.e. fine-tuning of the parameter values. This is an incorrect interpretation of calibra-

tion in the context of modeling water quality-ecological systems, and it is incorrect because it overlooks the significantly soft character of such systems. How, then, should we summarize the details of this soft nature?

First, field data from water quality-ecological systems are generally scarce. When data are available they are subject to high levels of error and uncertainty. Halfon (1979), for example, gives an indication of just how many sources of error there can be in data obtained from large lakes. These errors, however, are not the only causes of difficulties in the calibration of water quality-ecological models.

Young (1978) suggests that the inability to perform planned experiments is a distinctive feature of the problem of modeling badly-defined systems; and, clearly, water quality-ecological systems would be listed under such a category of systems. Second, therefore, success in model calibration is obstructed by the conditions under which field observations are obtained. Since the term 'planned experiment' has been introduced, it should be defined, as an experiment in which the responses of some of the system variables (i.e. ouputs, or effects) are recorded and are assumed to be unambiguously related to changes in other (input, causative) disturbance variables. In such a planned experiment, all variables but the chosen variables of cause, and any response variables thereby disturbed, are deliberately maintained at constant values. That is to say, the environment of the system is held constant, the causative variables can be manipulated so as to conform with a desired pattern of changes, and the experiment is planned such that unambiguous relationships between the system's variables can be determined. A planned experiment of this kind is virtually impossible for water quality-ecological systems. But this does not imply that complex natural systems do not permit the observation of natural experiments. For example, the hydrological sciences place much emphasis on the identification of catchment characteristics through analysis of the response of stream discharge to a storm event. The importance of the storm event is that it provides a significant input disturbance of the system behavior, and an output response can be relatively unambiguously related to that input. The environment of the hydrological system is not entirely constant, nor is the input disturbance manipulated at will. Natural experiments of this kind, however, are quite rare in water quality-ecological systems. Imagine by way of contrast, the 'extreme' response of a phytoplankton bloom in a lake. The bloom occurs because a specific but apparently commonplace sequence of environmental (input) conditions forces the state of the system into a region (of the state space) in which a non-linear mode of behavior is excited and becomes dominant. Unlike the example of the hydrological system, the response of the lake is probably not unambiguously related to a notably 'extreme' input disturbance. Rather, it may follow as a consequence both of subtle changes in the system's environment and of a very particular combination of circumstances within the lake at that point in time (or space).

The problems of error-corrupted data and the lack of planned or unam-

biguous natural experiments are technical, rather than fundamental problems of model calibration. They are purely technical in the sense that if the analyst knew, a priori, how the system ought to behave, then it would still be comparatively easy to distinguish the estimated patterns of behavior in the observed field data. The third, and basic problem of water quality-ecological model calibration is, therefore, the limited degree of a priori knowledge about expected system behavior. It may be asserted that, in spite of very many laboratory-scale experiments and a number of major field studies, knowledge of the relationships between the mineral, organic and microbiological components of water quality-ecological systems is actually quite uncertain. A sophisticated, but particularly apt, example of this type of uncertainty is illustrated in a recent study of Saginaw Bay, Lake Huron (Bierman et al., 1980) in which it was noted that the output response of their model was especially sensitive to the choice of hypothesis for the growth rate of phytoplankton. The model had originally been calibrated against field data from Saginaw Bay with phytoplankton growth expressed according to the threshold hypothesis, namely, that growth rate is governed only by that factor which is determined to be rate-limiting. There was, in fact, additional evidence from laboratory experiments to support the chosen hypothesis. But, Bierman et al. have subsequently admitted that an alternative hypothesis (the multiplicative growth hypothesis, where all factors contribute to an overall rate of growth) could probably have been calibrated against the Saginaw Bay data. Calibration of this differently structured model with the alternative growth-rate expression would almost certainly have resulted in different estimates for all the other parameter values in the model. The significance of the example is, of course, in demonstrating how there is sufficient uncertainty in a priori knowledge of system behavior to allow considerable speculation about the precise structure of the mathematical model. In short, there are ambiguities in the a priori knowledge of behavior patterns in water quality-ecological systems.

It is evident, therefore, that calibration of models for water quality-ecological systems is unlikely to be a simple matter of making minor adjustments to a well-designed 'instrument'. Instead, even before asking the question "Can I estimate the model parameters accurately?", the analyst must first ask himself whether he knows how the variables of the system are related to eachother, and whether information about these relationships can be identified from the in situ field data. In short, the analyst is faced with the problem of model structure identification (Beck, 1979). Put another way, and in a simpler form than the sophisticated example of the study of Saginaw Bay (Bierman et al., 1980): it is a fine idea to estimate the slope and intercept of a straight line drawn through a set of data points (i.e. parameter estimation), if it has already been established that a straight line, and not a curve, will give the best fit to those data (i.e. model structure identification). Hence, model structure identification logically precedes parameter estimation.

HOW THE SYSTEM VARIABLES ARE RELATED

Focussing on the problem of model structure identification in greater detail, the discussion may be prefaced by the following remarks. It has been stated that model structure identification is a problem because of the relatively soft nature of water quality-ecological systems, i.e.: the high levels of error associated with field data; the lack of planned or natural experiments; and the considerable limitations of a priori theory. One might, therefore, turn to the methodological armoury of applied systems analysis for assistance in solving this problem. But there we find only an impressive array of techniques for solving problems in the analysis of relatively hard systems. In other words, many methods are available for parameter estimation, whereas there is a strategic weakness in the number, variety and effectiveness of techniques for model structure identification. That weakness arises precisely because model structure identification can be assumed not to be a problem in calibrating models of better-defined systems.

Suppose the patterns of system behavior exhibited in the (historical) field data can be represented by the set A in the set P of all possible patterns of behavior (Fig. 1). This pictorial representation has its origins in the work of Mankin et al. (1977); in a qualitative, but perhaps not quantitative fashion it is a powerful medium in which to express the following arguments. For reasons that will be apparent later, care must be taken to qualify P as being the set of all behavior patterns that one would expect to observe in 'reality'. The
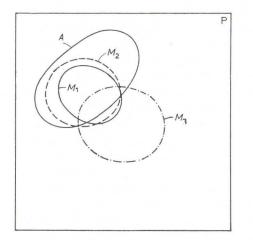

Fig. 1. Uncertainty and model structure identification. P is the set of all possible behavior patterns; A is the historically observed pattern of behavior; M_1 is the behavior pattern simulated by the first model hypothesis; M_2 and M_3 are alternative models hypothesized after assessing the suitability of M_1.

first hypothesis for a model (e.g. M_1) might be rather modest in size, allowing only a somewhat restricted type of behavior, although a reasonable proportion of the set of behavior patterns simulated by the model (M_1 in Fig. 1) is contained in the set A. Again, one must be careful about misinterpretation. Terms such as 'small model' or 'limited variety' of behaviour patterns should not be equated too literally with a small number of variables, equations, or relationships. Moreover, it should be noted that, strictly speaking, A and M_1 represent observation and simulation under exactly equivalent conditions. An example of a model typifying M_1 might be the Streeter—Phelps model of stream dissolved oxygen (DO) — biochemical oxygen demand (BOD) interaction. This model is a good starting point for analysis, although we are aware that its ability to describe system behavior is limited. Thus, given Fig. 1 as a pictorial representation of the problem, what does the analyst do? His first model may not be bad, for it has captured part of the essence of reality (A and M_1 have an intersection), but it is far from being good, because it does not simulate half of what was observed in practice. The crucial issue of model structure identification is that a method is required which provides a useful feedback of diagnostic information from analysis of the first hypothesis (M_1) so that a second hypothesis (M_2) can be cast more fully within the set of observed patterns (A). It would be undesirable at this stage of the analysis to suggest a revised model (e.g. M_3), probably both greater in size and with relationships different from those of M_1, that merely simulates more apparently spurious behavior.

The search for such an 'intelligent' model structure identification method (intelligent because it should indicate which parts of the structure are inadequate and how they might be corrected) is extremely difficult. It is not altogether desirable to use an analog of the curve-fitting technique that sequentially tests the goodness of fit provided by a straight line, a quadratic curve, a cubic curve, etc. This would be rather abstract, and we are especially interested in much less abstract forms for the relationships between the water quality variables under study. A more promising approach is to restate the problem of model structure identification in terms of the problem of parameter estimation (Beck and Young, 1976; Young, 1978; Beck, 1979; Whitehead, 1979). Calibration of each successive model hypothesis against the data then provides diagnostic information in the form of parameter estimates and residual errors of mismatch between the model and reality. If it transpires that the parameter estimates are as bizarre as something equivalent to an estimated constant for the earth's gravitational acceleration that acts upwards instead of downwards, it is clearly time to rethink the model.

This is, of course, an oversimplification. What has been described above is only a conceptual outline of the solution to the problem. Seldom are the field data likely to be sufficiently precise as to afford clear-cut rejection of the model, since the problem of model structure identification, as stated earlier, is beset with ambiguities. On occasion, however, the analyst is fortunate. For example, when calibration of a Streeter-Phelps model for stream

DO—BOD interaction yields a negatively-valued reaeration rate constant, as it did in the study by Beck and Young (1976), the analyst can be reasonably confident about rejection of the associated model structure. In such a situation, the analyst if forced to support an absurd hypothesis if he wishes to obtain correspondence between the given model and the data. But, when eventually the diagnostic evidence favors rejection of the model, is it really possible to formalize the procedure for generating the next hypothesis? In fact, isn't this a procedure that demands that spark of creative thought characteristic of scientific discovery? Perhaps, therefore, the analyst should be rather modest in searching for the intelligent algorithm of model structure identification.

Let us summarize the discussion thus far. Because of the lack of planned experiments, because field data are highly uncertain, and because a priori definition of the mathematical forms for relationships among the important system variables cannot be made categorical, the calibration of water quality-ecological models is not a straightforward exercise of parameter estimation. The prior problem of model structure identification has to be solved before accurate estimation of the parameter values is attempted. So what progress has been made in solving this problem of model structure identification? The basic aim of model structure identification is to seek plausible hypotheses for 'unexplained' relationships in a set of field data. Some of the case studies already conducted (e.g. Beck, 1978b) demonstrate that one can indeed attempt to solve this problem (in part), and that meaningful diagnostic evidence can be obtained in order to determine whether a model is falsified. Such an approach to model structure identification, by reference to the in situ field data, exploits the idea of curve-fitting as a 'means-to-an-end' and not as an 'end' in itself. Experience shows that approaching the problem from a variety of angles, e.g. using different types of models and different estimation algorithms, can yield different clues about why a given hypothesis is falsified and how it might subsequently be modified. Falsification of the model, or components thereof, rests partly upon judgements about absurd parameter values, or about implausible variations in the parameter values. Unless these variations and values can be defended by logical argument, then it must be conceded that the structure of the model does not match the structure of the observed patterns of behavior. Even in a relatively simple context, however, these kinds of solution to the problem are not easily derived, and in more complex situations (e.g. as described in Beck, 1980b), the basic process of absorbing and interpreting all the diagnostic information generated by the analysis becomes itself very much more difficult. The evidence cannot be sharply focused in order to reveal the absurd hypothesis. But to believe that such a sharp focus might be possible is arguably an illusion, since the field data are subject to high levels of uncertainty. The consequences of highly uncertain field data may become apparent in at least two different ways. First, there is the example quoted earlier (Bierman et al., 1980), in which either of two phytoplankton growth-rate hypotheses can be

calibrated against in-situ field data. As Bierman admits, there is usually not a unique set of parameter values and, it should be added, neither is there necessarily a unique model structure that will give a significantly superior fit between the data and the simulated responses of a complex model. Second, it is quite probable that field observations reflect just a small number of dominant modes of behavior, for example, that phytoplankton bloom in the spring. Because they are dominant, these modes of behavior may well obscure more subtle, minor modes of behavior, for instance, that the size of the bloom is determined by changes in temperature rather than by a specific rate of grazing by zooplankton. Both of these consequences of uncertain field observations are, in fact, indications of the problem of identifiability, a classic problem in model calibration. Young (1978) describes the same problem as follows: "There can, in other words, be a basic ambiguity; a situation in which a number of possible explanations for the observed behavior seem feasible, but where there exists little a priori evidence as to which of these explanations seems most plausible". The purpose of model structure identification is thus to allow a posteriori evidence (a posteriori, in the sense of having calibrated the model) to be brought to bear on distinguishing among one or another of the possible a priori explanations as (conditionally) the most plausible. The difficulty lies in focusing and interpreting the a posteriori evidence.

Given thus that one acknowledges all the limitations and inadequacies of the formal procedures for model structure identification, is it yet possible to point towards avenues for further progress? Our answer is positive, although for obvious reasons this is a cautious affirmative. For instance, it is not expected that any general solution for the problem can be reduced to the level of pure technique. But let us consider two conceptual views of the nature of the analysis required for model structure identification. For the first view, let us suppose that the ultimate objective is to recover 'natural experiments' from the observed data, and by analytical methods. It has, therefore, been assumed that reconstruction of 'planned experiments' from the field observations is not possible, because, according to the definition given earlier, none of the causative variables will have been manipulated so as to conform with a desired sequence of changes. It would, however, be reasonable to attempt to design the analysis of model structure identification such that it compensates for the unsteady environmental conditions of the 'experiment'. Once again there is a particularly apt example, that is, the recovery of an 'in situ chemostat experiment', where the objective is to identify the structure of the relationship between substrate and phytoplankton growth. In this example the skill of the analyst would lie in arranging the analysis such that extraneous interference with the 'experiment' (e.g. disturbances from the observed fluid mixing, vertical stratification, and seiche behavior of the lake) can be filtered out. This presupposes, of course, that that part of the model required to compensate for the 'experimental environment' is known a priori with sufficient confidence to permit the full power of the analysis to

be directed towards the problem of substrate — phytoplankton interaction. Such assumptions themselves have to be evaluated. By extending the analogy with laboratory experiments one intuitively reaches the conclusion that the analysis would attempt to define and identify ever more complex and detailed 'experiments'. This suggests, in turn, a rather natural, and fundamentally significant, approach to model structure identification: an approach that starts from a simple model (as described earlier with reference to Fig. 1) and progressively increases model complexity when the diagnostic evidence of analysis precludes acceptance of any simpler model structure. It is easy to imagine, however, that the sheer complexity of system behavior, and the uncertainties associated with the data, would soon impose constraints on the depth of such an analysis. But starting with a complex model and identifying those components of the structure that are essentially redundant (i.e. surplus content) is an approach seemingly fraught with many more difficulties. One of the key problems is that ambiguities arise in determining whether the a posteriori evidence supports rejection of an inadequate model structure. In the face of these ambiguities, and acknowledging the additional difficulties of interpreting large amounts of evidence, the analyst should respond by making particularly prudent choices for the postulated model structures. If the model is a vehicle for asking questions about the nature of reality (and if it is also a vehicle for recovering natural experiments), then it is advisable to make those questions as few, at least initially, and as unambiguous as possible.

The second conceptual view of model structure identification depends upon interpreting a parameter estimation algorithm as an information processing mechanism: information in the observed patterns of behaviour is translated into information about the model parameter values. Recall here that it has previously been said that model structure identification can be specified as a problem of parameter estimation. Recall also that the values estimated for the parameters can be used as diagnostic evidence of an inadequate model structure. Such evidence may well indicate what is wrong with the model, but probably it will not indicate why it is wrong, and almost certainly it will not indicate how a revised and better model might be hypothesised. Imagine, however, that at the interface between the model and the data there exists a matrix of 'information transfer channels'. In other words, within the estimation algorithm there is a correcting procedure that translates information about the perceived mismatch between model and reality into a revised set of parameter estimates. For the purposes of model structure identification, not only is it important to establish which parameter values are absurd, but it is equally important to know from which sources of mismatch these values derive. That is, it is instructive to record which of the information transfer channels is, or is not significantly 'activated', and over what periods of the observed data this activation occurs.

PREDICTION AFTER CALIBRATION

The unresolved ambiguities of model calibration may not be important when the model is used for making predictions. But in principle such ambiguities, if not easily recognisable, are undesirable. To examine why this should be so, let us construct a discussion around three conclusions about model calibration. These conclusions are deliberately stylized for the purposes of argument and are:

(1) that only relatively small and therefore, by association, black-box models can be calibrated against field data;

(2) that, in the face of the inevitable ambiguities that arise in calibrating larger models, rigorous calibration can be partially circumvented by a judicious use of information from sources other than the in-situ field data (for exmple, information from independent laboratory experiments or from in-situ observations of similar systems);

(3) that rigorous, quantitative calibration of a model is not necessary if the model embodies every detail of all possible behavior patterns that might be of relevance to all water quality-ecological systems.

Such conclusions are not particularly interesting if calibration is the sole objective of a given study. But what significance do these conclusions have if prediction is an ultimate objective? This is a much more interesting question, and one which will be considered with respect to each conclusion.

While it is generally true that only relatively small models have been calibrated against field data, it is certainly mistaken to dismiss small models as essentially suspect, black-box representations of reality. The aim of model structure identification, as we have described it, runs counter to that kind of dismissal. Nevertheless, whatever label one wishes to assign to such a model there will indeed be problems of prediction to which that model may not be well suited. The most pertinent criticism of a small model calibrated by reference to past behavior patterns is that is will be incapable of predicting a future dominated by conditions substantially different from the conditions of the past.

In respect of the second conclusion, let us assert that the ambiguities of model calibration, i.e. the existence of many combinations of parameter values that fit the data 'equally well' is due to the surplus content of the model (see also Beck, 1980). By 'surplus content' is meant therefore any parts of a model that cannot be identified from the in-situ field data. The small, fully identified model, in the sense used above, can thus be said to contain no surplus content. The analyst who justifies a model's surplus content on the basis of evidence from independent laboratory experiments should clearly admit the uncertainty associated with extrapolation from laboratory to field conditions. Quite apart from such an extrapolation, it is dangerous to believe that a value for a specific maximum growth-rate constant, for example, exists in some absolute sense. That value for the growth-rate constant is only defined relative to the model (the kinetic

expression) that was assumed and calibrated against observed nutrient and phytoplankton concentrations in the laboratory experiment. Likewise, the analyst who justifies surplus content on the basis of previous observations of similar field systems must support two possible arguments: either the surplus content of the model had originally been unambiguously identified in a prior calibration exercise with that other system, which is unlikely; or else his justification is built upon a chain of similar justifications with an original extrapolation from laboratory to field conditions. But this is not to dismiss the accumulation of experience. Rather, the important point is to be able to distinguish and account for the effects of surplus content on model-based predictions.

It is particularly difficult to argue against the third conclusion. There is a tendency to believe that a large, comprehensive model must be correct, for how can it be incorrect if every detail has been included? Without sufficient empirical evidence, it is difficult to advance a cogent case for disputing such a belief. And it is typical of the analysis of relatively soft systems that opinions can count more strongly than incontrovertible empirical evidence. But cogent argument is precisely what is necessary, if the limitations in constructing large models are to be appreciated. Suppose the question is posed: can a large simulation model predict future behavior under substantially changed conditions in the real system? An answer that avoids the point of the question, yet an answer that is commonly encountered, might be one that denies prediction as an intended objective; instead, scientific understanding is the goal. The analyst who subscribes to such an attitude should not, by the same standards of judgement, argue against small, well-calibrated models on the grounds that their predictive capacities are limited. He must also be aware of the following possible conclusion. If scientific understanding is the goal, then presumably at some stage the model (the hypothesis) must be confronted with observations from the field system, and not from the laboratory system. Since the field observations are likely to be sparse and inadequate for such purposes, and since the design of a new experiment (laboratory or otherwise) may be a primary objective of evaluating the model against field data, the analyst has to be able to determine unambiguously which part of the model should be revised in order for scientific understanding to progress. This is, as the reader will conclude, nothing other than the problem of model structure identification, and it implies that rigorous and systematic model calibration cannot ultimately be avoided.

In answer to the original question, therefore, let us assume that scientific understanding is not the only goal, that the quest for the general, all-embracing model is an objective that some analysts cherish, and that one reason for this quest is indeed the desire and ability to make predictions. In which case, how is it possible to expose a prediction that may reflect a pattern of behavior which is a highly unlikely attribute of the real system, and which is probably a spurious artefact of the model? How, also, is it possible to begin debating, on a sound basis, the benefits and limitations of the

various approaches to modeling that are represented in our three conclusions on the subject of model calibration? In order to answer these questions, it is necessary to consider the problem of prediction in greater detail. For it is only in the context of prediction that the limitations of small and large models are thus revealed.

ACCOUNTING FOR UNCERTAINTY

Let us suppose that in an 'ideal' study, the problem of model structure identification has been solved and that it merely remains for calibration to be completed by estimation of the model parameter values. Now, recall the earlier interpretation of a parameter estimation algorithm as an information processing mechanism. After a successful calibration exercise it would be expected that the degree of uncertainty in any given parameter estimate would be less than the uncertainty associated with the prior estimate of that parameter value before calibration. The amount by which the uncertainty in the parameter estimate is reduced should be roughly consistent with the degree of relevance that that parameter, and its associated sector of the model's behavior patterns, has to the observed system behavior. The reduction in the uncertainty of the parameter estimates will also be approximately inversely related both to the number of field observations and to the levels of uncertainty and error associated with those observations. But the a posteriori estimates of the parameters will still be subject to uncertainty; their estimation errors are, as it were, a kind of 'fingerprint' of the calibration procedure; and the effects of these errors will propagate forward with predictions about the future.

In Fig. 2. therefore, let us assume that the set of behavior patterns M_1 belongs to a model characteristic of the class of large simulation models, the type of model that simulates a much greater variety of behavior patterns than has actually been observed in the historical field data, A (i.e. a large part of M_1 does not intersect with the set A). For such a model the many parameters not associated with those modes of behavior in the set A (i.e. that part of M_1 lying outside A) would have relatively large a posteriori estimation errors. In other words, and with other factors being equal, there was no information in the field data with which to reduce this uncertainty. Large a posteriori estimation errors reflect surplus content in the model; they may also reflect ambiguities in the model structure.

The complement, or opposite, of the large simulation model is a more compact kind of model, the 'fully-identified' model as it has been called earlier. If one is optimistic, this fully-identified model might be represented by the set M_2 in Fig. 2. Its a posteriori parameter estimates ought to be much less uncertain than many of those of M_1; and, since this model contains no surplus content, the set M_1 is contained completely in the set A.

How might these two models perform when applied to the problem of prediction? The most interesting case to consider is that in which future
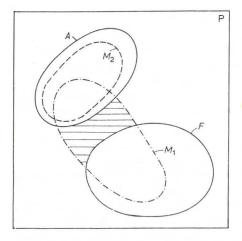
Fig. 2. Uncertainty and predicted behavior. P is the set of all possible behavior patterns; A is the historically observed pattern of behavior; M_1 is the behavior pattern simulated by the typical 'large simulation model'; M_2 is the behavior pattern simulated by the small, fully identified model; F is a set of possible future behavior patterns of the actual system.

input disturbances of the lake or river, such as different meteorological conditions and modified effluent discharges, force the variations of water quality into patterns of behavior (e.g. F in Fig. 2) quite different from the historically-observed patterns. Transactions of uncertainty are now operative from the model parameter uncertainties, the uncertainty in the estimated present state of water quality, and the uncertainty of future input disturbances, to the uncertainty in model forecasts of future response patterns. In its simplest form, the uncertainty of a forecast is understood in terms of the variance of the forecasting error. Thus, if the postulated input disturbances (which may themselves be highly unlikely events) were to drive the behavior of the large simulation model (M_1) into parts of the set F, we would expect the following. The predicted responses of the model should become relatively much more uncertain because relatively uncertain sectors of the model are being brought into play. And this is significant, for it says that there is no good empirical basis for expecting this kind of behavior. Moreover, the fact that we might believe one part of the model to be more accurate (certain) than another does not imply everlasting confidence in that part of the model. An uncertain quantity of zooplankton, whose behavior has not been well identified, preying upon an initially certain quantity of phytoplankton leads to an increasingly uncertain quantity of remaining phytoplankton.

In contrast, would a small model that captures only the dominant modes of past behavior (as does the model M_2 in Fig. 2) tend not to predict different future conditions? After all, its parameter values have been well identified and would thus be associated with relatively small estimation errors. Hence, given the kind of argument presented above, we might always be mistakenly confident about its predictions. There is, for example, no intersec-

tion between M_2 and F in Fig. 2, which suggests that F is outside the scope of behavior patterns simulated by M_2.

Hence, we arrive at the dilemma that is to be the terminal point of this discussion. With a large model (M_1) it may well be possible to predict the 'correct' future, but one would have little or no confidence in that prediction. With a small model (M_2) it may be that a quite 'incorrect' future is predicted, and, worse still, one might place considerable confidence in that prediction.

Of course, this dilemma has perhaps been stated in an exaggerated and overly simplistic fashion, but this was intended to give sharp definition to the problem and not to obscure the inevitable areas of grey between such a black-and-white statement of the problem. For example, consider the shaded portion of Fig. 2, where there are patterns of behavior simulated by M_1 that do not have any correspondense with past (A) and future (F) observations of reality. What confidence should be attached to predictions reflecting, in effect, this spurious content of the model? Indeed, it has been suggested elsewhere (Beck, 1980) that such behavior patterns might lie completely outside the frame of Fig. 2, that is, they lie outside the set P of all possible behavior patterns of the real system.

To begin with less complex issues, however, it is clear that the dilemma poses more problems for the analyst. Nothing has been said of the mechanics of undertaking analyses of the propagation of forecasting errors, although there is a growing body of literature on the subject (e.g. O'Neill and Gardner, 1979; Reckhow, 1979; Beck et al., 1979; Fedra et al., 1980). Nor is there evidence from case studies in calibration and prediction of how one could assess the performances of the two types of model on the basis suggested. In fact, to clarify what that basis is, let us point out that the analysis of prediction error propagation can be viewed as a kind of a posteriori sensitivity analysis. It provides a check on the relative levels of confidence associated with the assumptions made in developing, calibrating, and applying a model. Such an analysis should reveal when, and to what extent the model's predictions rely upon these assumptions and upon each component of uncertainty. It ought to be possible to distinguish among the effects of uncertainty propagated from surplus content in the model, the effects of unresolved ambiguities of calibration, and the effects of uncertainty associated with the extrapolation of knowledge about laboratory systems' behavior to knowledge about the field system's behavior. And in order to have relevance any method of analysis should be applicable to large, complex models, yet retain simplicity of use.

CONCLUSIONS

It is difficult to escape the suspicion that many analysts place too great a faith in their models. Is it possible that decision-makers are influenced by such faithful promotion of the model's forecasts? For there are undoubtedly serious dangers in disconnecting model application (to the problem of pre-

diction) from model calibration, and of separating forecasts from forecasting errors. Model calibration is, as we have described it, an exercise in reducing some of the uncertainties in the model and in discriminating against the unwanted effects of errors in the field data. When predicting the future it is equally important to account for the inevitable uncertainty in the calibrated model's behavior and to discriminate against the deceptive certainty of a forecast without its error.

The discussion of this paper has been set against the broader background of systems modelling. It is argued that water quality—ecological systems are in part hard systems and in part soft systems. Two schools of thought on the subject of water quality-ecological modeling (and there may be other schools of thought) can therefore develop and flourish. One school of thought works principally from the in-situ field data: it maintains that this is the only source of objective evidence for evaluation of model performance; it believes that the underlying relationships that govern the observed system behavior can be identified from those data; and for this purpose of system identification it draws upon analytical methods that have proven successful in their application to the analysis of hard systems. The onus of this school of thought is to demonstrate that it can, in fact, construct a 'meaningful' picture of the fabric of water quality-ecological systems. Here the term 'meaningful' has to be judged according to the principles of the second school of thought if a dialogue is to be established. The second school of thought works principally from what may be termed 'a priori knowledge', i.e. it admits extrapolations from laboratory-determined relationships to a model of the field system; it seeks to overcome the difficulties of unidentifiable surplus content in the model by relying upon previous experience; and it seeks fundamentally general 'laws' that will permit prediction of the future under substantially changed circumstances. The burden of this school of thought, in the event that rigorous calibration is not feasible, is to make only those predictions that allow inspection of the sensitivity of the prediction to the uncertainties of the inevitable surplus content and possible spurious content of the model. Otherwise, assessment of the confidence in the model will be reduced to the level of debates about strongly subjective opinions; a kind of debate, in other words, that relies upon the relatively soft nature of analysing water quality-ecological systems.

knew it or not, have contributed to the discussion of this paper. I am especially indebted to Gerrit van Straten, Peter Young, Dominic Di Toro and Efraim Halfon for their invaluable contributions.

REFERENCES

Beck, M.B., 1978a. Mathematical modeling of water quality. Int. Inst. Appl. Syst. Anal. Conf. Publ. No. CP-78-10, Laxenburg, Austria, p. 50.

Beck, M.B., 1978b. Random signal analysis in an environmental sciences problem. Appl. Math. Modelling, 2: 23—29.

Beck, M.B., 1979. Model structure identification from experimental data. In: E. Halfon (Editor), Theoretical Systems Ecology. Academic Press, New York, NY, pp. 259—289.

Beck, M.B., 1980a. Applications of system identification and parameter estimation in water equality modeling. In: Proc. Oxford Symp. Hydrological Forecasting, IAHS—AISH Publ. No. 129, pp. 123—131.

Beck, M.B., 1980b. Uncertainty, system identification, and the prediction of water quality. In: Proc. Task Force Meet. Uncertainty and Forecasting of Water Quality. Int. Inst. Appl. Syst. Anal., Laxenburg, Austria (in preparation).

Beck, M.B. and Young, P.C., 1976. Systematic identification of DO—BOD model structure. Proc. Am. Soc. Civ. Eng., J. Environ. Eng. Division, 102: 909—927.

Beck, M.B., Halfon, E. and van Straten, G., 1979. The propagation of errors and uncertainty in forecasting water quality. Working Paper WP-79-100. Int. Inst. Appl. Syst. Anal. Wkg. Pap. No. WP-79-100, Laxenburg, Austria, p. 54.

Bierman, V.J., Jr., Dolan, D.M., Stoermer, E.F., Gannon, J.F. and Smith, V.E., 1979. The development and calibration of a multi-class phytoplankton model for Saginaw Bay, Lake Huron. Environ. Prot. Agency (U.S.) Ecol. Res. Ser. (in press).

Fedra, K., van Straten, G. and Beck, M.B. 1980. Uncertainty and arbitrariness in ecosystems modeling: a lake modeling example (submitted to Ecol. Modelling).

Halfon, E., 1979. The effects of data variability in the development and validation of ecosystems models. In: B.P. Zeigler, M.S. Elzas, G.J. Klir and T.I. Ören (Editors), Methodology in Systems Modelling and Simulation. North-Holland, Amsterdam, pp. 335—343.

Jørgensen, S.E. and Harleman, D.R.F., 1978. Hydrophysical and ecological modeling of deep lakes and reservoirs. Int. Inst. Appl. Syst. Anal. Conf. Publ. CP-78-7, Laxenburg, Austria, p. 38.

Karplus, W.J., 1976. The future of mathematical models of water resources systems. In: G.C. Vansteenkiste (Editor), System Simulation in Water Resources. North-Holland, Amsterdam, pp. 11—18.

Maciejowski, J.M., 1979. Model discrimination using an algorithmic information criterion. Automatica, 15: 579—593.

Mankin, J.B., O'Neill, R.V., Shugart, H.H. and Rust, B.W., 1977. The importance of validation in ecosystem analysis. In: G.S. Innis (Editor), New Directions in the Analysis of Ecological Systems. Simulation Council Proc. Ser., 5: 63—72.

O'Neill, R.V. and Gardner, R.H., 1979. Sources of uncertainty in ecological models. In: B.P. Zeigler, M.S. Elzas, G.J. Klir and T.I. Ören (Editors), Methodology in Systems Modelling and Simulation. North-Holland, Amsterdam, pp. 447—463.

Patten, B.C. (Editor), 1975. Systems Analysis and Simulation in Ecology, Vol. 3. Academic Press, New York, NY, p. 622.

Patten, B.C. (Editor), 1976. Systems Analysis and Simulation in Ecology, Vol. 4. Academic Press, New York, NY, p. 608.

Reckhow, K.H., 1979. The use of a simple model and uncertainty analysis in lake management. Water Res. Bull., 15: 601—611.

Russell, C.S. (Editor), 1975. Ecological Modeling in a Resource Management Framework. Resources for the Future, Wkg. Pap. QE-1. Johns-Hopkins University Press, Washington, DC.

Scavia, D. and Robertson, A. (Editors), 1979. Perspectives on Lake Ecosystem Modeling. Ann Arbor Science, MI.

Thomann, R.V., 1978. Size-dependent model of hazardous substances in the aquatic food chain. Report EPA 600/3-78-036. Environ. Prot. Agency (U.S.), Washington, DC.

Thomann, R.V. and Winfield, R.P., 1976. On the verification of a three-dimensional phytoplankton model of Lake Ontario. In: Proc. Conf. on Environmental Modeling and Simulation, Environ. Prot. Agency (U.S.), Washington, DC Rep. EPA 600/9-76-016, pp. 568—572.

Vansteenkiste, G.C. (Editor), 1975. Computer Simulation of Water Resources Systems. North-Holland, Amsterdam, p. 686.

Vansteenkiste, G.C. (Editor), 1978. Modeling, Identification and Control in Environmental Systems. North-Holland, Amsterdam, p. 1028.

Whitehead, P.G., 1979. Application of recursive estimation techniques to time-variable hydrological systems. J. Hydrol., 40: 1—16.

Young, P.C., 1978. General theory of modeling for badly defined systems. In: G.C. Vansteenkiste (Editor), Modeling, Identification and Control in Environmental Systems. North-Holland, Amsterdam, pp. 103—135.