ENVIRONMENTAL MODELING UNDER UNCERTAINTY: MONTE CARLO SIMULATION

K. Fedra
International Institute for Applied Systems Analysis, Laxenburg, Austria

RR-83-28
November 1983

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

International Standard Book Number 3-7045-0061-5

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Cover design by Anka James

Printed by Novographic, Vienna, Austria

FOREWORD

In recent years, there has been considerable interest in developing models for environmental systems, and for aquatic systems in particular. Much of this effort has been directed toward large and complex simulation models. However, this trend has given rise to a number of concerns, notably those of accounting for the effects of uncertainty. Testing model structures, calibrating complex simulation models under uncertainty, and propagating this uncertainty in the predictions of models are essential steps in establishing model validity and credibility for practical applications.

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

This report, based on a series of earlier papers on the subject, discusses the use of Monte Carlo methods when the available field data are sparse and uncertain. It examines the problem of constructing, calibrating, evaluating, and applying a model for prediction – and ultimately for management (K. Fedra (1980) Mathematical modelling – a management tool for aquatic ecosystems? Helgoländer Meeresuntersuchungen 34:221–235, also reprinted as IIASA Research Report RR-81-2). In particular, it emphasizes the importance of model testability (K. Fedra (1981) Hypothesis testing by simulation: an environmental example. IIASA Working Paper WP-81-74) and the close relationship between the processes of model calibration and the predictions obtained subsequently (K. Fedra, G. van Straten, and M.B. Beck (1981) Uncertainty and arbitrariness in ecosystems modelling: a lake modelling example. Ecological Modelling 13:87–110, also reprinted as IIASA Research Report RR-81-26).

Thus, uncertainty and the reliability of models and forecasts based on Monte Carlo simulation are the key concerns of this report.

Janusz Kindler
Chairman
of the former
Resources and Environment Area

CONTENTS

	SUMMARY	1
1	INTRODUCTION	2
	1.1 Monte Carlo Methods: Computerized Trial and Error	3
	1.2 The Theoretical Framework: Models, Knowns, and Unknowns	4
	1.3 Model Structure, Parameters, Inputs, and Observations: Some Implications of Uncertainty	5
2	THE METHOD	7
	2.1 The Concepts of Behavior Space and Model Response Set: Defining a Problem-oriented Objective Function	7
	2.2 The Concept of Model Parameter Space	9
	2.3 A Formal Presentation of the Method	11
	2.4 A Very Simple Illustrative Example: Estimating Regression Coefficients	13
	2.5 Some Technical Details	16
3	APPLICATION EXAMPLES	19
	3.1 Hypothesis Testing: A Marine Pelagic Food-web Example	19
	3.2 Estimation and Prediction with Parameter Ensembles: A Lake Modeling Example	36
	3.3 The Limits of Estimation: A Simple Rain–Runoff Model	50
4	DISCUSSION	61
	4.1 A Philosophical View: Pragmatic Instrumentalism	62
	4.2 Uncertainty Analysis: Alternative Approaches	64
	4.3 Consequences and Implications: Uncertainty and Forecasting	68
	REFERENCES	69
	BIBLIOGRAPHY	73

ENVIRONMENTAL MODELING UNDER UNCERTAINTY: MONTE CARLO SIMULATION

K. Fedra
International Institute for Applied Systems Analysis, Laxenburg, Austria

SUMMARY

The study of environmental systems as ecological and physicochemical as well as socioeconomic entities requires a high degree of simplifying formalism. However, a detailed understanding of a systems function and response to various changes for the explicit purpose of systems management and planning still requires fairly complex hypotheses, or models. Such models can hardly be subjected to rigorous tests without the aid of computers. Systems simulation is a powerful tool when subjecting complex hypotheses to critical tests of their logical structure and their performance over the range of plausible input conditions.

Based on a formalized trial-and-error approach using Monte Carlo methods, this report presents and discusses an approach to simulation modeling under uncertainty. An introduction to the causes and implications of the problem, namely uncertainty, and a short formal presentation of the methodology proposed are followed by some more technical remarks on Monte Carlo simulation. Using three different application examples, the author discusses the role of uncertainty in the formal testing of model structures, in parameter estimation, and in prediction. In the last example, the limits of estimation and, with it, prediction are demonstrated. In a comparison of Monte Carlo simulation with alternative approaches to including and evaluating uncertainty in simulation modeling, the discussion section examines the implications of uncertainty for model application in a broader framework.

1 INTRODUCTION

Environmental modeling may conveniently be understood as a tool – a tool for the study of systems that are large, complex, difficult to observe, and experimentally more or less inaccessible. It is a formal way of organizing knowledge (or the lack thereof) at the intersections of ecology and the life sciences, geography and the earth sciences, the social and political sciences, economy and engineering, and usually a few more of the classical disciplines.

Environmental modeling and simulation is also a tool for developing and testing the hypotheses on which any organization of knowledge is based, and is therefore just one instrument of scientific research. This tool may be used for making "predictions," experiments with possible futures, exploring alternative courses of action. It thus has potential to aid management and decision making and to help design and explore policies.

In the core of any comprehensive environmental system, there is usually an ecological system or an ecosystem in the more classical sense (Haeckel 1870, E.P. Odum 1971); and a close look at the kinds of data that are available on ecosystems shows mainly uncertainties, variability, and sampling errors (more often than not of undetermined magnitude). In addition, ecological theory (and whatever part of it may be relevant within the more comprehensive framework of environmental science) is full of contradictory hypotheses, and it is mostly impossible to rule out any of those because of lack of reliable and sufficient data. Consequently, the coexistence of competing and eventually contradictory model formulations (contradictory in the sense that they will produce significantly different predictions from the same set of inputs) is notorious. A nice illustration is given by Simons and Lam (1980), when they observe in their critique of models used in the Great Lakes studies that "these results illustrate quite clearly that one can accommodate a wide range of primary production formulations in a model as long as there are additional degrees of freedom to 'play with,' in this case the uncertainty associated with respiration and other forms of nutrient regeneration." This phenomenon, by the way, can also be observed in the social or political sciences as well as in economics, which, unfortunately but significantly, are also basic components of applied environmental research.

Experimental evidence, as a rule, stems from microscale physiological approaches, contradictory in their very design to the richness and variety of ecosystems, and deliberately neglecting a main feature of any even moderately complex ecosystem, which is the simultaneous interaction of large numbers of variables. Traditional concepts and approaches are merely extrapolations of ideas that proved to be successful in physics and chemistry. However, ecosystems are quite different from electrical networks, the frictionless pendulum, and controlled chemical reactions of some compounds. All these incompatibilities can seemingly be overcome only with numerous more or less arbitrary assumptions, often enough implicitly hidden in a hypothesis, or model formulation. The information available is of a jigsaw puzzle structure, and at best we can deduce fuzzy patterns, semiquantitative relationships, ranges, and constraining conditions, unless we blindly believe in numbers once they are printed, preferably by the computer.

Chance, or random variability, plays an important and sometimes dominant role in environmental systems. This is true not only for the microscopic, elementary level (Monod 1970), but also for living, evolving, dissipative systems and structures in general (e.g. Eigen and Winkler 1975). All these features, including the consequences of haphazard human interference, contribute to one prominent aspect of environmental systems and thus modeling: uncertainty. Clearly, under these circumstances the applicability of traditional, fully deterministic techniques, with all their implicit and explicit assumptions on the distributions and functional properties of the variables observed (or rather sampled), and a firm belief in numbers have to be questioned. Forcing environmental systems into a mathematical framework developed for vastly different systems, for the sake of the ease and elegance of the analysis, seems to me not only a futile but also a dangerous line of work. And as a consequence, many model-based predictions on environmental systems are either trivial or false or, at best, computerized intuition of the analyst.

Alternative approaches are needed, if environmental modeling is to improve its so far meager record of impact on environmental decision making and public reasoning. One possibility is a formal and computer-based application of probably the simplest and most straightforward approach, but maybe also the only possible approach to scientific research: trial and error.

1.1 Monte Carlo Methods: Computerized Trial and Error

"Our whole problem is to make the mistakes fast enough ..."
(Wheeler 1956)

Monte Carlo methods, as used and discussed in this report, are nothing more than computerized trial and error. It is a technique, however, to make extremely high numbers of errors, and to make them very fast – and, it is hoped, to learn from these errors. As indicated by the name, it is a form of gambling – picking random numbers from appropriate distributions and using them for numerous trials (and errors). A system of filters is then used to separate the solutions – if there are any winning numbers – from the failures.

The method is characterized by a very appealing simplicity. This may be best exemplified by the fact that this report is written by an ecologist, not a mathematician. No implicit, abstruse statistical assumptions have to be made, either on the available data describing the system to be modeled, or on the concept of agreement or "goodness of fit" between model output and the observations modeled, which is the deviation or error to be minimized in "classical" approaches (Section 4.2). Arbitrary assumptions have to be made, like in all other approaches, but the simplicity of the method allows for an explicit statement and treatment of all the assumptions. None of the assumptions are hidden within the method, they can all be made "externally." A high degree of flexibility in constructing an appropriate estimation scheme for a given application problem allows one to structure the tool according to the

problem – and not force the problem into the constraints of the method.

Any simulation model can, with a minimum amount of programming skills, be easily incorporated into an appropriate framework for the Monte Carlo estimation, including the generation of trial runs, their monitoring, and the most crucial part, the evaluation of the trials. The model can be as complex and nonlinear as deemed necessary by its builder, and there is no limit, in principle, to the number of parameters for simultaneous estimation.

The price for all these advantages has to be paid in terms of computer time: excessive trial and error, when done simply (blindly and "unintelligently," i.e. without learning from the errors within a series of trials), requires a comparatively large amount. In addition, the time requirements grow exponentially with the dimensionality of the problem, that is, the number of parameters estimated simultaneously. Computer time, however, is becoming cheaper and cheaper, and in many cases is no real constraint for the analysis, as compared with, for example, the much more demanding and expensive collection of field or laboratory data.

1.2 The Theoretical Framework: Models, Knowns, and Unknowns

Some conceptual clarifications seem to be unavoidable in order to introduce the terminology used in the following sections. Calibration, in a nontechnical definition, is the tuning of a model in order to improve the agreement of model-generated output with the observations from the system to be modeled. Tuned or adjusted are coefficients describing the relationships between the model elements, i.e. state variables, inputs, and outputs (the boxes and cycles in flow diagrams), and auxiliary values such as thresholds, carrying capacities, stoichiometric constants, or any other "adjustable" values. If a model deals with "simple" systems and well established laws of nature, no tuning should be necessary, since all the parameters required are well known constants. If we want to model the fall of a pebble, we certainly would not attempt to calibrate the constant of gravity, but would take it from the literature.

In epistemological terms, the modeling process involves:

a. a theory or universal statement (the model structure), together with
b. a set of initial conditions (the initial conditions *sensu stricto*, i.e. the state of the elements of the system at time $t=0$; the parameters, i.e. measures quantitatively describing the relationships of these systems elements and any auxiliary coefficients; and, in the case of dynamic models, inputs into the system, or forcings or driving variables, which can be viewed as a time series extension of a certain subset of the initial conditions), to derive
c. a set of singular statements (the model output), which then has to be compared with appropriate observations.

In a pragmatic (ab)use of the usual terminology, I will split the union set of parameters, initial conditions, and inputs (forcings) into two

complementary subsets, namely the "knowns" (e.g. site constants, such as the volume of a lake or the length of a river reach, or any number in which we can place enough confidence to consider it "known") and the "unknowns." The latter have to be estimated, and will, for simplicity, be referred to as parameters; the "knowns" I will call constants.

1.3 Model Structure, Parameters, Inputs, and Observations: Some Implications of Uncertainty

If a system or process to be modeled is well known, as, for example, in classical mechanics, if the initial conditions can be manipulated or observed without error, and if the elements of the system and thus the outcome of an experiment can be observed directly and without (or with very small) error, calibration would, if at all necessary, be a simple undertaking. One could, to exploit a simple example given by Popper (1959), calibrate a material constant for a thread. However, one would rather call this process a direct experimental determination of the magnitude in question, as it can usually directly and analytically be inferred from the experiment. If, however, the required value for the material constant would have to be found by iteration, one might call this calibration.

In environmental modeling, however, the problems are much more muddled and diffuse, and we have neither a well established theoretical framework (allowing us to set up an indisputable model structure *a priori*) nor known constants. Even the observations available *in situ* or from experiments are difficult to use, since they are generally made on a different level of complexity and on a different scale than used in our models. There are several generic problems associated with ecological modeling, or any large-scale modeling of systems and processes that are complex, difficult to observe, and almost impossible to manipulate.

The first and probably most important problem is in the discrepancy between the scale of model conceptualization and the scales of measurement, observation, and experimentation. Our knowledge of large and heterogeneous systems is always derived from "samples," and even these samples, generally associated with a certain error, are always ranges. Observations and experiments are usually made on a micro-scale, involving individual cells, monospecific cultures, or extremely small samples from the system (just consider the proportion of the volume of a sampling bottle to that of a lake). There exists, of course, a well established theory of sampling, and statistics will tell the observer or experimenter how many and which size of samples should be drawn to reach a certain level of confidence for the resulting estimates. However, for reasons that can only partly be attributed to logistic problems and resource limitations, sampling statistics seem to be one of the most neglected fields in ecological research.

A somewhat different interpretation of the discrepancy between theory and observations – anathema to the pure empiricist – could be a claim that the relevant observational and experimental techniques are just insufficient or unreliable (e.g. Feyerabend 1975, Lakatos 1978, and Section 4.1). Empirical

evidence and theory can eventually be even incommensurable.

The units dealt with in formal conceptualizations of environmental systems, i.e. the models, on the other hand, are usually large, lumped, and inaccessible to direct experimentation. They are idealized functional entities, whereas experiment and observation usually concentrate on entities that are systematic (in the biological or chemical sense). The units in the models are lumped and heterogeneous, such as "primary producers," "zooplankton," or "available nutrients." Therefore, their functional characteristics, described by the "parameters," can only crudely be estimated from the eventually measurable characteristics of their elements, e.g. an individual species (ignoring the additional complications of age groups, sexes, physiological states, etc.). As these functional attributes cannot be measured directly, and there is no way of reliably deriving them from the properties of the microscale components, they have to be calibrated, i.e. adjusted to values that result in an acceptable performance of the model. Such heterogeneous assemblages tend to exhibit a fairly, and sometimes surprisingly, simple behavior. This phenomenon, often referred to as the "linear" response of highly nonlinear systems (in terms of their microelements), allows one to treat such heterogeneous elements as functional units.

It is important to recognize that neither model structures, nor initial conditions, inputs, and parameters, nor the observations used as the testing ground for a model are without error. They are all uncertain, usually to an uncertain degree, and all ought to be formulated in terms of ranges or probability distributions. Parameter estimation, as a consequence, is mostly an art. Seemingly exact approaches that reduce the problem to the minimization of an objective function are based on numerous simplifying and often implicit arbitrary assumptions. Since almost everything, including the reference values (the observations) used for calibration, is somewhat fuzzy and error-corrupted, derived from subjective interpretation of information rather than indisputable measurements and experimental design, an exact and "best" solution to the parameter estimation problem is only obtained when at least parts of the uncertainty are ignored, thereby reducing the number of unknowns, although in a disputable and arbitrary fashion.

Both parameters and model structure are uncertain, and intimately depend on each other. Their estimation should therefore be made concurrently. This will be demonstrated in the first application example (Section 3.1), based on a marine pelagic food-web simulation for the German Bight in the southern North Sea. This example illustrates the close dependency of parameter estimates on the model structure chosen and, vice versa, attempts to show how parameter space characteristics can be utilized to modify a model structure.

In the next step, the simple application of Monte Carlo methods for parameter estimation can be extended for predictions. Obviously, predictions and especially prediction uncertainty will depend on model and parameter uncertainty. The second example of application (Section 3.2), based on a lake water quality model, demonstrates how the uncertainty in the parameter estimates obtained by Monte Carlo estimation can be preserved, and included in the predictions, in order to estimate the reliability of predictions.

Finally, in a third application example, the interdependence between parameter estimates and performance criteria, or objective function (which is derived from the available observations) used in the estimation procedure, will be shown (Section 3.3). By use of a simple example based on a rain–runoff model, two alternative parameter vectors, both minimizing plausible objective functions but resulting in quite different model behavior, can be generated. These obvious limits to calibration can only be resolved with additional information from the system, that is to say, with an additional set of (specific) observations.

2 THE METHOD

The basic principle of Monte Carlo methods, as used and discussed here, is a trial-and-error procedure for the solution of the inverse problem, i.e. estimating the "unknowns" in the input of the model (the parameters) from the required output. Since complex dynamic simulation models cannot be solved analytically, the solution of the inverse problem demands a more complicated procedure.

The basic steps of this estimation procedure are as follows (Figure 1): for a given model structure, performance criteria describing the expected, satisfactory behavior of the model, based on the available data, are formulated. For all the unknowns to be estimated, allowable ranges or probability density functions are defined. From these ranges or distributions a sample vector is drawn randomly and substituted in the model for one trial run. The performance criteria of this trial run are then compared with, or classified according to, the predefined target values or ranges of the performance criteria. The process is then repeated for a sufficient number of trials. After some initial trials and their analysis, the ranges to be sampled may be redefined, criteria may be added or deleted, or the model structure changed. This whole process is repeated iteratively until the model performance is satisfactory, in light of the original problem to be solved, or until the user's computer account is exhausted.

2.1 The Concepts of Behavior Space and Model Response Set: Defining a Problem-oriented Objective Function

From a model run, a simulation, one obtains a vector of output values, a singular statement, or prediction, which has to be testable, i.e. comparable (and compared) with corresponding observations from the system in order to determine whether or not the model (and its parameter set) is acceptable under the constraints of the predefined performance criteria.

If one recognizes that the entities used in a simulation model and those measured in the field or in a laboratory experiment are quite different, it is obvious that they cannot be compared directly, and then used to estimate one from the other. One has to take into account the differences in scale and aggregation, and the resulting uncertainties. Models, because of their high

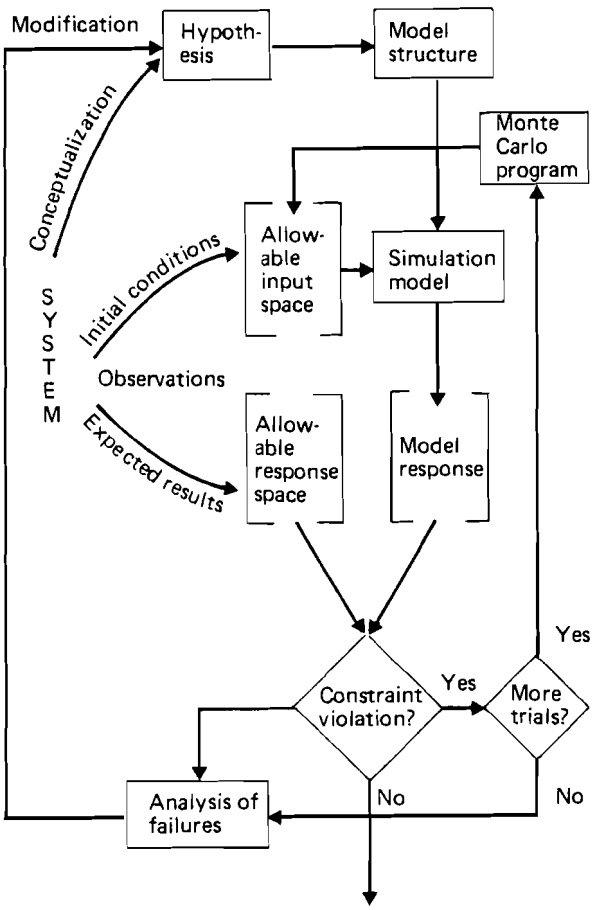

FIGURE 1 Flow diagram of the approach.

degree of abstraction, simulate average patterns or general features of a system (as conceptualized in the model). These patterns have to be derived from the available information at an appropriate level of abstraction and aggregation. Only such derived measures can then be compared with the magnitudes generated with the model, in order to test and improve model performance.

The original set of observations of the system to be reproduced by the model output can conveniently be thought of as a region in an n-dimensional behavior vector space. Clearly, each observable property of the system can form one dimension. Time, in the case of dynamic systems and models, can be thought of as just one attribute of an observation; that is, algal biomass at a certain time, say spring turnover of a lake, might form one dimension, and algal biomass at another time, say summer solstice, could be another. Also,

observable properties could be independent of time, such as algae biomass maximum, whenever it was observed during the year. Another class of observable properties comprises integrated properties, such as total yearly primary production, or relational properties, such as the ratio of maximum to minimum algal biomass. Each of these properties – and many more, certainly depending on the kind of system in question – can be used in defining the behavior space of the system, and with it, as a subset, the set of desired, "realistic" model responses.

Obviously, the great flexibility in these constraint conditions allows for tailoring a very detailed and problem-specific set of constraint conditions. Violating none of them can be understood as analogous to minimizing an objective function. Besides criteria that can be easily and directly derived from the available set of specific observations on a given system, one might want to constrain more and other elements of model response, such as flows and relationships between integrated flows or efficiencies in ecological jargon. Since such magnitudes are usually not observed, one would have to resort to the ecological or environmental literature for appropriate ranges. However, such additional constraints can only help to rule out ecologically or physically implausible behavior of a model, but not to identify the parameters for a given, specific system as such.

The concepts of system behavior space and model response set are quite versatile and, in fact, can even accommodate measures such as the sum of squares of deviations of model output from corresponding observations. A traditional squared-error criterion can be understood as a measure of distance, in the response vector space, between any singular model response and the required target, the behavior region of the system. The latter, however, is represented by a singular point (Section 4.1 includes a discussion of the different concepts and their relationships).

Along each of the relevant dimensions of the behavior space, the set of available observations can now be used to define a range, or a probability density distribution, within which the (observed) state of the system was found and, consequently, within which the simulated state ought to be. Each of the ranges in the model response space therefore constitutes a constraint condition imposed on an allowable model output. The defined allowable model response set can be understood as a filter that will separate the class of all model responses into allowable ones – contained in the allowable model response set – and its complementary "unrealistic" subset (Section 2.3). Figure 2 gives an example of projections of model response on to planes of two constraining response variables, with each allowable range forming a darkened rectangle in the projection plane.

2.2 The Concept of Model Parameter Space

Similar to the behavior vector space and response set associated with the output side of the model, one might conceive an input or parameter vector space on the input side. Each of the unknowns to be fed into the model for a simulation run again defines one dimension in this vector space. The

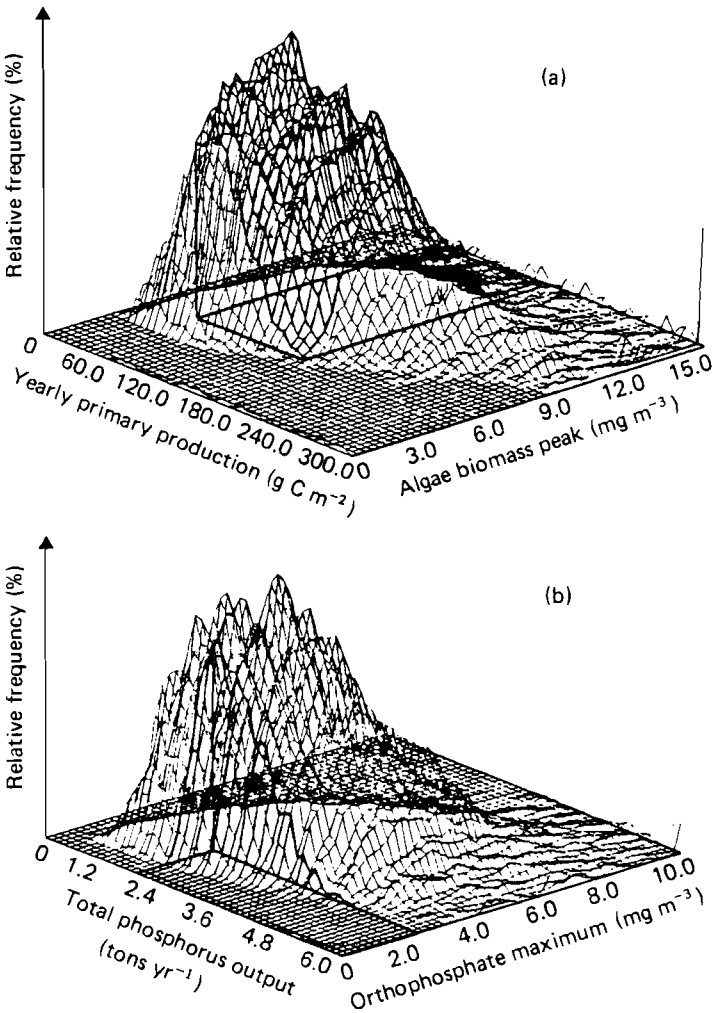

FIGURE 2 Model response space projection on to a plane of two response variables, indicating the position of the constraint conditions for (a) a pair of uncritical conditions, (b) a pair of critical conditions.

allowable values of this unknown define a range or probability density function on each of the coordinate axes. To define such ranges requires that each of the unknowns is physically meaningful, or measurable in principle, so that such a finite range will exist. Only if all the unknowns (and the classical parameters in particular) have a physical function that can be interpreted well, can they be reasonably constrained. The ranges within which a certain parameter has to be are – in the worst case – given by physical limits, e.g. a lower limit of zero for most rate constants, or an upper limit of one for a

limiting factor. In many cases, however, one will be able to find parameter values in the appropriate literature (e.g. Jørgensen *et al.* 1978), or even information from specific experimentation or observations from the specific system modeled. They can all be utilized to define allowable ranges for the unknowns.

The rationale for defining these ranges as narrowly as can be justified, without too much arbitrariness, is twofold. On one hand, narrow ranges increase the sampling density and reduce the number of trials necessary to explore the parameter space sufficiently. On the other hand, if no satisfactory solution can be found within the ranges deemed plausible *a priori*, this will indicate that the model does not function as the analyst thought it would. Obviously, the parameters do not influence the model behavior as assumed, they function differently from the analyst's perception of their function; in other words, there is something wrong with the model structure (Section 3.1).

Besides such straightforward information to be derived from the relationships between parameter set and response set, the subregions of the parameter space corresponding to certain subregions in the response space can give valuable insight into the model behavior. Parameter correlations, or any structural properties of parameter space regions for a certain class of response, can be interpreted in terms of a sensitivity analysis (Section 3.2). Figure 3 gives examples of projections of parameter space regions with certain response characteristics on to planes of two parameters.

2.3 A Formal Presentation of the Method

Let us suppose that a given model structure is assumed. The model can be represented by a vector function $\boldsymbol{f}$ with domain $D(\boldsymbol{f})$ and range $R(\boldsymbol{f})$. If RD is a subset of R, then the inverse image of RD under $\boldsymbol{f}$ is the subset of $D(\boldsymbol{f})$ given by

$$\boldsymbol{f}^{-1}(RD) = \{\boldsymbol{x}:\boldsymbol{f}(\boldsymbol{x}) \in RD\}$$

This subset will be called PM, and represents the set of all parameter vectors resulting in the defined, acceptable model responses RD.

To identify PM, we have to define RD by a series of constraint conditions, which can include more classical objective functions, e.g. a least-squares criterion (Section 3.1). From the plausible ranges for each of the parameters to be estimated, the set of allowable parameter vectors, PD, is formed as the direct or Cartesian product. Random samples are then drawn from PD, forming a set of trial parameter vectors. Each of these vectors is used for one trial run of the model, and the resulting model response is classified according to the set of constraint conditions into those giving the defined behavior:

$$RS' = \{RS_i:(RS_i \in RD)\} \qquad n(RS') = M$$

and those violating at least one of the constraint conditions, thus not giving the defined behavior:

$$RS'' = \{RS_i:(RS_i \notin RD)\} \qquad n(RS'') = N - M$$

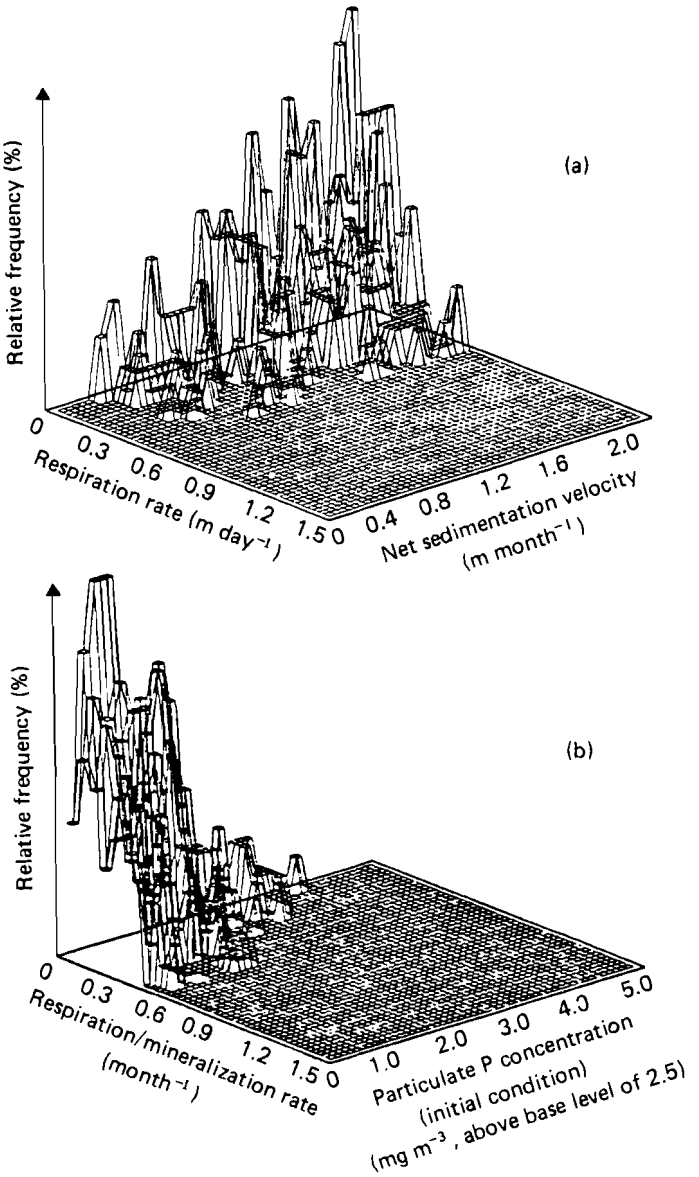

FIGURE 3 Parameter vector space projection for a behavior-giving set of parameter vectors. Projection from the 22-dimensional parameter vector space on to a plane of two model parameters; extension of the individual axes indicates the range used for sampling.

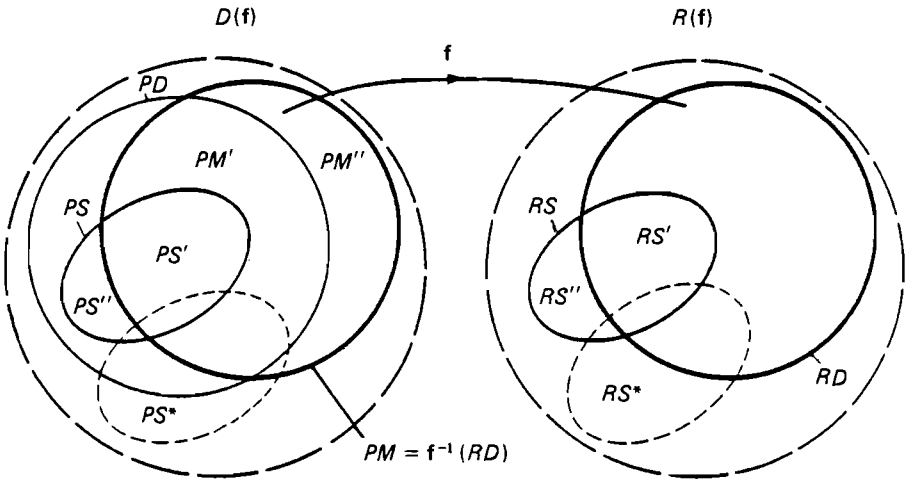

FIGURE 4 Set diagram of the relationships between parameter space and response space. $D(\boldsymbol{f})$ is the set of all possible parameter vectors (domain of $\boldsymbol{f}$); $R(\boldsymbol{f})$ is the set of all possible model responses (range of $\boldsymbol{f}$); $\boldsymbol{f}$ is the model (vector function); *PD* is the defined set of plausible parameter vectors; *RD* is the defined realistic response region; *PM* is the inverse image of *RD*; *PS* represents the character vectors sampled in the Monte Carlo procedure; *RS* is the direct image of *PS*; *PS'* is the subset of *PS* that generates plausible realistic response *RS'*; *PS''* is the subset of *PS* resulting in an unrealistic response *RS''*; *PS** is the modified *PS'* used for prediction, resulting in *RS**.

The N parameter vectors used for the trials are thus split into the complementary subsets *PS'* and *PS''* with M and $N-M$ elements, respectively. The set of parameter vectors *PS'*, resulting in the defined model behavior, is then the solution to the estimation problem. It is a subset sampled from the parameter space region *PM*. These relationships are summarized in Figure 4.

2.4 A Very Simple Illustrative Example: Estimating Regression Coefficients

To illustrate the method very simply, let us consider a data set (Figure 5(a)), with only one dependent state variable (y) plotted as a function of an independent one, which could, for example, be time. Let us also assume that *a priori* information about the system represented allows us to construct a model for it. To make the example as simple as possible, I will propose a model of the form

$$y(t) = at$$

(the reader might try to find a meaningful ecological example for this) with only one parameter (a), to be estimated from the data. Let me assume that, for reasons of "ecological plausibility," a can be constrained to the range

$$0.5 < a < 2.5$$

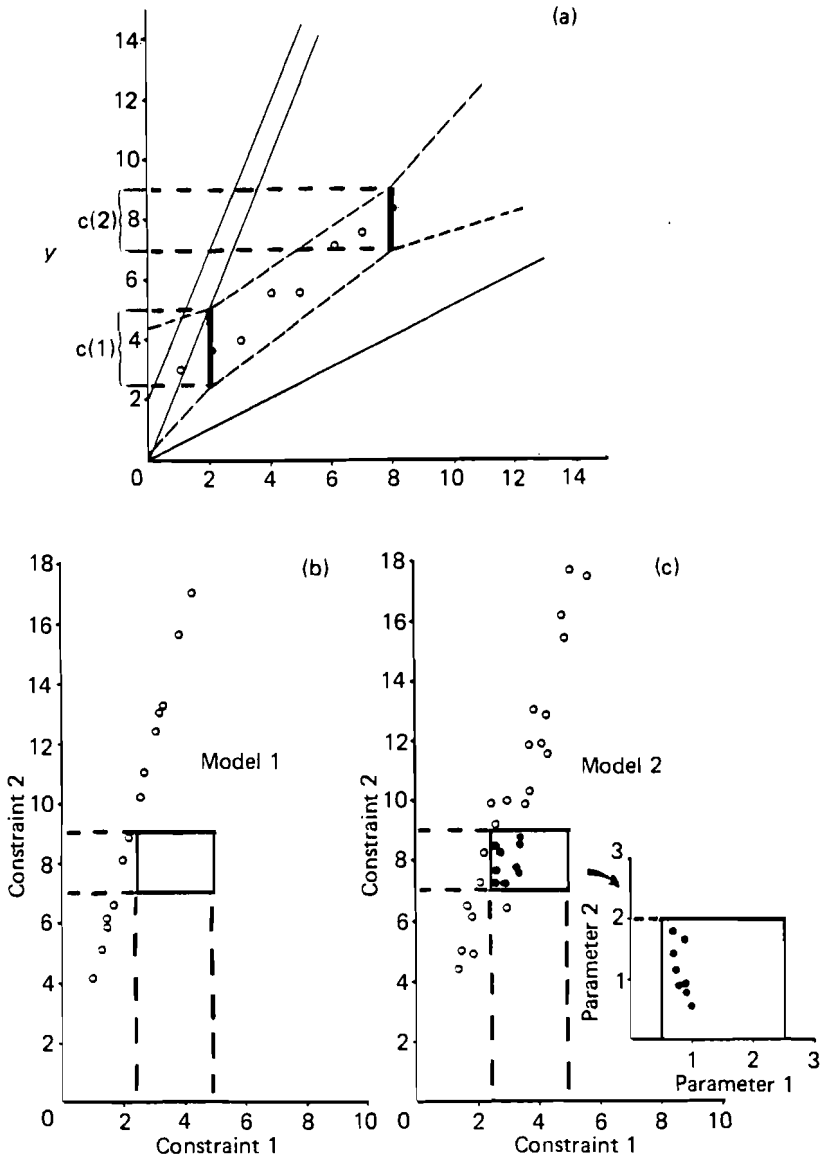

FIGURE 5 (a) Data set indicating the positions of the constraint conditions c(1) and c(2); thin lines represent envelopes over the responses of models 1 and 2; broken lines show allowable model response for model 2 (note the divergence outside the constraining bars). (b) Projection of model 1 response space, defined by the two constraint conditions; the box delimits the allowable response. (c) Projection of model 2 response space, indicating the positions of successful trials in the parameter space; the parameter box represents the ranges sampled.

which, in fact, defines a region in the one-dimensional parameter space or a set of plausible, allowable a's. On the output side, we can formulate a number of constraint conditions or performance criteria based on the data in Figure 5(a), which explicitly include the uncertainty around the observations (thought of as samples from a real-world system), indicated by the bars extending from two of the points. (The points without bars represent the typical singular observations or measurements without replica, which are somewhat difficult to interpret.)

The constraint conditions or performance criteria c(i) to be met are summarized as:

c(1): $2.5 < y(2) < 5.0$

c(2): $7.0 < y(8) < 9.0$

According to the terminology introduced above, the two criteria define a region in a two-dimensional behavior space (Figure 5(b)), or a set of allowable model responses.

To estimate values for a, we simply draw random samples from the interval defined around a, substitute these values in the model, "simulate" for the desired range of the independent variable, and determine the values of the two performance criteria, namely the values of y at t=2 and t=8. To no surprise of the reader, none of the values of a that can be sampled from the predefined interval will satisfy both of the constraint conditions. Consequently, the model will be rejected. Similarly, other one-parameter alternatives to the proposed model, namely

$$y(t) = t^a$$

and

$$y(t) = e^{at}$$

also fail to meet the constraint conditions imposed on the model output. Modifying the model by introducing one more parameter will lead to an alternative two-parameter model, which is then subjected to the same test and estimation procedure. The simplest form of the model would be

$$y(t) = at + b$$

with a simple additive second parameter (b). This could be thought of as, for example, the initial state of y at $t = 0$, which, in the first models, was implicitly forced to take values of 0 and 1, respectively. Constraining b to the range

$$0 < b < 2$$

we repeat the sampling and simulation procedure. This time, some of the simulation runs will meet both constraint conditions (Figure 5(c)). As can also be seen from Figure 5, the corresponding parameter vectors are found clustered in the parameter space region sampled randomly. The two parameters are clearly correlated (Table 1), indicating their mutual dependency or, in other words, the fact that a change in one of them can, within a certain

range, be balanced by a corresponding change in the other. Table 1 summarizes some basic statistics of the parameter subset PS' (Figure 4), i.e. the subset resulting in acceptable model behavior.

TABLE 1 Parameter statistics and prediction (example); 5000 runs evaluated.

	Range sampled	Mean	Min.	Max.	Standard deviation
Parameter a	0.500 – 1.500	0.85	0.630	1.080	0.097
Parameter b	0 – 2.000	1.33	0.340	2.000	0.407
Prediction (t 12)		11.45	9.54	13.30	0.924
Correlation matrix					
	1		2		
2	Significant ($p < 0.05$) negative correlation				
3	Significant ($p < 0.05$) positive correlation		Significant ($p < 0.05$) negative correlation		

In a final step, the set of allowable parameter vectors can now be used for predictions of y, for example, for $t=12$. A set of estimates results. If enough vectors are used, a frequency distribution or a probability density function can be constructed for the prediction, allowing for a probabilistic interpretation (Figure 5(a)). The variability of the parameters results directly from the uncertainty in the observations, and is again reflected in the output variability.

2.5 Some Technical Details

One of the major drawbacks of Monte Carlo methods is their insatiable demand for computer time. Although they are very efficient in terms of the time required by the analyst or modeler to set up an appropriate scheme for estimation and evaluation, this efficiency is traded against computer time and, eventually, storage capacity.

There are a few basic rules that can help to make Monte Carlo techniques more efficient in terms of computer use.

a. *Minimize the number of trials*

A reduction of the number of trial runs necessary to identify a set of parameter vectors for a certain class of model response can be achieved in several ways. First, a given estimation problem can be split into several cycles of trial runs in an iterative way. Each cycle is analyzed before the next one is started. This eventually allows corrections to be made, the ranges that are to be sampled to be redefined, constraint conditions to be modified, etc. After a relatively small number of trial runs (which certainly will depend on the number of unknowns estimated simultaneously) one might, for example, find a clear clustering of the "good" vectors in the parameter space already. If, consequently, certain regions in the parameter space seem "empty" (in

terms of solutions), they can be discarded (by redefining the ranges sampled) to improve the efficiency of the sampling. Another example would be constraint conditions, which are always violated. This should lead to the reconsideration of these conditions and the parameter ranges sampled (here they might have to be extended), or a modification of the whole model structure itself. Clearly, if after a first screening of the parameter space all model responses are off their target in a systematic way (as in the example above), an increase in the number of trials will probably not be worth while.

Some intelligent check on the number of runs can be made by defining complex stop rules for a cycle instead of simply using a fixed number of trials. Such stop rules, for example, can monitor the means, standard deviations, and ranges of parameters of a certain response class, and stop the estimation if new samples no longer change these values, i.e. when the estimates converge. Table 2 refers to the example described above.

TABLE 2 Convergence of parameter estimates with increasing number of samples (independent cycles).

Number	a			b		
of samples	Mean	Minimum	Maximum	Mean	Minimum	Maximum
2	0.89	0.85	0.92	1.10	0.69	1.52
4	0.84	0.77	0.90	1.51	1.14	1.72
6	0.86	0.82	0.91	1.48	1.13	1.93
8	0.84	0.73	1.00	1.48	0.98	1.90
10	0.85	0.65	0.97	1.35	0.63	2.00
20	0.82	0.68	1.00	1.42	0.90	1.98
50	0.87	0.71	1.03	1.24	0.51	1.96
100	0.84	0.64	1.06	1.34	0.48	1.97
200	0.84	0.64	1.05	1.33	0.43	1.99
500	0.85	0.63	1.06	1.33	0.43	1.99
1000	0.85	0.63	1.08	1.33	0.34	2.00
2000	0.85	0.63	1.08	1.33	0.34	2.00
5000	0.85	0.63	1.08	1.33	0.34	2.00

b. *Speed up the trial runs*

Since a simulation program may run several thousand times in a Monte Carlo framework, streamlining the code will pay off. This includes, for example, the inactivation of all statements that are not essential for the determination of performance criteria. Examples might be auxiliary output variables, which are not used in the testing procedure. Also, parts of the model that are unchanged within a cycle of trial runs (for instance, setting up the geometry of the lake in the second application example, Section 3.2) should not be executed more than once in such a cycle. This, of course, requires more programming effort than simply calling the entire model as a subroutine of the Monte Carlo program – a compromise between programming effort and computer resource utilization has to be found.

A somewhat simpler possibility is to abandon a run as soon as it is obvious (even during run-time) that a given constraint condition will be violated. Since this may happen within the first few time steps, savings in computer time can be considerable.

c. *Reduce input/output*

As even a small simulation program, when run several hundred or thousand times, can produce an absolutely incomprehensible mountain of output, the reduction of output is essential for more than one reason. First, there will rarely be enough space to store it all; second, nobody is going to look at it all anyway; and third, I/O is time-consuming. Therefore, it is essential to reduce output to a minimum and do whatever processing has to be done with the output (e.g. classification, and calculation of certain statistics) within the Monte Carlo program. Again, there is a trade-off between the size a program can have on a certain machine, setting an upper limit to what can be done simultaneously, on-line, and storage capacity. Designing "intelligent" programs for the automatic analysis of Monte Carlo runs is probably the most demanding – and most challenging – part of the technique.

Similarly, input should clearly also be reduced to the absolute minimum. The most obvious examples are time-variable inputs or forcings to a dynamic simulation model, which should not be read at each time step of each trial, but only once for a cycle of trials, and then stored in an appropriate form within the program. Again, this calls for a compromise between time and core requirements.

d. *Think first*

As trivial as this last "rule" might seem, it is probably the most important one. It is most tempting to just let the program run (specifically when computer time is a free commodity) – and then to discover a little bug, somewhere, that makes thousands of runs worthless. Time spent in carefully considering the estimation scheme will certainly pay off in the long run. For example, if the parameter ranges sampled are fairly large, most complex models are bound to "crash" sooner or later – unless care is taken of zero divides, overflows, and underflows. Also, since operating systems tend to fail sometimes, provisions should be made that, in case of the unavoidable crash, only a minimum amount of information is lost, and an estimation cycle can be restarted. The Monte Carlo approach is very forgiving and helpful in this respect, as sample runs can always be pooled.

3 APPLICATION EXAMPLES

3.1 Hypothesis Testing: A Marine Pelagic Food-web Example*

The study of environmental systems as ecological and physicochemical as well as socioeconomic entities requires a high degree of simplifying formalism. However, a detailed understanding of a systems function and response to various changes for the explicit purpose of systems management and planning still requires complex hypotheses, or models, which can hardly be subjected to rigorous tests without the aid of computers. Systems simulation is a powerful tool for subjecting complex hypotheses to rigorous tests of their logical structure, as well as a possible means for rejecting or corroborating the underlying hypotheses.

The complexity and variability of environmental systems, the scarcity of appropriate observations and experiments, problems in the interpretation of empirical data, and the lack of a well established, comprehensive theoretical background make it difficult to test any possible conceptualization, or hypothesis, describing a given system. A formal approach to hypothesis testing, based on numerical simulation and Monte Carlo methods, which explicitly considers the above constraints, is proposed in this section.

Based on a data set from the North Sea, a series of hypotheses on the structural relations and the dynamic function of the pelagic food web is formulated in terms of numerical models. Hypotheses of various degrees of aggregation and abstraction are tested by comparing singular statements (predictions) deduced from the proposed hypotheses (the models) with the observations. The basic processes of primary production, consumption, and remineralization, driven by light, heat, and advection/diffusion, are described in systems models ranging in complexity from two compartments to many compartments and species groups. Yearly cycles of systems behavior are simulated with each of the proposed models. A comparative analysis of the response of each of the models allows conclusions to be drawn on the adequacy of the alternative hypotheses, including their "unknowns" or initial conditions (i.e. parameters). This analysis also allows one to reject inadequate constructs, and provides some guidance on how to improve a certain hypothesis, even in the presence of a high degree of uncertainty.

Universal statements, describing those properties of a system that are invariant in space and time, may be called models, whether they are of an informal (e.g. verbal or mental) or a formalized mathematical structure. Such models, viewed as scientific theories, have to be *testable*. When one feeds or substitutes a set of specific singular statements into the model (the initial conditions, which, in the case of a mathematical model, also include the model parameters in a general sense, as discussed in Section 2.2), it must be possible to deduce or predict testable singular statements (i.e. possible observations or the outcome of possible experiments). Disagreement between the prediction deduced from the hypothesis or model and the available

*This section is largely based on Fedra (1981a, b).

observations would then require rejection of the hypothesis, modification and improvement, or the search for alternative hypotheses, which would then have to be subjected to the same procedure. This method, which would basically represent the strategy of scientific research proposed by Popper (e.g. 1959), labeled falsificationism by critics such as Feyerabend (1975) and Lakatos (1978), however, has a major drawback when applied to complex simulation models or dynamic hypotheses describing ecological systems, in that the so-called initial conditions to be used with the basic structure of the theory to deduce the testable predictions are not exactly known. In one simple example given by Popper (1959), where he refers to a mechanical experiment (breaking a piece of thread), the initial conditions to be specified are simple enough: a weight and the characteristics of the thread (e.g. material, diameter etc.), which are measurable without considerable error (it is significant that many examples used in epistemological analyses refer to relatively simple physical systems). Measurements "without" error, however, are not usually possible when we are dealing with the complex aggregates conceptualized as "units" in large-scale systems thinking and models. This can certainly be seen as the result of two basic shortcomings, one in the measurement techniques available, another in the formulation of the models themselves: if the models require unknowns as inputs, they are not well formulated. The latter is certainly a generic shortcoming of environmental models and the underlying theoretical understanding.

The same line of argument can be followed with regard to the observation used for comparison with model output in hypothesis testing. The breaking of a thread, the singular prediction in Popper's example, is readily observable. It either happens, or does not. In most environmental applications, however, we have to compare predictions with measurements (as a rule, samples) of the system, which always include some measurement error, that is to say, these are ranges. Also, in environmental systems the degree of abstraction and aggregation is quite different for measurements and for model conceptualization. Therefore, the observations and measurements can serve only as samples of the properties or the state of the units conceptualized. As these units are generally heterogeneous (in terms of their measurable properties) and are generally characterized by a high degree of variability, further uncertainty has to be dealt with in the hypothesis-testing procedure.

Retaining the logical structure of testing a proposed hypothesis, but including at the same time the appropriate (or rather unavoidable) way of describing uncertain "initial conditions" as well as the expected outcome of the experiment, involves the following procedure. It is possible to describe the initial conditions or inputs by several numbers (forming a vector, determining a point in the n-dimensional input vector space) and to do the same for the expected result of the experiment (the observed behavior of the system), resulting again in a point in an n-dimensional output or behavior space. In the presence of uncertainty, the two points will have to be extended to regions in their respective spaces. Instead of the two vectors, we have to deal with sets of vectors with certain statistical properties and probability structures.

To test any specific hypothesis, we now examine whether, for a set of admissible initial conditions (i.e. the parameters), predictions (members of the set of allowable outcomes) can be made. The rejection of a hypothesis, whenever no allowable outcome can be generated, is based on a statistical argument, as the number of possible initial conditions forming the admissible set is infinite, and only samples can be examined. Also, the set of admissible initial conditions will rarely be well defined on the basis of *a priori* knowledge (*a priori* in relation to the specific experiment to be carried out). Generally, it will be possible to specify allowable ranges for the individual initial conditions. The admissible set, however, is also characterized by the correlation structure, which determines the "shape" of the admissible region in the parameter vector space.

This method of testing a given hypothesis does not indicate how such a hypothesis can be arrived at in the first place — by "conjecture." Popper's rejection of inductive reasoning does not provide much help, but in practice hypotheses (and simulation models) are rarely generated randomly but are always based on empirical knowledge. However, the process of testing and rejecting a given hypothesis can also provide some diagnostic information about the causes of failure and about possible ways to improve the hypothesis.

One possibility is strict parsimony: to start with the simplest possible conceptualization, or the least complex model one can formulate bona fide, which still may capture the relevant features of the system in view of the problem studied. Certainly, each hypothesis tested should be an honest candidate for success: "What then is the point of setting up a [Poisson] model like a skittle, just to knock it down again?" (Finch 1981). If this simple version fails to give an acceptable behavior over the allowable parameter ranges, the model structure is modified. Complexity is increased by adding elements and more complex process descriptions to the model (Figure 6), until a satisfactory behavior can be achieved. However, there is in any case more than one way to increase the complexity of a model. A general formalization of this "adding of complexity" seems to be most difficult, if not impossible. Some guidance for this process can be expected from the analysis of a series of errors, as will be shown below. Also, as I am only considering "conceptual" models (as opposed to purely statistical models, they are based on physical processes and only include terms directly interpretable in the "real world"), additional observations can be exploited in many cases. Knowledge accumulated from the study of similar systems may also be helpful in changing a given model structure.

Building up complexity and iteratively subjecting each version or level of the model to extensive tests should allow one to learn about the way structural changes influence model response. At the same time, the intricate connection between structure and the parameters has to be emphasized, since model behavior is certainly responsive to both. As changes in the model structure will, in almost every case, also necessitate changes in the parameters (their numbers, admissible ranges, and interpretation), comparisons of different versions are quite difficult. Although the approach described below is clearly far from being ideal, any attempt at a formalization of the modeling

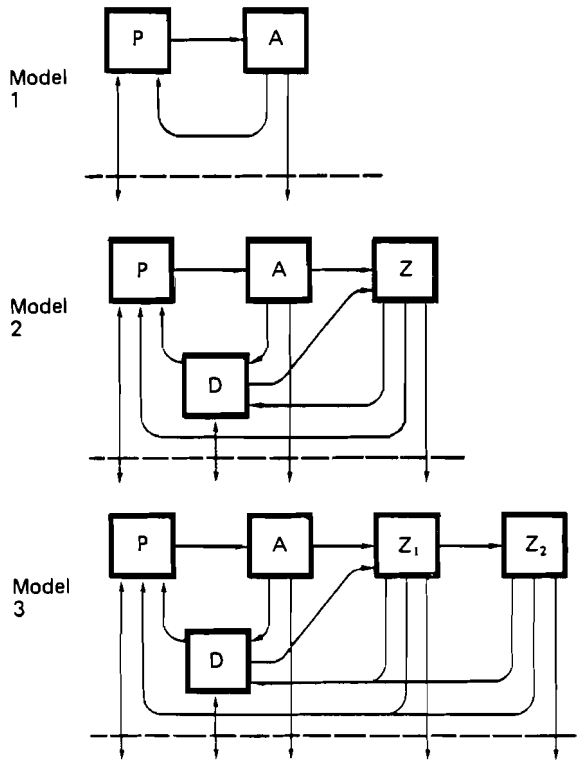

FIGURE 6 Flow diagrams for the models compared: P, phosphate; A, phytoplankton; D, detritus; Z, zooplankton; Z_1, herbivores; Z_2, carnivores.

process seems preferable to a purely arbitrary and subjective procedure.

3.1.1 The Empirical Background: Describing the Environmental System

Considering the above constraints, the direct use of the raw data available on any ecosystem seems to be rather inappropriate and difficult for the testing of complex and highly aggregated dynamic hypotheses. Consequently, we have to derive from the available data a description of the system and the processes we want to study at a more appropriate level of abstraction and aggregation. This description, which already has to be formulated in terms of the hypothesis to be tested, should take advantage of all the available information, and at the same time provide an estimate of the reliability of this information at the required level of abstraction.

To illustrate the approach, a data set from the southern North Sea was used. Most of the information utilized stems from the yearly reports of the Biological Station Helgoland, and describes physicochemical as well as biological variables at the sampling station "Helgoland-Reede" for the period 1964–79.

Figure 7 summarizes the data used. The driving environmental variables, water temperature and radiation, were found sufficiently smooth and well

behaved for a direct utilization of the long-term averages, approximated by simple sine waves. Data for nutrients (PO_4-P) and algae (measured as chlorophyll *a* as well as in terms of carbon, recalculated from counts) showed consistent yearly patterns. However, when the year-to-year variations (as well as the implicit sampling errors) are included, the high variability of the observations as well as the difficulty in averaging over time (several years) becomes obvious. Although the average phytoplankton dynamics show a single, but extended peak around July/August, the individual years exhibit at least two peaks in the summer. As a result of their variable timing, the peaks are averaged out when one looks at the long-term mean. Also, the long-term mean is about one order of magnitude below the spiky peaks of the data for the individual year. Little information was available on zooplankton biomass. However, some additional information from independent experimentation, mainly on primary production, was also found. Also, the (time-variable) ratio of phytoplankton carbon to chlorophyll was used for the models described below, and approximated by a simple exponential curve.

Among the invariable, generalizable conditions derived from the observations are the following:

1. Primary producers are below a level of 4.0 mg m^{-3} chlorophyll during the first three months of the year.
2. Between Julian days 120 and 270 there is at least a twofold increase in biomass.
3. There have to be at least two peaks within that period, with a reduction of more than 25% of the first peak value in between the two peaks.
4. After day 270, biomass must be below 4.0 mg m^{-3} chlorophyll again.
5. The higher of the two peak values must not exceed 25 mg m^{-3} chlorophyll.
6. Yearly primary production must be above 300 and below 700 g C m^{-2}.
7. Herbivorous consumers (zooplankton) reach their first biomass peak value (defined as at least a twofold increase of their initial biomass before a subsequent decline) after the phytoplankton.
8. The maximum density of herbivorous consumers must not exceed 1000 mg C m^{-3}.
9. PO_4-P concentration has to be above 20 mg m^{-3} between days 1 and 90.
10. The average PO_4-P concentration between days 120 and 240 has to be below 20 mg m^{-3}.
11. PO_4-P concentration has to be above 20 mg m^{-3} after day 270.
12. PO_4-P concentration must never exceed 50 mg m^{-3} and it must never be below 2 mg m^{-3}.

13–17. All state variables must be cyclically stable (± 25% tolerance level).

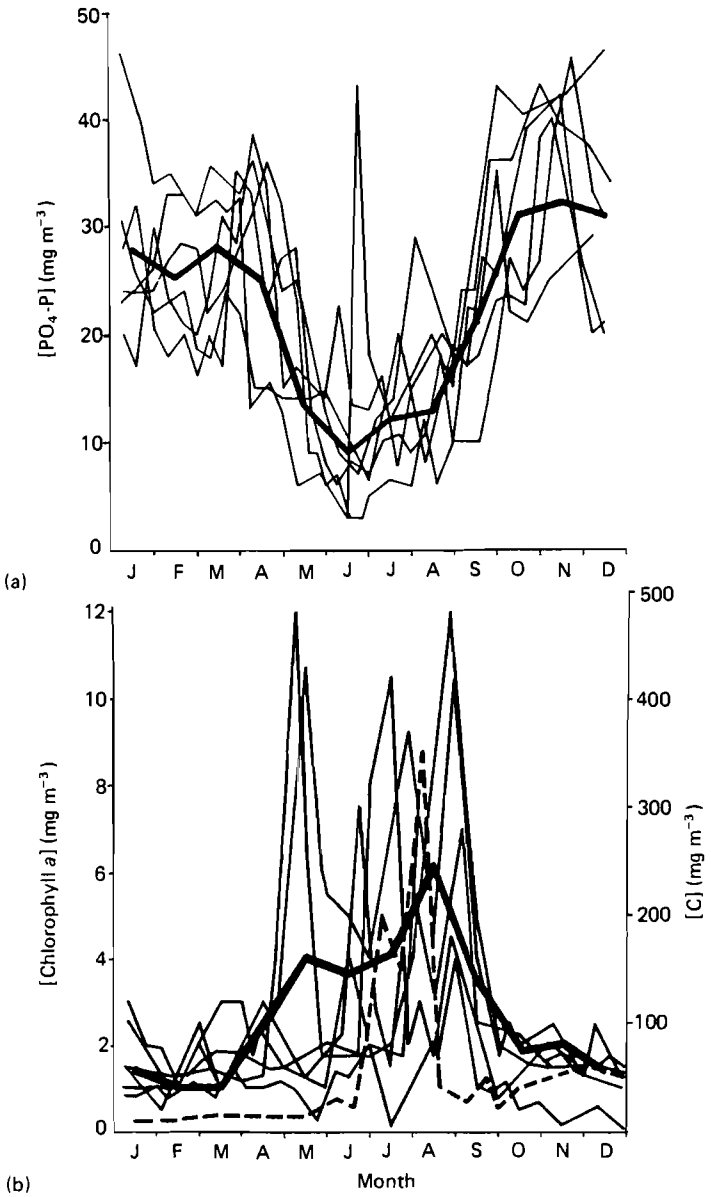

FIGURE 7 (a) Phosphorus dynamics (PO_4-P) for selected years from 1964 to 1979; the thick line indicates monthly averages for the years 1965 to 1975 (after unpublished data from Weigel and Mangelsdorf; Harms; Harms and Hagmeier; Harms, Mangelsdorf, and Hagmeier; Mangelsdorf). (b) Chlorophyll dynamics for selected years from 1964 to 1979; the thick line indicates monthly averages for the years 1965 to 1975; the broken line shows microzooplankton carbon for the year 1975 (after unpublished data from Weigel, Hagmeier, and Treutner; Hagmeier, Kanje, and Treutner (yearly reports of the Biological Station Helgoland, 1964–79)).

This description of the observed systems features, defining a region in the behavior space of the system, has to be understood as a semiquantitative description of persistent patterns rather than a quantitative description of the system for any specific period. Of course, more resourceful analysis of the available data and the incorporation of additional information would allow this description to be refined.

3.1.2 Hypothesis Generation and Testing: Designing Alternative Models

In the literature, one can find many conceptualizations or models of aquatic ecosystems, and of the pelagic, productive upper part of lakes or the oceans in particular. Several books have dealt with such conceptualizations for marine systems (e.g. Steele 1974, Nihoul 1975, Cushing and Walsh 1976, Goldberg *et al.* 1977, Parsons *et al.* 1977, Kremer and Nixon 1978, Barnes and Mann 1980). Some contributions deal with the North Sea specifically (e.g. Pichot and Runfola 1974, 1975, Radach and Maier-Reimer 1975, Radach 1980). A wide range in detail and complexity has been covered both with respect to biological and physiological factors (e.g. Steele and Frost 1977, Steele and Mullin 1977, Morris 1980, Greve 1981) and with the emphasis on the physical and spatial aspects (e.g. Walsh 1975, Steele 1976, 1978, Dubois 1976, Dubois and Closset 1976). Against this background, the models presented and discussed below are not to be understood as further contributions to the study of the southern North Sea; rather, they are extremely simplified examples, primarily designed to illustrate the approach.

3.1.3 Hypothesis Number 1: Two Compartments in a Simple Physical Framework

Let me now try to formulate one very simple hypothesis about the pelagic food web described by the data set in Section 3.1.1. Again, it should be stressed that the model described below is proposed not as a useful representation of the southern North Sea, but only as an illustrative example to demonstrate the approach. The system is conceptualized as consisting of only two compartments, namely particulate, photosynthesizing organic matter, and mineral nutrients, which are coupled by the processes of primary production and nutrient uptake, mortality, and respiration/mineralization; the system is driven by light and heat and by turbulent mixing (eddy diffusivity). Controlling mechanisms are light and nutrient limitation of primary production, self-shading of algae, and temperature dependency of all the biological processes. A detailed description of these models is given by Fedra (1981b, c).

Monod kinetics are used to describe nutrient limitation of primary production, with a constant half-saturation concentration; maximum growth rate is described as an exponential function of temperature, with a Q_{10} of about 2; light limitation is described using the double time–depth integral of Di Toro *et al.* (1971) of Steele's (1962) equation (the implications of this formulation are discussed by Kremer and Nixon (1978)). Mortality is described as a nonlinear, concentration-dependent function of algal biomass and is directly coupled to remineralization, without any time lag or further control.

Mixing with a "deep layer" is described as the exchange of a constant fraction of the volume of the upper layer (10 m deep), where the PO_4-P concentration of the deep layer equals the initial (winter) concentration of the upper layer, and the algae concentration is zero, that is to say, algae can only be lost from the system. The rate of mixing is changed by a step function, triggered by temperature, such that the initial high (January) value is set to one-tenth as soon as the surface temperature reaches three times its starting value; mixing rate is reset to the high value as soon as the surface temperature drops below the trigger level. This extremely simplified variation of the mixing coefficient over the year comes close to the patterns used by Lassen and Nielsen (1972), and is also frequently used for the description of seasonal thermal stratification in lakes.

This model requires six parameters to be estimated, the initial conditions and the driving variables being "known." For each of these parameters or rate coefficients a possible, allowable range can be specified, depending on the available knowledge. In the worst case, a mortality rate, for example, has to be greater than zero and smaller than one. To circumvent the problem of uncertain initial conditions, a set of likely values (estimated from the available data) was taken and allowed to adjust by letting the model simulate a period of three years. This strategy (using the results of the third year after arbitrarily specifying the initial condition for year one instead of adding more dimensions to the parameter search space) was followed with all the models described below. The models are formulated in terms of phosphorus, with constant stoichiometric conversions to carbon and a time-variable carbon: chlorophyll ratio.

3.1.3.1 Testing hypothesis number 1

To test the hypothesis formulated in Model 1, the model was incorporated into a Monte Carlo framework, which randomly sampled a parameter vector from the allowable ranges (Table 3), ran the model through a period of three years, to allow the arbitrary initial values of the state variables to adjust, and finally tested for violations of the constraint conditions in the third year of simulation. This process was repeated for a sufficiently high number of trials (more than 100,000 runs were performed with each of the models). Since 100,000 runs of even a comparatively simple simulation model produce a large amount of almost incomprehensible information, several auxiliary programs for the automatic analysis of the simulation results were used. Table 3 shows an example of the output of one of these analysis programs, which includes the parameter ranges sampled and the basic statistics of the parameter ensemble used to generate the model response shown in Figure 8.

In summary, Model 1 could fulfill all of the constraint conditions but one: it was not possible to reproduce two algae peaks during the summer period (without violating several other conditions). Figure 8 shows a sample output from Model 1.

Hypothesis number 1 consequently had to be rejected. To build an improved hypothesis, the distributions and correlation structure of parameters and output variables from those runs violating only condition 3 (the two

TABLE 3 Automatic parameter estimation analysis program: parameter statistics for Model 1. Monte Carlo run selection: violation of condition 3 only.

	Mean	Min.	Max.	Standard deviation	Range sampled	
Parameters						
1 Michaelis constant	6.08	2.23	13.84	3.33	2.00	15.00
2 Phytoplankton mortality	0.36	0.25	0.50	0.07	0.05	0.50
3 Light optimum	410.10	301.51	497.40	63.15	300.00	500.00
4 Mixing coefficient	0.18	0.11	0.23	0.03	0.05	0.50
5 Maximum growth coefficient	1.03	0.70	1.42	0.18	0.50	2.50
6 Temperature trigger	3.24	2.56	3.85	0.29	2.00	4.00
Output constraint variables (concentrations in mg m^{-3})						
7 Chl. high, days 1–90	0.	0.	0.	0.	(not violated)	
8 Chl. summer peak	8.16	7.84	8.72	0.24		
9 Chl. first peak	8.16	7.84	8.72	0.24		
10 Chl. low between peaks	3.89	3.49	4.09	0.13		
11 Day of first peak	189.81	188.00	191.00	0.83		
12 Chl. second peak	0.	0.	0.	0.		
13 Day of second peak	270.	270.	270.	0.		
14 Chl. high after day 270	0.	0.	0.	0.	(not violated)	
15 Chl. maximum	8.16	7.84	8.72	0.24		
16 PO_4 maximum	30.00	30.00	30.00	0.		
17 PO_4 minimum	15.47	12.86	16.87	1.07		
18 PO_4 low before day 90	29.92	29.67	29.99	0.092		
19 PO_4 low after day 270	27.84	27.56	28.36	0.18		
20 PO_4 average, days 120–240	19.02	17.27	19.93	0.73		
21 Primary production (g C m^{-2})	406.20	303.05	625.44	86.58		

Correlation matrix of parameters

	1	2	3	4
2	−0.2			
3	0.1	−0.6		
4	0.1	0.7	−0.5	
5	0.5	0.6	−0.0	0.7
6	−0.2	0.5	−0.2	0.2

algae peaks) were analyzed. Some technical details of this kind of analysis are described below, in the next application example, and in Fedra *et al.* (1981). However, the analysis indicates that phytoplankton mortality is a critical process, and its representation consequently deserves refinement. This can be deduced from the significant correlations between the mortality rate coefficient and the other parameters, as well as from different output variables in groups of simulations violating different constraint conditions.

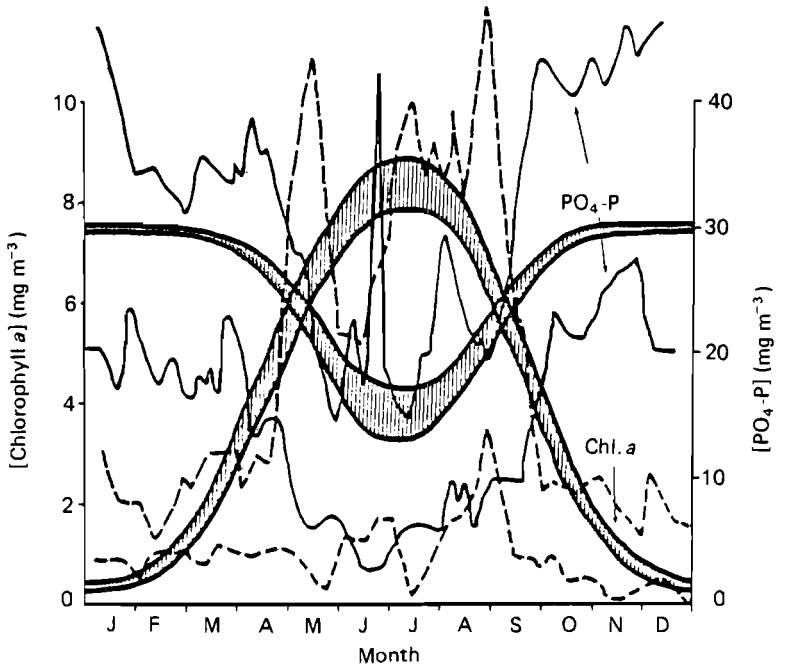

FIGURE 8 Output from Model 1 (Figure 6): envelope over 31 runs (for the state variables "algae" and "nutrients") contrasted with the envelopes over the field data set used (Figure 7). The runs shown fulfill all behavior requirements but the condition of two peaks of algae biomass during the productive season.

3.1.4 Hypothesis Number 2: A Four-compartment Web

As a slightly more complex alternative to Model 1, a second version was formulated to incorporate detritus and omnivorous zooplankton. The description of primary production and the physical framework are essentially the same as in the first version. Model 2, however, splits the phytoplankton mortality into a natural, background mortality, which is described as concentration-dependent, and losses due to grazing. Background mortality as well as zooplankton mortality now feeds into the detritus pool, which in turn feeds (temperature-dependent) back into the nutrient pool; detritus is also available for zooplankton, for which, however, a certain preference for living algae is assumed. Zooplankton respiration also feeds into the nutrient pool. Figure 6 shows the flowchart for this model. The description of grazing was based on a simple encounter theory. With this inclusion of a herbivorous zooplankton compartment, a choice had to be made of how to describe grazing. Numerous formulations abound in the literature, and to give one example, Jørgensen (1980a, table 3.9) lists 14 different formulations of zooplankton grazing rates. Given that there is no additional information available to support a decision on which construct should be used, one can start with as

simple an assumption as possible, and subsequently test it. In the test, the resulting model performance was not satisfactory either: for low values of the grazing rate constant, the zooplankton did not survive phytoplankton lows in winter, and died away. For high values of the feeding rate, in contrast, phytoplankton was removed very quickly, as soon as it started growing in the spring, with a consequent collapse of the zooplankton population itself. This does not rule out the possibility that features of the model other than the formulation of grazing are responsible for these failures, or at least contribute to them.

However, after "rejection" of the encounter theory, description of grazing was based on a saturation curve, similar to Michaelis–Menten kinetics, using a temperature-dependent maximum feeding rate coefficient, with the same temperature dependency as for respiration and remineralization. The governing equations are given in Fedra (1981b, c).

This version was also subjected to the Monte Carlo simulation procedure described above. The resulting response was analyzed accordingly. The introduction of a second trophic level in Model 2 now allowed a reproduction of the well known oscillatory behavior of predator–prey systems, and thus permitted fulfillment of condition 3, requiring two phytoplankton peaks. However, this version was incapable of producing enough algal carbon over the year, thus violating condition 6 (Table 4). This is simply due to the fact that only at comparatively low primary productivity levels was the system stable enough to stay within the behavioral bounds specified. The output or constraint variable yearly primary production showed a strong positive correlation with the zooplankton grazing coefficient (parameter 6 in Table 4) and zooplankton respiration (parameter 8), which is a major source of nutrient recycling. This directly points to the positive feedback loop in these processes, and the resulting stability problems in this version of the model.

3.1.5 Hypothesis Number 3: One More Trophic Level

Model 2 was used as the basis for yet another modification, namely the introduction of another trophic level of carnivorous zooplankton, to explore its importance in controlling the herbivores (Greve 1981). A sample output of Model 3 is shown in Figure 9, and the equations are given by Fedra (1981b). Another five parameters had to be introduced for the additional detail in Model 3, leading to further problems in the estimation and analysis. For example, the proportion of runs aborted during run-time (due to the violation of some run-time checks on the state variables, confining them within certain plausible ranges, or to numerical instabilities in solving the system of differential equations) grew dramatically to almost 99.9% of the trial runs when the broad initial parameter intervals given in Table 5 were sampled.

The second trophic level of carnivorous zooplankton feeds on the herbivores in structurally the same way as the herbivores feed on the phytoplankton; herbivores, however, have the additional source of detritus available. Owing to its higher complexity, Model 3 was able to generate a broad spectrum of behavioral features (Table 5); it could not, however, fulfill all of the test conditions imposed on its behavior at the same time. Obviously, the simple inclusion of a structurally similar additional compartment did not

TABLE 4 Automatic parameter estimation analysis program: parameter statistics for Model 2. Monte Carlo output run selection: 43 runs violating condition 6 only (primary production).

	Mean	Min.	Max.	Standard deviation	Range sampled Low	High
Parameter values						
1 Michaelis constant	9.88	5.06	14.88	3.17	5.00	15.00
2 Phytoplankton mortality	0.07	0.03	0.10	0.02	0.00	0.10
3 Light optimum	429.54	317.58	499.78	52.08	300.00	500.00
4 Mixing coefficient	0.05	0.01	0.10	0.02	0.01	0.10
5 Max. growth coefficient	1.53	0.82	2.35	0.41	0.50	2.50
6 Zooplankton grazing	1.01	0.19	1.87	0.48	0.01	2.00
7 Zooplankton detritus uptake	0.10	0.005	0.20	0.06	0.01	0.50
8 Zooplankton respiration	0.09	0.02	0.28	0.06	0.01	0.25
9 Zooplankton mortality	0.11	0.012	0.28	0.07	0.01	0.50
10 Remineralization	0.25	0.015	0.49	0.13	0.01	0.50
11 Temperature trigger	3.04	2.51	3.49	0.29	2.50	3.50
12 Grazing half-saturation constant	13.80	3.94	23.44	5.07	0.00	25.00
Output constraint variables (concentrations in mg m^{-3})						
13 Chl. high, days 1–90	0.56	0.20	1.07	0.23		
14 Chl. summer peak	7.60	5.86	9.05	0.69		
15 Chl. first peak	7.60	5.86	9.05	0.69		
16 Chl. low between peaks	1.35	0.06	3.91	1.00		
17 Day of first peak	155.37	138.00	192.00	13.07		
18 Chl. second peak	3.13	1.00	4.45	0.72		
19 Day of second peak	270.	270.	270.	0.		
20 Chl. maximum	7.60	5.79	9.05	0.69		
21 PO_4 maximum	25.76	24.76	26.72	0.433		
22 PO_4 minimum	4.83	2.16	11.97	2.19		
23 PO_4 low before day 90	25.36	24.73	25.91	0.23		
24 PO_4 low after day 270	22.19	20.06	25.11	1.12		
25 PO_4 average, days 120–240	17.11	13.70	19.86	1.67		
26 Primary production (g C m^{-2})	40.60	19.08	75.76	14.08		
27 Day of zooplankton peak	165.63	145.00	220.00	16.20		
28 Zooplankton at algae peak	0.69	0.	3.48	0.92		
29 Zooplankton peak value	14.33	7.66	18.74	3.00		

resolve the basic problem; since the process rates of both zooplankton compartments are only determined by external driving variables (temperature, food availability) but not by internal control mechanisms (e.g. developmental stages, size and age classes, etc.), the resulting zooplankton response was not adequate over the whole range of driving conditions for a yearly cycle. The model does well for part of the year, or for part of the required behavioral features over a full year; if, however, the model behaves well during the productive season, zooplankton will starve and collapse during the winter.

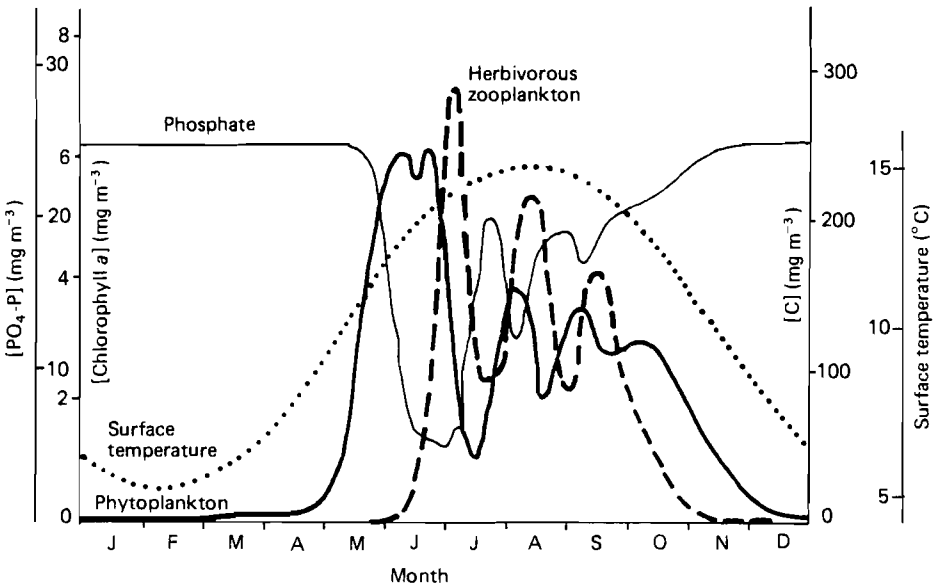

FIGURE 9 Sample output from Model 3; the run violates only the condition that yearly primary production should be above 300 g C m^{-2}. Thick line: phytoplankton (chlorophyll a); broken line: herbivorous zooplankton (carbon); thin line: phosphate; dotted curve: surface temperature.

Alternatively, if all plankton groups survive the winter well, the onset of high primary productivity will quickly lead to explosive growth and consequent collapse.

Another possible explanation, although less appealing, might be that some of the constraint conditions are just too narrow or badly placed. For example, the lower bound for yearly primary production set at 300 g C m^{-2} could seem unrealistically high. Another comparable estimate, given by Pichot and Runfola (1975) for the Southern Bight off the Belgian coast, is 17.5 g N m^{-2}, which amounts to less than half the German estimate (from yearly reports of the Biological Station Helgoland, 1964–79) when converted to carbon units. Reducing the constraint of minimum yearly primary production to, say, 100 g C m^{-2}, would make the model "acceptable." Quite obviously, the definition of the constraint conditions can be critical, and thus indicates where further effort in data analysis (or collection) would be worth while.

This points to one of the principal problems in environmental systems modeling, namely the problem of interpretation of "micro-scale" observations and experiments, eventually performed in the laboratory, on a macro-scale compatible with the level of aggregation and abstraction used in the conceptualization of the system.

As was argued above, parts of the systems behavior can be reproduced relatively easily. The major problem was found to lie in the reproduction of

TABLE 5 Parameter statistics for Model 3. Monte Carlo output: 250,000 trial runs made, 219 runs evaluated.

	Mean	Min.	Max.	Standard deviation	Range sampled Low	High
Parameter values						
1 Michaelis constant	11.13	2.36	19.82	4.82	2.0	20.0
2 Phytoplankton mortality	0.19	0.	0.50	0.46	0.0	0.5
3 Light optimum	419.04	300.00	548.95	74.60	300.0	550.0
4 Mixing coefficient	0.05	0.00	0.20	0.05	0.0	0.2
5 Max. growth coefficient	7.10	0.76	9.98	2.11	0.5	10.0
6 Zooplankton grazing	1.01	0.00	1.99	0.57	0.0	2.0
7 Zooplankton detritus uptake	0.51	0.04	0.80	0.19	0.0	0.8
8 Zooplankton respiration	0.18	0.	0.72	0.17	0.0	0.8
9 Zooplankton mortality	0.21	0.	0.78	0.18	0.0	0.8
10 Remineralization	0.21	0.	0.50	0.14	0.0	0.5
11 Temperature trigger	2.94	2.50	3.49	0.99	2.5	3.5
12 Grazing rate, carnivores	1.15	0.04	1.99	0.53	0.02	2.0
13 Mortality rate, carnivores	0.15	0.00	0.79	0.11	0.0	0.8
14 Respiration, carnivores	0.07	0.	0.62	0.09	0.0	0.8
15 MM constant, algae	16.42	0.43	29.99	7.90	0.0	30.0
16 MM constant, detritus	10.77	0.00	29.80	8.80	0.0	30.0
17 MM constant ,herbivores	15.07	0.11	29.65	8.82	0.0	30.0
Output constraint variables (concentrations in mg m^{-3}, production values in g C m^{-2})						
Chl. high, days 1–90	4.049	0.058	9.492	2.293		
Chl. summer peak	6.149	2.385	10.111	1.589		
Chl. first peak	5.867	1.686	10.111	1.687		
Day of first peak	131.416	120.000	182.000	17.468		
Chl. low between peaks	4.227	0.539	8.306	1.840		
Chl. second peak	4.675	0.	8.551	1.739		
Chl. high after day 270	4.678	1.039	9.112	1.689		
Chl. maximum	6.423	2.385	10.111	1.488		
PO_4-P maximum	23.616	6.541	27.994	3.250		
PO_4-P minimum	1.975	0.062	18.797	2.098		
PO_4-P low until day 90	15.387	1.351	25.170	8.266		
PO_4-P low after day 270	6.457	0.531	23.395	5.126		
Zooplankton peak value	132.307	0.001	1259.682	229.683		
Carnivores peak value	402.631	8.680	1422.003	272.741		
PO_4-P average, days 120–240	3.791	0.294	21.827	3.490		
Primary production	142.038	6.016	374.074	74.652		
pp January	0.207	0.	3.976	0.528		
pp February	0.932	0.	12.258	1.892		
pp March	5.218	0.	23.553	5.450		
pp April	15.463	0.	38.713	8.502		
pp May	19.989	0.239	46.851	10.749		
pp June	26.289	1.038	76.145	15.214		
pp July	23.977	0.842	71.133	14.238		
pp August	20.277	0.698	63.635	12.717		
pp September	18.847	0.557	59.717	12.386		
pp October	8.838	0.114	32.067	6.462		
pp November	1.735	0.003	7.953	1.697		
pp December	0.266	0.	3.600	0.480		

TABLE 5 *(continued)*

	Mean	Min.	Max.	Standard deviation	Range sampled Low	High
Secondary production	82.737	1.284	571.125	91.024		
Tertiary production	3.745	0.	23.717	4.149		
Algae carbon, end	11.608	0.001	241.564	32.676		
Algae carbon, start	11.610	0.001	241.525	32.676		
Phosphate, end	22.584	4.600	26.336	3.870		
Phosphate, start	22.584	4.600	26.336	3.870		
Zooplankton, end	8.470	0.002	221.002	23.783		
Zooplankton, start	8.459	0.002	219.280	23.701		
Zooplankton 2, end	1.139	0.	41.086	3.933		
Zooplankton 2, start	1.140	0.	41.051	3.931		
Detritus, end	32.840	1.217	473.174	60.060		
Detritus, start	32.850	1.253	472.995	60.078		
Total P, end	23.94	5.14	33.26	3.30		
Total P, start	23.09	4.96	28.68	3.58		

the full range of systems behavior over the yearly cycle, that is, over a wide range of the physical driving conditions.

Quite obviously, none of the models discussed above is entirely satisfactory in light of the constraint conditions defined. The constraint conditions, although seemingly liberal, are quite demanding when compared with many examples of arbitrary judgment, so-called "satisfactory" or "reasonably good" agreement between (some) output variables of a model and the observations one can frequently find in the literature. However, this section does not attempt to propose an elaborated dynamic model of the pelagic food web of the southern North Sea, but rather attempts to demonstrate (using the example of admittedly quite simplistic models) a formal Monte-Carlo-based approach to model or hypothesis testing.

3.1.6 The Generalizable Lesson

To build complex hypotheses, used to describe and explain the structural and behavioral features of ecological systems, a formal approach and rigorous testing procedures are required. As has been demonstrated, parts of the observed behavior of a system may easily be reproduced. This then goes parallel with unrealistic behavior in other parts of the system. A complex hypothesis or model, however, can only be accepted as a valuable working tool, with explanatory value and predictive capabilities, if it fulfills ***all the constraints*** one can formulate as defining the observed systems behavior. Violation of one single condition necessitates the rejection of such a model, which should be just one step in an iterative process of analysis (Figure 1).

One basic idea of the approach is to use the available information according to its relevance to the model's (that is, the theory's) level of abstraction or aggregation. For a given model structure, this information is grouped into a set of singular statements that are to be substituted for the variables of the universal statement (the theory or model), and – as the analysis is usually done ***ex post*** – a set of singular statements (the observations already available from the system) describing the expected outcome of

the simulation experiment.

These constraint conditions, which generally will describe allowable ranges, have to be understood as replacing the arbitrarily precise observations that are possible, for example, in classical mechanics. The formulation of these constraints provides a high degree of flexibility. In addition to the direct utilization of individual measurements (including the measurement or sampling error to define a range), derived measures, relations, integrals, averages, etc. can be used. Whatever can be inferred from the observations is a valid constraint on the allowable model response. In addition, certain bounds, although not observed in a specific case, are obvious, deducible from some basic laws, such as mass and energy conservation, or from more empirical rules like maximum efficiencies or process rates.

Obviously, the description of the (lumped) states of a system can be accomplished much more easily on the appropriate level than the description of (lumped) process rates and controls (just think in terms of phytoplankton biomass versus production rate). Consequently, we turn the argument of the hypothesis-testing process around: instead of putting the "known" initial conditions (the rates, among others) into the model structure and deriving the response for comparison, we use the allowable response as a constraint to identify possible initial conditions. This is to say, we map a given region in the behavior space of a system back into the parameter vector space (Section 2.3). The test is then whether or not this region in the parameter space exists within the specified possible or plausible bounds.

Hypothesis generation, that is, the conjecture of the initial hypothesis or of an alternative hypothesis after the failure of a previous one, is a crucial step: the hypotheses we are using in environmental systems research are fairly complex or, rather, composite, that is, they are built from numerous individual constructs, each of them being a hypothesis in itself. Their complex, dynamic, and nonlinear interactions make it difficult to relate a failure in the overall performance of the model to any of the individual constructs used. The kind of sensitivity analysis provided by the method described above, although involving all input values (or parameters) simultaneously, only relates model performance to the inputs, and not to the structural features of the hypotheses *per se*. In principle, structure and input values are inseparable in their effect on the model response. Also, it is impossible to test any isolated process descriptions against observations – as has been proposed by some authors – as soon as feedbacks exist between the isolated process and the remainder of the system. In complex environmental examples, this will almost always be the case. Obviously, the same holds true for calibration, which, in the above approach, is part of the hypothesis-testing process.

To arrive at an alternative model structure, the diagnostic information provided by the analysis of failures is only of limited value. However, I assume that whoever builds simulation models has a fair understanding of the systems he is modeling. In the previous examples, it is ecological common sense that predation is usually a critical component in algae mortality and thus the control of algae biomass dynamics; and that herbivorous zooplankton may very efficiently be controlled by carnivorous species. Thus, the basic direction in which to go when adding complexity is obvious. Within a

given level of complexity, that is, when comparing alternative process descriptions for the same number of elements (such as in the comparison of different descriptions of grazing in hypothesis number 2), formulating an alternative description is again based on a simple and intuitive understanding of the observed as compared with the required functioning of the model: when the simple encounter theory used for the description of grazing failed under the typical overshoot symptoms of lack of feedback, an additional level of feedback was included with a nonlinear density dependence. If the parameter describing this effect (i.e. the grazing half-saturation constant, Table 4) is allowed to vary from zero to a relatively high value, the resulting test does in fact test the effect of the construct: since all the (conditionally) successful runs have significantly nonzero half-saturation values, the nonlinear process description is obviously superior to its linear counterpart. Thus, by "parameterization" of alternative model structures, they can be tested and compared in a single test series. This approach has been used by Steele (1974) in the context described above. In general, by parameterizing a larger number of structural alternatives, one could use the method to identify feasible model structures in the same way as it is here used to find feasible parameter combinations. Working with sets of alternative model structures (where usually only one process description is changed at a time), rather than extensively parameterizing the model structure, simple reduces the dimensionality of the problem and facilitates interpretation.

If a given hypothesis does stand up to all the tests one can design on the basis of the available data, that is to say, the hypothesis cannot (yet) be rejected, one can legitimately use it as a working hypothesis. However, quite easily we can imagine a situation where the uncertainty inherent in the behavior definition for a system is large enough to allow for more than one alternative hypothesis, without the possibility of discriminating or ruling out any of them (see the application example in Section 3.3). Although the two or more hypotheses do not differ significantly in their behavior in the descriptive, empirical test case (that is why no discrimination is possible, since the concept of significance here is related to the extent of the allowable behavior range, which in turn depends on data uncertainty and systems variability), they might well differ significantly when used for further predictions, that is, extrapolations outside the empirical range used for tests so far. Here the only possible approach would be to look for predictions from the alternative versions that clearly (and supposedly measurably) differ, and then perform the required observation or experiment in the field. The simulation of alternative hypotheses could thus provide some guidelines for measurements and fieldwork as well, allowing for a more precise formulation of questions to be addressed during expensive field observations.

3.2 Estimation and Prediction with Parameter Ensembles: A Lake Modeling Example*

In cooperation with the Austrian Lake Eutrophication Program, Project Salzkammergutseen, the Attersee, a deep, stratified, oligotrophic lake of almost 4000 million m^3 and a theoretical retention time of seven to eight years, was subjected to another version of the Monte Carlo approach. Basic data are compiled in Table 6. Investigations, carried out since 1974, and initially within the frame of the OECD Lake Eutrophication Program, indicated increasing eutrophication of the lake. Increasing phytoplankton peak biomass and decreasing transparency of the water signaled a trend toward eutrophication; however, the variability in the measurements and the comparatively short time span of observations make it difficult to identify significant changes. Nevertheless, a preliminary study of the nutrient-loading/production relations seemed promising. Primary production per unit lake area, algae peak biomass, and maximum epilimnetic concentration of available phosphorus were taken as approximate measures of the trophic state of the lake. Although the problem setting is somewhat diffuse from the point of view of possible management and water quality control measures, two principal features of the lake system allowed us to address practical problems. First, the major nutrient input stems from one point source, the upstream Mondsee (surface area 4.2 km^2, volume 510 million m^3, catchment area 247 km^2, and a retention time of about two years). More than 50% of the total phosphorus loading of the Attersee is attributable to the Mondsee discharge, and a very high fraction of particulate phosphorus is contained in this discharge (Müller 1979). The impact of possible changes in the trophic state of the Mondsee on the Attersee is therefore of considerable interest. Second, sewer systems and associated treatment plants for the sewage discharge into the Attersee and the Mondsee have been recently constructed (Flögl 1974, 1976a,b). Again, the impact of these installations on the water quality of the Attersee is of interest.

The data available for our analysis, comprising estimates of nutrient inputs and outputs as well as lake nutrient concentrations, collected roughly on a monthly basis, were found to show a high degree of variability both within and between years, and this variability was especially pronounced for the phosphorus measurements (Figure 10). This is, at least in the case of the orthophosphate, due to its low concentration, around 1 mg m^{-3}, which is approximately at the same level as the absolute measurement error. This also led us to the simplifying assumption of a horizontally completely mixed water body; the data would not support a more detailed spatial resolution for the model. It was also decided to combine the observations of several years to obtain a picture of a typical Attersee behavior pattern.

*This section is based on Fedra (1980) and Fedra *et al.* (1981).

TABLE 6 Attersee: basic lake data (after Flögl 1974).

Geographic position	47° 52' N
	13° 32' E
Catchment area	463.5 km^2
Surface area	45.9 km^2
Maximum depth	171 m
Mean depth	84 m
Volume	3934 million m^3
Length	20 km
Average width	3 km
Total shore length	53 km
Retention time	7–8 years
Average outflow	17.5 $m^3 s^{-1}$

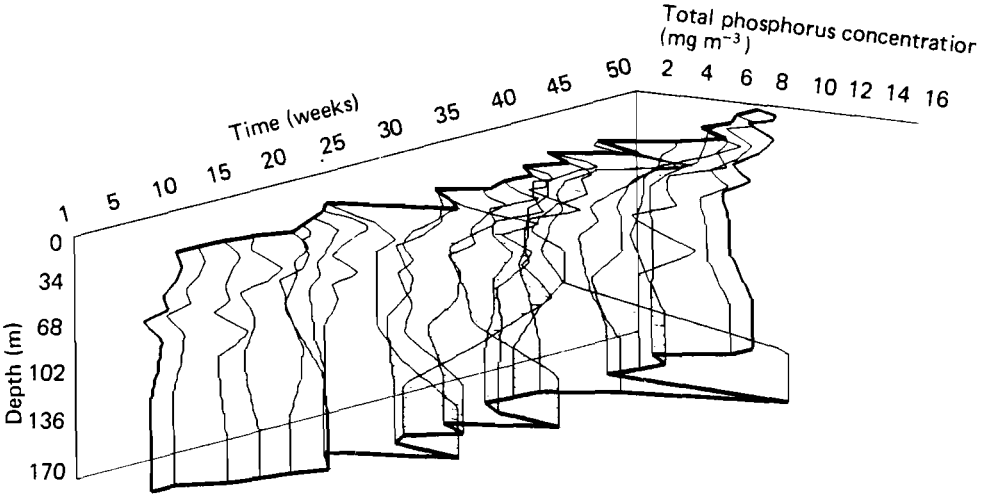

FIGURE 10 Total phosphorus in the Attersee: average yearly pattern (1975–79) of the depth distribution. Since total phosphorus is essentially conservative in the hypolimnion of such a large lake, the observed data variability represents sampling errors due to patchiness (raw data: ÖEP – Salzkammergut).

3.2.1 The Simulation Model

Rather than developing one more simulation model for this study, the dynamic lake phosphorus model by Imboden and Gächter (1978) was chosen for the prediction of the relationships between nutrient loading and water quality. The model predicts primary production per unit lake area as related to imports of soluble (reactive) as well as particulate (algae biomass) phosphorus, various forcings, and model parameters. The relationship between loading and primary production is described by means of a dynamic, one-dimensional, vertical (multilayer) diffusion model for the two state variables, particulate phosphorus and soluble reactive phosphorus concentrations. The

model uses Michaelis–Menten kinetics and self-shading by algae, together with a production rate that varies in time according to the seasonal variations in irradiance and water temperature. Respiration, sedimentation, stratification with vertical eddy diffusivity and variable thermocline depth, lake morphometry, and hydraulic loading are all accounted for in the model. A homogeneous, well mixed epilimnion is assumed, and phosphorus export is determined by its epilimnion concentration and by hydraulic loading. Zooplankton is not explicitly included in the model; its effects on phytoplankton are included in the first-order loss term that describes respiration/remineralization. Consequently, the model is designed more for the simulation of yearly aggregate features than for the simulation of short-term algal population dynamics. Figure 11 shows a flow diagram for the model.

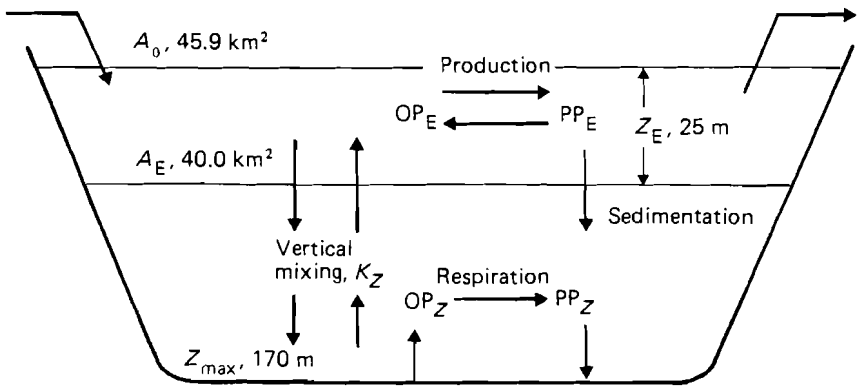

FIGURE 11 Attersee model: flow diagram.

Some minor modifications of the model were made in order to allow for a parameterized description of time-varying forcing functions (production rate and thermocline depth). Rather than specifying these coefficients in the form of tables, as was done originally for the model, we approximated the time patterns by simple analytical functions of time. Thus, the dynamic pattern of the production rate is described by a sine function with the minimum, maximum, and time of maximum as auxiliary parameters. Similarly, thermocline depth is a linear function of time, characterized by the depth and time at the onset of stratification and the depth and time at the end of the stratification period. Other potentially time-varying data (e.g. nutrient inputs, hydraulic loading, eddy coefficients) were kept constant, since the available field data did not allow a meaningful yearly pattern to be specified. In view of the morphology and the current trophic state of the Attersee, the back-flux of phosphorus from the sediments was set to zero in the model.

Ultimately, a total of 22 parameter vector elements (e.g. rate constants, forcing-function-related parameters and initial conditions) were required in this application. These are listed by name in Table 7, together with the ranges sampled in the Monte Carlo simulation. The minimum and maximum values,

which define the ranges, were obtained either from the known variability of available estimates (e.g. particulate phosphorus loading) or from expansion around values given in the literature. The results of the method are not critically influenced by the ranges selected, as long as they are ecologically or physically feasible. However, reduction of the ranges wherever possible is useful for increasing the efficiency of the computation. Thus, for several of the parameter vector elements the ranges in Table 7 were obtained after reduction on the basis of an initial set of 10,000 pilot runs (Section 2.5).

TABLE 7 Parameter vector elements and the ranges used.

Data type	Minimum	Maximum
Parameters sensu stricto		
1 Michaelis constant (phosphorus) (mg m^{-3})	0.20	2.00
2 Respiration/mineralization, epilimnion (day^{-1})	0.02	0.20
3 Respiration/mineralization, hypolimnion (day^{-1})	0.01	0.025
4 Net sedimentation rate, epilimnion (m day^{-1})	0.01	0.75
5 Net sedimentation rate, hypolimnion (m day^{-1})	0.025	2.00
6 Diffusion coefficient, hypolimnion (cm^2 s^{-1})	0.02	0.50
7 Diffusion coefficient, thermocline (cm^2 s^{-1})	0.01	0.25
8 Extinction coefficient (m^{-1})	0.20	0.40
9 Self-shading coefficient (m^2 mg^{-1})	0.01	0.02
10 Thickness of thermocline (m)	5.00	10.00
Data describing imports and forcings		
11 Orthophosphate import (mg m^{-2} day^{-1})	0.01	0.20
12 Particulate phosphorus import (mg m^{-2} day^{-1})	0.25	1.50
13 Hydraulic loading (m day^{-1})	0.03	0.05
14 Minimum production rate (day^{-1})	0.25	0.50
15 Maximum production rate (day^{-1})	1.00	10.00
16 Time lag of production maximum (day)	180	270
17 Initial thermocline depth (m)	3.00	6.00
18 Final thermocline depth (m)	15.00	20.00
19 Start of stratified period (day)	120	280
20 End of stratified period (day)	280	330
Initial conditions		
21 Orthophosphate, mixed period (mg m^{-3})	0.20	2.00
22 Particulate phosphorus, mixed period (mg m^{-3})	2.50	7.00

The simulation model was incorporated as a subroutine in a control program, which generated random sample parameter vectors from the ranges specified. Since *a priori* information on the probability distributions and correlation structure of the parameters was absent, independent rectangular distributions were assumed. For each parameter vector, one simulation run was completed (for a period of one year) and the model response stored for subsequent analysis.

3.2.2 Behavior Definition

The output of any given model run has to be compared with the (defined) systems behavior in order to allow classification of the parameter vectors into a behavior-giving set and a set that does not result in the observed behavior. Obviously, the definition of the behavior of the system is a crucial step in the analysis. The systems behavior definition should reflect all knowledge of the system that is available and relevant (in terms of the problem and the conceptualization of the system, i.e. the model). It is worth noting that a definition of systems behavior (the empirically defined region in systems behavior space) derived from the observations does not depend upon the model. However, the allowable model response set has to be specified in terms of model output constraint conditions, or, in other words, the behavior definition must be cast within the framework of the model actually used.

The behavior definition uses ten constraint conditions describing a region in a seven-dimensional behavior space for the model: the constraints are defined for yearly primary production, algae biomass peak (maximum and timing), relative increase of algae, orthophosphate maximum during the mixed period, yearly phosphorus output, and cyclic stability of total phosphorus (maximum relative difference between beginning and end of the simulation year). The constraints placed on these indices for the purpose of behavior definition were specified such that the measurement uncertainty and the natural stochastic variability of the ecosystem (including variability among the years) were accounted for. The resulting behavior definition is given below:

1. Total primary production per year has to be between 50 and 150 g C m^{-2}.
2. Total phosphorus export per year has to be between 2 and 8 tons.
3. The peak value of particulate phosphorus in the epilimnion has to occur between Julian days 60 and 210.
4. The peak value of particulate phosphorus in the epilimnion must not exceed 15 mg P m^{-3}.
5. The concentration of orthophosphate during the mixed period must not exceed 2.5 mg P m^{-3}.
6. The peak value of particulate phosphorus must be at least twice the minimum value.
7. The maximum total phosphorus content of the lake during the year must not exceed twice the minimum value.

In this way the behavior definition can be viewed as a seven-dimensional box in the behavior space and a model simulation run has to lie completely within this box in order to be classified as a simulation belonging to the set of allowable model responses.

3.2.3 Analysis: Parameter Vector Space Structure

Out of 10,000 runs only 293 parameter vectors were found that gave rise to a model output fully within the behavior constraint conditions given in the previous section. Inspection of the sample ranges of individual elements of the 293 vectors showed that no further "rectangular" reduction of the parameter vector space (*PD* in Figure 4) was possible. In other words, the boundaries of the behavior-giving parameter vector space region (*PM*), as sampled by the 293 behavior-giving vectors (*PS*), extended to the boundaries of the 22-dimensional parameter vector box (*PD*). Figure 12 shows the distribution of the behavior-giving values for each of three parameter vector elements in order to illustrate this point. The figure also suggests that there are regions in the parameter vector space (*PD*) where one is more likely to find an allowable model response than in others.

The high dimensionality of the parameter vector space means that the geometry of the behavior-giving region is in general difficult to investigate. However, a tentative exploration of the distributions can be made by projecting these distributions on to a two-dimensional surface. An example was given in Figure 3. From this figure it is apparent that the behavior-giving parameter combinations are more densely clustered in certain regions. It is also evident that other regions in the 22-dimensional box are "empty." However, in view of the large number of combinations that are not behavior-giving, we must conclude that almost every individual value of a parameter vector element can give rise to the behavior or not, depending on the sample values of the other elements. Thus, as also suggested by Figure 3, it is rather the (multiple) correlations between the parameter vector elements that determine the shape of the behavior-giving parameter space.

To gain insight into the structure of the model (and, it is hoped, of the system) a correlation analysis was performed. Consequently, 13 of the 22 parameter vector elements were found to be significantly correlated with one or more of the other parameter vector elements. The most complex relations, with four or five significant pairwise correlations, were found for the respiration/mineralization rate in the hypolimnion, the net sedimentation velocities, the particulate phosphorus import, and the hydraulic loading. Also, production rate maximum and time lag showed more than one significant correlation.

Correlations between the parameter vector elements of the behavior-giving class reflect the ability of the model to balance one extreme with another, while still fulfilling the behavior definition constraints. This immediately makes it obvious that they cannot be estimated individually. A typical example was presented by Simons and Lam (1980), who demonstrated the ability of a lake model to balance a considerable change in the nutrient loading with some change in the settling rate in order to obtain basically the same model response. Parameter vector elements that would force the behavior-defining variables in the same "direction" (relative to the boundaries of the seven-dimensional behavior box) can be expected to be negatively correlated, and vice versa for the positive correlations. In this way, the strong positive correlation of particulate phosphorus import and sedimentation velocity (epilimnion), for example, indicates that the constraint

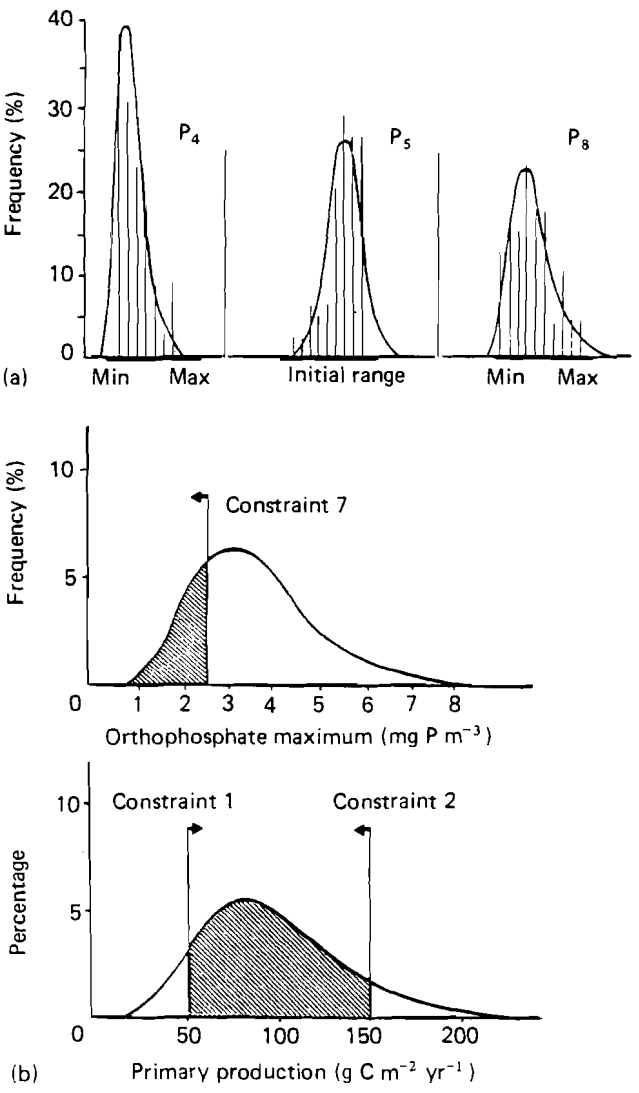

FIGURE 12 (a) Frequency distributions for individual parameters (thick bar indicates range sampled) for (b) the shaded response ranges. P_4, net sedimentation rate, epilimnion; P_5, net sedimentation rate, hypolimnion; P_8, extinction coefficient.

variables yearly primary production and/or the allowable algal biomass peak value are sensitive to the "net" effect of these counteracting processes. The constraint of maximum allowable orthophosphate concentration provides another example. Hypolimnetic remineralization – a major process affecting the orthophosphate concentration – is consequently negatively correlated with orthophosphate import and with the initial phosphorus concentrations.

In geometric terms it can be said that the correlations indicate the main axes along which the behavior-giving parameter vector region (*PM*) is

oriented. Consequently, the model response (of giving the defined behavior) is most strongly influenced by varying the parameter values in a direction orthogonal to these axes. In this sense the correlation matrix can also be interpreted in terms of a sensitivity analysis.

According to the parameter vector correlation structure, the Attersee system, as defined by its geomorphology and the behavior definition, is characterized by a delicate balance between the processes responsible for primary production and phosphorus export (which is mainly determined by the epilimnetic phosphorus concentration) and those determining the orthophosphate peak concentrations, namely (besides the imports) sedimentation to the comparatively large hypolimnion and hypolimnetic remineralization. This balance can only be achieved with a high phosphorus turnover in the epilimnion and comparatively slow net remineralization in the hypolimnion. For a lake in the geographic position of the Attersee and with the Attersee's morphometric features and associated temperature distributions, this seems to be a reasonable interpretation.

3.2.4 The Role of the Behavior Definition

Choosing values for the constraints on the allowable behavior patterns is subject to solving a two-sided problem. On the one hand, the constraint ranges should be sufficiently narrow to restrict the allowable patterns such that they unambiguously represent the empirical behavior of the system in a meaningful way. On the other hand, all the variability in systems behavior and the uncertainty in the observations should be taken care of with a minimum of arbitrariness. Since a reconciliation of these two objectives is rather difficult in practice, the effects of the constraint setting on the parameter vector classification were examined. For this purpose, the model response space was projected on to the individual model output variable axes. The positions of the constraints in relation to the resulting frequency distributions (Figure 12) give some indication of the relative importance of the individual constraints. Figure 2 shows two contrasting examples for a pair of critical constraints and a pair of uncritical constraints.

The original constraint values were altered and the effects on the parameter vector separation were studied by recording violations of the constraints. For the standard set of constraint values the numbers of violations together with a relative coincidence matrix of violations are shown in Table 8. Clearly, the allowable phosphate level and the first permissible day for the algal peak are the major constraints on achieving an overall "acceptable" model response. Some of the other constraints are either not violated at all, for example, minimum relative biomass increase, or are rarely violated, such as the upper limit of total phosphorus output. There are also some notable relationships in the violations observed. For example, in almost all cases in which condition 5 is violated (upper limit for biomass peak) so too are conditions 1, 3, and 8 violated (primary production too low, biomass peak too early, phosphorus export too low); however, only 1% of this subclass violates the most critical condition, 7, on maximum phosphate level. Excessive primary production always occurs together with an excessive level of phosphate, and about half of this subclass gives the biomass peak either in the required

TABLE 8 Constraint violations (standard definition, 10,000 runs).

Condition	Number of cases
1 Primary production too low	1,237
2 Primary production too high	904
3 Biomass peak too early	5,108
4 Biomass peak too late	1,492
5 Biomass peak too high	357
6 Relative biomass increase too low	0
7 Phosphate level too high	7,201
8 Phosphorus export too low	2,398
9 Phosphorus export too high	1
10 Relative change in phosphorus content too high	2,148

Coincidence matrix of constraint violations (%)

	1	2	3	4	5	6	7	8	9	10
1	100.0	0.0	77.8	8.2	28.5	0.0	35.6	67.7	0.0	17.5
2	0.0	100.0	13.3	39.5	0.4	0.0	100.0	0.0	0.1	9.5
3	18.8	2.4	100.0	0.0	6.9	0.0	60.5	36.5	0.0	34.1
4	6.8	23.9	0.0	100.0	0.3	0.0	89.1	8.1	0.1	13.7
5	98.9	1.1	98.9	1.1	100.0	0.0	1.1	98.9	0.0	0.6
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	6.1	12.6	42.9	18.5	0.1	0.0	100.0	10.1	0.0	13.6
8	34.9	0.0	77.7	5.1	14.7	0.0	30.2	100.0	0.0	36.0
9	0.0	100.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0
10	10.1	4.0	81.0	9.5	0.1	0.0	45.7	40.2	0.0	100.0

interval of time or too late (40%). Excessive variations in the total phosphorus content of the lake are mostly paralleled by too early a biomass peak, but almost never occur in conjunction with an excessive peak value.

Changing condition 7 from the initial 2.5 mg P m^{-3} to 3.0 mg P m^{-3} allowable in the mixed period decreased the number of violations of this condition from 7,201 to 5,680, and resulted in 665 "behavior-giving" parameter vectors. Thus, 372 of the 1,521 vectors located in that "interval" (compare with Figure 2) do not violate any other condition. Further change in the allowable phosphate level from 3.0 to 3.5 mg P m^{-3} increased the number of behavior-giving vectors to 1,127, with 4,126 residual violations of the constraint condition, indicating a fraction of about 500 potential "behavior" vectors within that interval. For comparison, a reduction of the allowable value from the original 2.5 to 2.0 mg P m^{-3} decreased the number of behavior runs quite dramatically to 68, with a corresponding number of constraint violations of 8,565. Again, a considerable number of the vectors in that interval (more than 1,000 of the total of 1,350) were already violating at least one other constraint condition. In addition, as another example, changing condition 3 from day 60 to day 50 resulted in only two additional "behavior" vectors, although the number of violations of condition 3 dropped from 5,108 to 5,074. The remaining 32 samples thus give a model response that violates at least one other condition.

In conclusion it can be said that although the specification of some of the constraint conditions is rather critical for the resulting parameter vector separation, the high degree of coincidence makes the method less sensitive to the individual conditions. This analysis may give some indication of where further efforts in data collection or analysis should be concentrated. Admittedly, however, evaluation of the sensitivity of the approach to the behavior definition should be carried out in terms of response probability distributions for predictions. This remains to be done.

3.2.5 Projections into the Future

Having established a set of "model calibrations" for the range of empirical conditions covered by the behavior definition, we can now use this ensemble for making predictions of the response of the lake system to changes in nutrient loading. The mean total phosphorus loading in this "empirical" ensemble was estimated to be $1\,mg\,P\,m^{-2}\,day^{-1}$ (standard deviation 0.33), which corresponds well with independent field estimates (Müller 1979). For the predictions, the loading (parameter vector elements 11 and 12 in Table 7) was varied systematically from 0 to 5 mg P m^{-2} day^{-1} in steps of 0.25. The proportion of available phosphorus in the total loading was set to 10% after a series of runs in which ratios of 0, 10, and 25% were compared. For each of the 21 new loading values the first 150 sample parameter vectors from the behavior-giving set were taken, thus generating a set of 150 estimates for several output variables (yearly primary production, algae peak biomass, phosphate maximum, phosphorus export, and phosphorus sedimentation) for each loading value for a series of 10 years. Figure 13 summarizes the results for primary production, showing the situation after years 1 and 10.

Another set of predictions of future systems response to changes in the phosphorus loading conditions was made by subsets of the behavior-giving parameter set, where the load-determining parameter values were changed by a certain factor. This "relative" change not only accounts for the uncertainty in the parameters, but also preserves the correlation structure of the behavior-generating set of parameter vectors. Input changes representing increases of 50 and 100% (to simulate the effect of no control actions but increasing nutrient release in the catchment area) and reductions to 75, 50, and 25% of the 1975–78 empirical range of loading were simulated for a 10-year period. Some examples of these scenarios, again showing the stochastic mean with a minimum/maximum envelope, are given in Figure 14.

To estimate prediction accuracy as related to the changes in the phosphorus loading (the degree of extrapolation in input space), and as related to simulation time (the extrapolation in time), the coefficient of variation was plotted against these extrapolations. Figure 15 shows one example for the model output variable, yearly primary production. The plot shows an increase of prediction uncertainty with time, stabilizing when a new equilibrium is reached after a transient period following the change in the phosphorus loading. The plot also indicates an increase of uncertainty with the amount of change in the input conditions, showing a minimum of the coefficient of variation in the empirical range. In summary, prediction uncertainty (measured as the coefficient of variation of the Monte Carlo

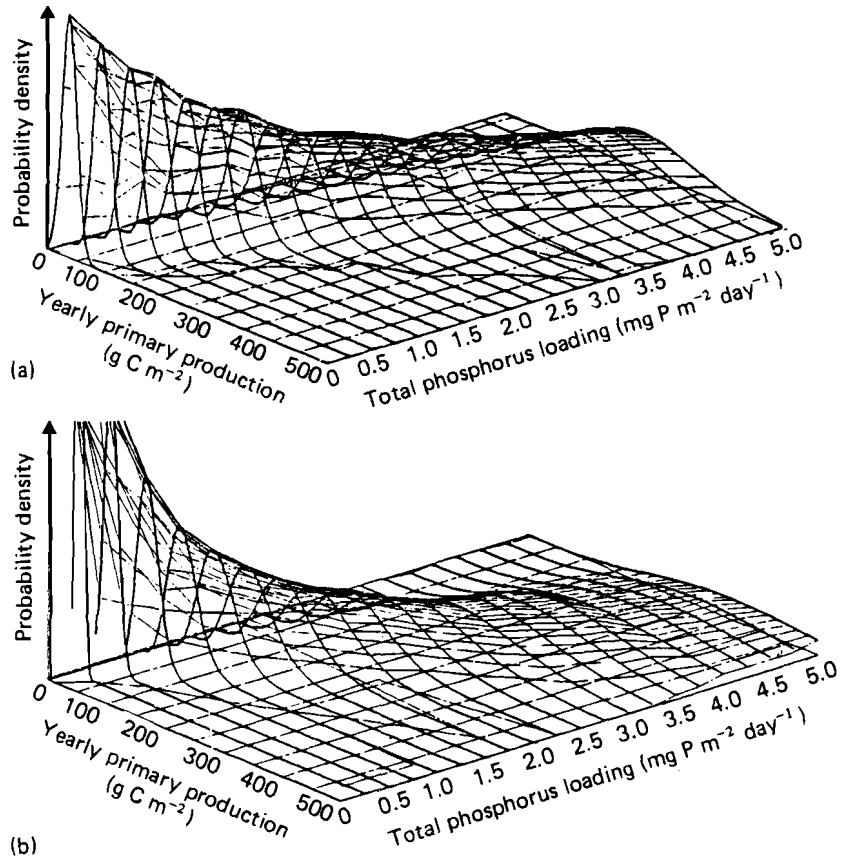

FIGURE 13 Probability distributions for model output variable, yearly primary production, for different total phosphorus loadings, in the Attersee phosphorus budget model. (a) First year of lake response; initial state represents the empirical range of lake behavior. (b) Lake response after 10 years of changed phosphorus loading. Note the extremely flat distribution in the high-loading classes.

ensembles) increases with the extrapolation in time as well as in input space. Since prediction reliability is related to the initial variability in the descriptive empirical case, there is an obvious (and intuitively to be expected) relation of prediction reliability or nontriviality to the magnitudes of the input variability (incorporating data uncertainty and systems variability in time), the degree of extrapolation in the controlling inputs, and the degree of extrapolation in time. Obviously, the more precise the original knowledge about the system is, the larger the extrapolation in the controlling conditions and in time can be, before the limits of predictability are reached; or, the larger the change that is to be simulated, the better the knowledge about the system has to be.

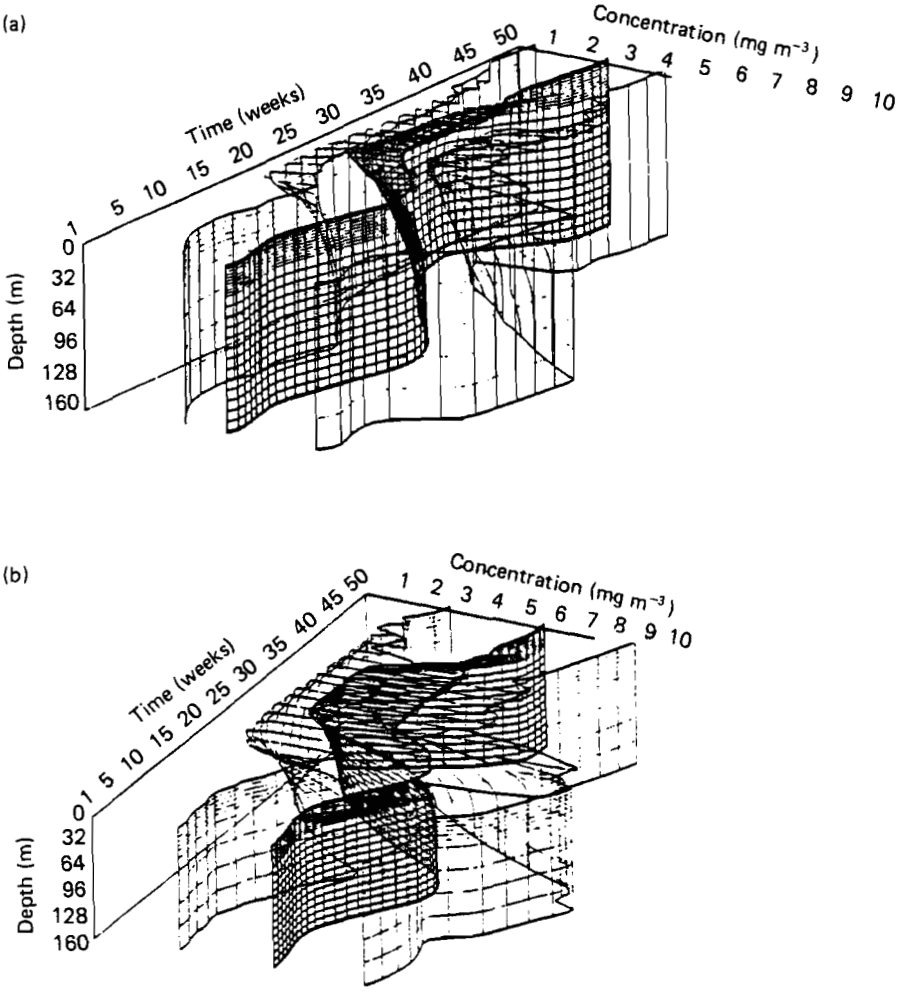

FIGURE 14 Ensemble of behavior runs for the Attersee model. (a) Mean of 39 runs with minimum/maximum envelope. (b) Prediction in terms of envelopes.

A different representation of prediction accuracy was shown in Figure 13 (where prediction refers to the mean estimate, and accuracy is measured in terms of confidence intervals). The probability distributions fitted for the response variable frequency distributions can be read in the above terms. These probability distributions are not primarily to be understood as the probabilities of certain systems states in the future – they are rather representations of prediction uncertainty, or the propagation of the initial uncertainty and variability in the available information.

The above analysis and the generalizing conclusions to be drawn are certainly biased with regard to the model used and, to a lesser extent, with regard to the data set used. The arbitrary selection of any model for a given

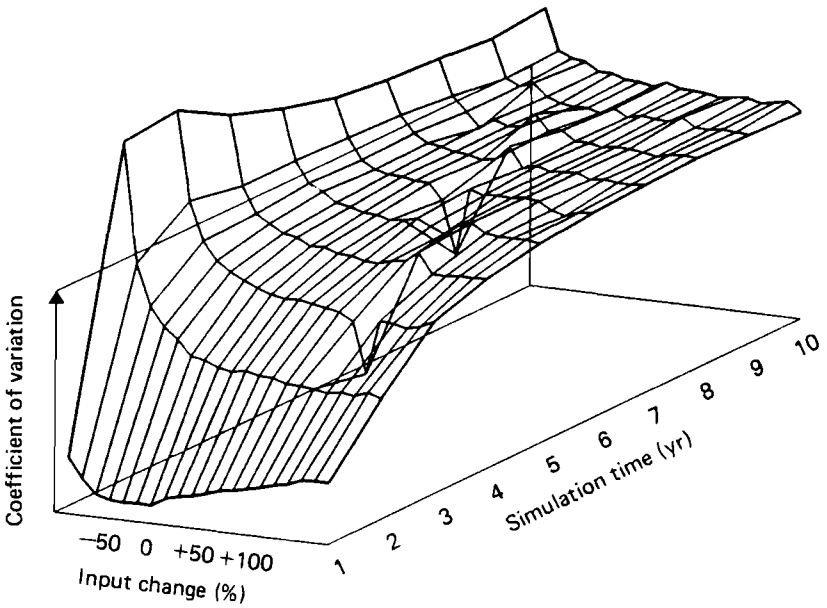

FIGURE 15 Representation of prediction uncertainty (coefficient of variation for an ensemble of plausible model runs) plotted against input change (extrapolation) and time.

system seems to be unavoidable in light of the meager data available; the model order and structure cannot be derived from the available data, and one has to use *a priori* information about the system to be described. However, the thus conjectured model might well turn out to be inadequate, and changes in the model structure will become necessary. One indication of inadequate model structure, in terms of the above approach, would be that no behavior-giving parameter combination can be found in the specified region (compare the above example); or that the distributions of the single parameters within the ranges sampled suggest a high number of possible solutions outside the specified "plausible" bounds. If a combination of unrealistic inputs still results in a realistic behavior of the model, one has to question the validity of the model structure. This of course requires that the expected behavior is defined in sufficient detail.

As indicated in Figure 15, the probability distributions in the higher-loading classes level out with time. If one takes the coefficient of variation as a measure of prediction uncertainty, a saturation-curve-type pattern in time can be observed for this measure (see also below). When this coefficient of variation is plotted against loading for the first year's response, a distinct minimum – in the empirical (observed) range of loadings – can also be observed (Figure 15). One may conclude, therefore, that prediction uncertainty increases with extrapolation away from both the present time and the presently observed input loading conditions. Certainly, the predictions for larger deviations from the empirical situation are rather trivial after only a

few years: an estimate of yearly primary production between 100 and 1000 g C m^{-2} for a phosphorus loading of 5 mg m^{-2} day^{-1} is certainly of little value as a prediction. However, it should serve as a warning to the analyst that the uncertainty in the data available or the variability of the system itself simply does not support such an extrapolation.

The samples of predictions can now be interpreted in terms of the original problem setting. Only the loading values close to the empirical range result in meaningful distributions in the 10-year projections, but these are of course the most interesting and "realistic" alternatives to be studied. At the end of the 10-year period of simulations, the lake system is found to be in a kind of new dynamic equilibrium with regard to the output variables considered. Whereas the variability of the predictions rapidly increases during the transient stage of the first six years (or less in some cases) after a change in the loading, this variability stabilizes by the end of the simulation period (the somewhat unrealistic zero-loading class is omitted from these evaluations). The time to reach a new equilibrium was found to be connected with the relative change in the loading. The comparison of different loadings in terms of primary production, peak biomass, or phosphate level can now be made by comparing either the mean levels and their confidence intervals or the probabilities for reaching or exceeding certain levels. This is especially interesting, as in fact almost all of the probability curves fitted are skewed, which clearly implies that a simple comparison of mean values might be misleading for certain problems.

The above analysis indicated that for a model with 22 input data (which, at least for ecological models, is a rather low figure) or "degrees of freedom" in the estimation procedure, behavior-giving values can be found all over the ranges (independently) sampled. On the other hand, only a small percentage of the possible combinations resulted in a satisfactory model response. As a consequence, the ranges for the search should be constrained as much as possible, for reasons of efficiency as well as to avoid "unrealistic" input data combinations (where the unrealistic value in any of the parameters or inputs will be "balanced" by some changes in all the other values) in the behavior ensemble. This of course requires that all the parameters used in the model are physically interpretable and can be measured or at least estimated from field measurements or experiments. The same holds true for the state variables of the model and measures derived: only if they are measured (or are at least measurable) can their allowable values be reasonably constrained in the definition of the behavior of the system. Including unmeasured (and unconstrained) state variables will result in behavior runs (in terms of the constrained measures) where the uncertainty is all transferred to this unconstrained "leak" in the behavior definition (compare the following application example). The ability of even a simple model to balance its (constrained) response in terms of some variables by (unconstrained) changes in others requires that all model behavior (and, of course, output) should be interpretable in physical (measurable in the field) terms. Also, the approach described above raises some doubt whether models, by including more and more detail (requiring more and more state variables and parameters, and consequently more data for the "calibration"), become more realistic. Obviously,

increasing model complexity without increasing the available data for constraining input data ranges as well as allowable response ranges just adds degrees of freedom for the calibration or estimation procedure. Undoubtedly such models can be very useful, especially in more qualitative "hypothesis-testing" approaches. But their value for prediction might well be questioned.

3.3 The Limits of Estimation: A Simple Rain–Runoff Model

For the study of water quality problems of lake systems, generally due to eutrophication resulting from excess nutrient inputs, the more traditional approaches concentrated on load–response models of various degrees of complexity and resolution (Park *et al.* 1974, Chen and Orlob 1975, Imboden and Gächter 1978, Vollenweider and Kerekes 1980; see also Section 3.2). These were primarily designed to predict the changes in lake water quality as a result of changes in the nutrient loading. Lakes, however, are only one element in regional water resource systems; they have to be understood as being linked to a physical as well as a socioeconomic watershed, which affects them and vice versa; and, by their outflow, they affect also downstream water bodies.

As most of the control options for lake water quality are based in the lake catchment, this catchment has to be included in the analysis. Land use, i.e. agriculture and silviculture, and domestic, commercial, and industrial activities, tie up with lake water quality by affecting the quantity and quality of the runoff from the catchment. The pollutant loads in the runoff, with or without treatment, form the inputs to the lakes. To predict the effects of any change in land-use patterns, or the effectiveness of any control action such as sewering and interceptor systems, models to do so have to explicitly include these activities. Consequently, watershed models and lake water quality models have to be coupled, if the input to the lakes is to be traced back to its natural and man-made causes.

To be useful in a framework oriented toward planning and management, simulation models have to incorporate the relevant planning and management variables with a time and space resolution appropriate to the structure and dynamics of the real system. The time and space scales should be appropriate for the planning and management decisions that can be implemented.

On the other hand, there are always severe limitations on the data available. Hydrometeorological networks are usually very coarse in relation to the scale of regional or local problems, which is even more pronounced under complex orographic conditions such as in the mountainous areas of Austria. Also, measurements of precipitation, temperature, and flow are often made on a daily (or 12-hour) basis. Especially in the case of flow measurements, these data may contain considerable errors (e.g. Winter 1981).

Many hydrological models, in contrast, require extensive input data, but models calling for data that are just not available are of little practical use. Therefore, while a sufficient degree of disaggregation should be maintained on one side, modesty in data requirements on the other is an absolute must for a

useful approach. Consequently, the simulation model discussed below operates on a daily time step, using daily input values for precipitation and temperature, and a very coarse and lumped representation of major processes. Only one precipitation value for the whole watershed is used (which, however, is corrected for changes in altitude within the catchment). If no observations are available from the catchment, data from the nearest hydrometeorological station will have to suffice. The same applies to daily average air temperatures. All the additional information required on basin morphology and land use can normally be obtained from standard maps. It is important that this information is readily available to the analyst and does not require the launching of an expensive and time-consuming measurement program in the field. A useful tool for those responsible for current planning and management should also be able to answer questions more or less immediately, without requiring additional years of research to compile the necessary, problem-specific input data. The basic design principle of the model is ruthless or sometimes naive parsimony in the description of the individual processes; the basic constraint is imposed by data availability and thus testability of the model formulations.

Another reason to strive for maximum simplicity is computational efficiency and cost. Whenever a simple version is sufficient within the limits of detail set by the purpose, a more complex one is obsolete. Although computer time is not a real constraint as compared with data availability, the linking of numerous modules for a comprehensive system simulation requires simple elements if the whole is to be operational. As one of the purposes of such simulation systems is to learn about the systems behavior by performing numerous and extensive numerical experiments, comparing management alternatives, or exploring the (model) systems sensitivity to different assumptions, parameters, data, or the lack thereof, small and quick is beautiful. Small and quick may also be an essential requirement for the effective communication of results. Simulation models are, at best, one element in an array of methods and approaches for planning and decision making; therefore, to be used, they have to be attractive to the user. This means they should be quick and easy to use and of comparatively low cost, and should generate results in a format that is interactive, attractive, and educational.

3.3.1 A Description of the Model: The Basic Concepts

The rain–runoff model used for this example is based on the principle of mass conservation. It performs a dynamic accounting of water in different parts of the watershed, i.e. the soil–channel system, transforming precipitation in the watershed into runoff at its outlet. Figure 16 is a flowchart of the model, indicating the basic elements and the major processes considered. A detailed description of the model, including several examples of application, is given by Fedra (1983).

The model is spatially dimensionless. The whole soil system is conceptualized as one single block (or rather column) of soil, with its lumped properties derived from land use as weighted averages. Designed for temperate mid-latitude zones, the model explicitly describes snow accumulation and snowmelt. Water available for runoff, that is, water in excess of infiltration

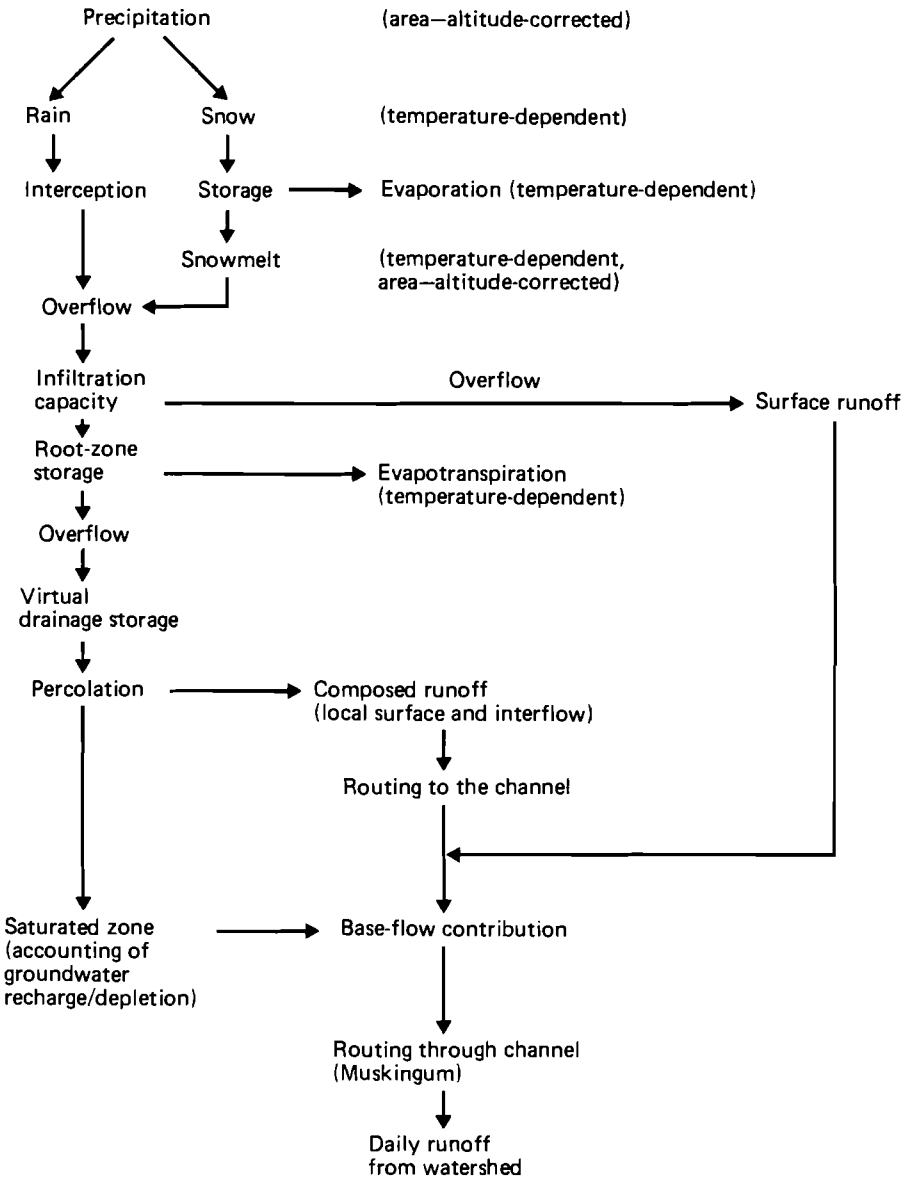

FIGURE 16 Flowchart of the model.

capacity, or, in the root zone, in excess of field capacity, and after allowing for loss due to evapotranspiration, is divided between percolation and runoff, using a concept of variable source areas. The runoff is routed from this column of soil to the channel, and from the channel to the outlet. Percolation feeds the groundwater reservoir, which in turn contributes to channel flow by a base-flow component, coupled to the groundwater budget. Groundwater is only simulated as a balance between percolation and base flow. A critical assumption of the model is that all groundwater will leave the catchment through the channel as base flow; no other groundwater inputs or outputs are considered, that is, surface and underground recharge areas are assumed to overlap fully.

The rationale for this simplification is mainly based on the problem of determining groundwater budgets experimentally; if estimated as the residual term in the water budget (which is exactly what the model does), they are as uncertain as all the factors considered in this water budget. Only if independent measurements exist, can groundwater interaction between neighboring catchments be reasonably included in the model.

3.3.2 Inputs and Parameters

The model requires the following inputs to be specified for the characterization of a watershed:

a. *Watershed characteristics*
   Drainage area (km^2)
   Landcover: forests (%)
   Landcover: agriculture (%)
   Landcover: pastures (%)
   Basin length (km)
   Altitude difference (m)
   Main-stem channel length (km)
   Total channel length (km)
   Number of elevation bands considered
   Height of each elevation band (m)
   Area in each elevation band (%)

Drainage area, or catchment size, can easily be determined from an appropriate map with contour lines. The landcover information can be obtained from the same source, or from aerial photography, or satellite remote sensing (e.g. Salomonson and Bhavsar 1980). The basin length represents the longitudinal (along the major channel) dimension of the catchment, as would result from a rectangular approximation of its shape. Similarly, the main-stem channel length should approximate the total (or average, if more than one) length of the major draining channel, which will usually be somewhat shorter than the basin length. Altitude difference is just the difference between the average elevation of the watershed boundary "opposite" the basin outlet and the lowest point (the outlet) of the watershed; however, since the estimates of average slopes will be based on these figures, some flexibility in their estimation, representing special shapes of a

catchment, may be necessary. The total channel length describes all observable permanent channels in the catchment, including the main channel. It is used to estimate drainage density. The areal altitude distribution, which is used for correcting precipitation, temperature values, and snow cover, is read in as a number of percentages of the areas of elevation bands, proceeding from the lowest (at the level of the outflow) to the top of the watershed.

In addition to these watershed characteristics, the model requires the specification of initial conditions.

b. *Initial conditions*
Initial flow ($m^3 s^{-1}$)
Initial base-flow contribution ($m^3 s^{-1}$)
Initial snow pack (rain equivalent) (mm)
Initial interception storage (full/empty)
Initial soil moisture (%)

Finally, six parameters have to be estimated. For the estimation procedure (Section 3.1) ranges have to be specified; acceptable parameter estimates have to be within these ranges.

c. *Parameters*
1. Altitude correction factor for precipitation
2. Altitude correction factor for temperature
3. Field capacity (root zone) (mm m^{-1})
4. Maximum percolation rate (mm day^{-1})
5. Average runoff speed (root zone + surface) (m day^{-1})
6. Groundwater response time lag (day)

1. The altitude correction factor for precipitation will increase the raw input value (thought of as representing the lower end of the watershed) for a given percentage per elevation band. Altitude correction factors for precipitation will usually be of the order of a few percent per hundred meters.
2. The altitude correction factor for temperature is used to represent the temperature gradient with altitude within the catchment. The parameter value represents the average difference between temperatures at the lowest and highest elevation bands. It can be estimated from the differences in average temperature between measurement stations at different altitudes, or approximated with a value around 0.5 °C per hundred meters, times the altitude difference of the basin.
3. Field capacity, i.e. the amount of water the soil can retain against the pull of gravity, is a function of soil properties. Depending on the type of soil, values can range from 20 mm m^{-1} for sandy soils to 300 mm m^{-1} for clays.
4. Percolation, i.e. the downward movement of water under hydrostatic pressure, conveys water in excess of field capacity from the root zone to the groundwater. Percolation rates again depend largely on the type of soil and on moisture content. Literature

values for the coefficient of permeability range from 10^7 mm day^{-1} for gravel to 10^{-3} mm day^{-1} for clay.

5. The velocity of runoff is used to estimate the travel time of water to the channel system. This composed runoff is based on the concept of a variable source area, and incorporates small-scale surface runoff (large-scale runoff for extreme precipitation events is accounted for in a different manner), return flow, runoff through small ditches and temporary channels, and finally subsurface storm flow or interflow (Chorley 1978). Since such a large variety of different processes with pronounced local variability on a scale much below the resolution of the model are lumped in this one number, any estimation from watershed characteristics will be difficult. However, the distribution of slopes, surface roughness, amount of detention storage, proportion of impervious areas, and the structure of the drainage system (i.e. drainage density) are important factors. A simple first estimate can be based on the average time lag of a flood wave after a rain event. The average distance to the nearest channel divided by this lag time is a rough estimate of the average runoff speed, which will be of the order of a few hundred meters to kilometers per day.
6. The groundwater response time lag determines the rate at which the base-flow contribution to the channel flow will react to changes in the groundwater budget. The value depends on catchment size and geology and is of the order of weeks or months.

3.3.3 Parameter Estimation

The model, as described above, requires six "free" parameters to be estimated. These six parameters represent physically meaningful magnitudes, and, at least in principle, are measurable or could be experimentally determined. Consequently, the values of these parameters have to be within physically or hydrologically feasible ranges, which should be related to certain watershed characteristics as discussed above.

In the estimation procedure, these ranges are used as the starting point for the calibration. For each parameter an allowable range is specified, within which acceptable solutions to the estimation problem are constrained. From these ranges, independent random samples are drawn and used for one run of the model. This run can extend over any arbitrary period in time, for which a set of inputs (precipitation and temperatures) and a set of outputs for comparison (runoff from the catchment) are available. This simple Monte Carlo method is repeated for a sufficient number of trials, in order to minimize a certain objective function or to meet a set of performance criteria. The result of these trials, which may be repeated in an iterative manner, will be either an "optimal" set of parameters (according to the objective function minimized) or an ensemble of admissible parameter sets, meeting certain performance criteria.

In an estimation run, the model is run several thousand times (each run requires less than half a second of CPU time on a VAX 11/780), and for each

run the parameter set and the performance criteria of the run are recorded for later analysis. In addition, a record of the "best" parameter set (according to a weighted combination of the performance criteria) is kept.

3.3.4 Parameter Ranges

Because of their physical interpretation, the six parameters to be estimated have to be within well defined bounds. These ranges will differ from catchment to catchment, but there certainly exists an overall admissible, or physically plausible, range for each of them. Table 9 summarizes these ranges.

TABLE 9 Admissible ranges for the model parameters.

Parameter		Minimum	Maximum
1	Altitude correction factor for precipitation [$(100\ m)^{-1}$]	0.000	0.500
2	Altitude correction factor for temperature [$°C(100\ m)^{-1}$]	0.000	1.000
3	Field capacity (mm m^{-1})	50.000	300.000
4	Percolation rate (mm day^{-1})	1.000	100.000
5	Speed of composed runoff (m day^{-1})	100.000	1000.000
6	Groundwater response lag (day)	1.000	1000.000

3.3.5 Performance Criteria and Objective Functions

The use of the simple Monte Carlo method described above allows maximum flexibility in the formulation of performance criteria or objective functions. For the parameter estimation runs used in the example described in detail below, a number of very simple criteria were formulated. From a daily error factor (err), an allowable range extending around the observed value (flow) was calculated:

$$\text{upper} = \text{flow} \times \text{err}$$

$$\text{lower} = \text{flow} / \text{err}$$

On any given day, the model-generated runoff is either within or outside that range, the latter being considered a violation. A subroutine of the simulation program keeps track of the total number of violations, violations for critical flow events (i.e. observed runoff above a certain level) or during a special, limited period, e.g. during snowmelt. In parallel, the sum of squares of the deviations from the observations is calculated, together with the first day of a failure to occur, and the maximum difference between observed and simulated runoff. Similarly to the daily error range, progressively narrower ranges for monthly and yearly totals are defined. Again, any subset (e.g. a certain month deemed critical in the context of the analysis) can be specified, for which a performance criterion can be formulated. As a criterion, a maximum allowable deviation from a certain reference value or range, a maximum in

absolute or relative terms, or a maximum allowable number of failures for a certain class of events during a certain period may be specified.

For the application described below, the following performance criteria were used:

Sum of squares
(simulated runoff – observed runoff)^2

Total number of violations
(number of days when simulated runoff was less than half the observed runoff or more than twice the observed runoff)

Number of significant violations
(number of days with an observed runoff above a certain critical level, when simulated runoff was outside the allowable interval)

Several other criteria were defined for some of the calibration runs, e.g. the maximum difference between observed and simulated runoff, the number of days until the first violation in a given run, the relative difference between observed and simulated yearly runoff totals, and a similar criterion for selected monthly totals.

3.3.6 Application Example: A Small, Mountainous Watershed: Kienbach, Upper Austria

The example of an application is a small, mountainous watershed in the catchment of the Attersee, Upper Austria. The catchment extends over only 12.5 km^2, spanning an altitude difference of 540 m on average, the highest point on the watershed being 1,600 m above sea level. Landcover is almost exclusively forest, with some rangeland and pastures. Table 10 summarizes the catchment characteristics.

The driving variables, precipitation and temperature, are taken from the neighboring meteorological station in Weyregg, situated also on the shoreline of the Attersee, some 8 km from Kienbach. Flow measurements are taken from daily gauge readings, converted to flows by means of a simple power function.

There exist interesting and somewhat puzzling relationships between the parameters and the performance criteria, and among the performance criteria themselves. Table 11 compares the parameter values for three small, neighboring watersheds in the Attersee catchment, resulting from 30,000 Monte Carlo runs, sorted for two different objective functions, namely the average sum of squares (ssq) and the number of significant (sv) violations (see above) for storm runoff events.

Whereas the first three parameters (precipitation correction, temperature gradient, and field capacity) are largely unaffected by the choice of objective function, the remaining three, namely the maximum percolation rate, speed of runoff, and the groundwater response lag and damping, change considerably. The two cases of extreme behavior, contrasted in Figure 17, clearly show the difference in the groundwater response.

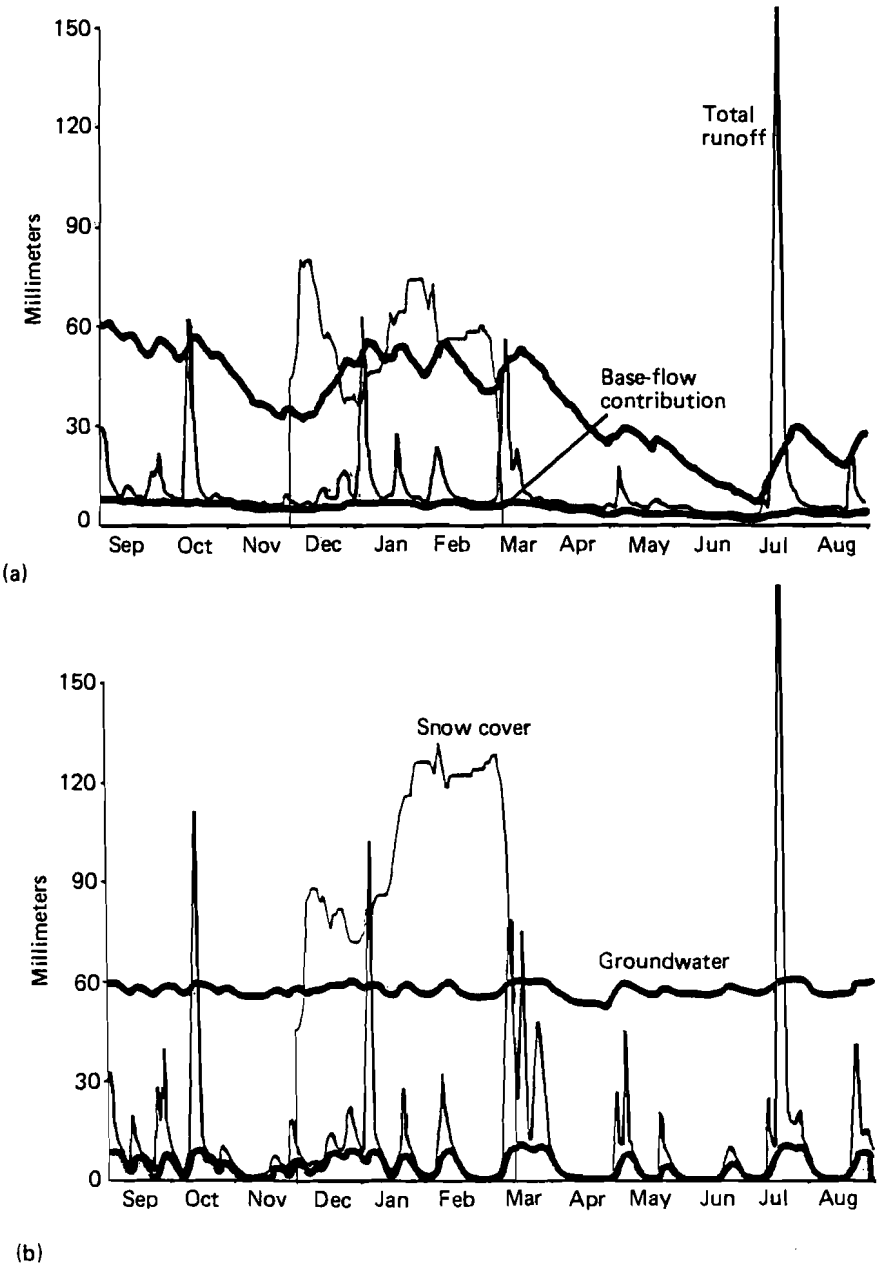

FIGURE 17 Output of rain–runoff model for Kienbach: (a) minimization of sum of squares; (b) minimization of number of violations (storm runoff).

TABLE 10 Watershed characteristics (sample output) for Kienbach.

Inputs and watershed characteristics	
Drainage area (km^2)	12.500
Landcover: forests (%)	80.000
Landcover: agriculture (%)	0.
Landcover: pastures (%)	10.000
Basin length (km)	5.500
Altitude difference (m)	542.000
Main-stem channel length (km)	6.300
Total channel length (km)	31.000
Initial conditions	
Initial snow (rain equivalent) (mm)	0.
Initial interception storage (0/1)	1
Initial soil moisture (relative)	1.000
Initial base flow ($m^3\ s^{-1}$)	0.300

	Range sampled:	
Parameters	Minimum	Maximum
1 Altitude correction for precipitation	0.050	0.110
2 Altitude correction for temperature	0.200	1.500
3 Field capacity (mm m^{-1})	150.000	250.000
4 Percolation rate (mm day^{-1})	3.000	60.000
5 Speed of composed runoff (m day^{-1})	100.000	250.000
6 Groundwater response lag (day)	3.000	90.000

TABLE 11 Best parameter set (from 30,000 Monte Carlo runs) for two different performance criteria: ssq, sum of squares; sv, number of significant violations.

	Kienbach 80/81		Weyreggerbach 80/81		Alexenauerbach 79/80	
Parameter	ssq	sv	ssq	sv	ssq	sv
1	0.08	0.10	0.17	0.21	0.08	0.12
2	0.2	0.6	0.2	0.2	0.9	1.2
3	192.	160.	230.	224.	161.	224.
4	6.4	3.7	12.6	52.1	12.6	52.1
5	100.	206.	118.	19.1	118.	355.
6	76.	4.	51.	12.	51.	12.

The explanation is trivial: since basically all of the state variables of the model are unconstrained as such, and the only constraint is put on the output "runoff" (which is a kind of weighted sum of the states that represent the water storages in the system), there is more than one cluster of "solutions" to the parameter estimation problem (Figure 18). For each performance criterion, all the error is pushed into the unconstrained parts of the models (by adjusting the relative role of the underground storage via percolation and base-flow response). If the estimation scheme optimizes "significant events," i.e. storm runoff events, the groundwater response has to be quick. If the

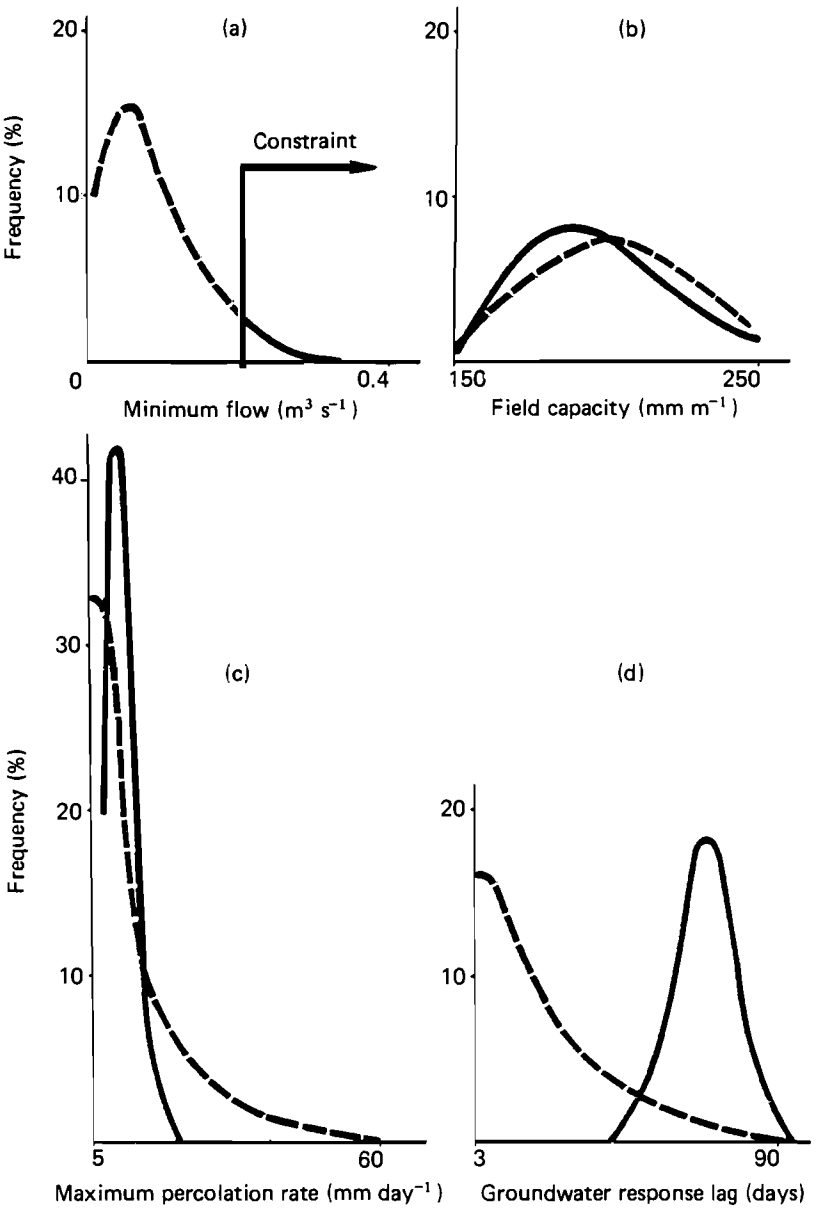

FIGURE 18 Frequency distribution of model response variable "minimum flow" and position of low-flow constraint; segregation of parameter distributions for constrained (thick lines) and unconstrained (broken lines) response classes.

overall performance is optimized, low-flow events play a more important role, and the groundwater response has to be damped in order to sustain low flows at a sufficiently high level.

Since the observed lower limit in the flow record looks very artificial (see Figure 17; in fact, it is more likely determined by the method of observation, namely a staff gauge, than by the "real" low flow), the problem is difficult to resolve without additional constraints, based on additional information. One possibility without having to resort to observations of the groundwater directly – which are just not available in the given case – is to exploit the low-flow conditions. Even if the records are unreliable, there is a certain minimum flow in the creek for a normal year. Clearly, this minimum flow is attributable, in terms of the model, to the base-flow contribution only. Therefore, by constraining the allowable minimum flow to, say, $0.2\,m^3s^{-1}$, the groundwater response will be constrained. Figure 18 contrasts the resulting parameter distributions with the "unconstrained" versions described above. Runs with six or less "violations" for storm runoff conditions were selected for the comparison. The two parameters (percolation rate and groundwater response lag) that largely determine the base-flow contribution show quite different frequency distributions for the two sets of estimation runs. With the minimum flow constrained, distributions are narrower – that is, the parameters are more easily identifiable. Also, whereas parameters 4 and 6 (Table 10) are found to show significant negative correlation in the first case, there is no more significant correlation after the introduction of the low-flow constraint.

The basic lesson from this example is simple enough: unconstrained state variables will just be acting as a residual term, where all the error can be dumped. Model response will be satisfactory for the constrained variables or the constrained part of the response, but at the expense of rather unrealistic behavior of the unconstrained variables and behavioral features. The calibration exercise degenerates into mere curve fitting. Parameters may turn out to be highly correlated (and thus at least one of a correlated pair is obsolete) and practically unidentifiable. Consequently, for a meaningful calibration, constraining all the state variables at least to a certain degree, and eventually in an indirect way (as demonstrated above), is essential. Unconstrained state variables will just increase the degree of freedom in "curve fitting," but they have no justifiable role whatsoever in a model.

4 DISCUSSION

"But as for certain truth, no man has known it ...
For all is but a woven web of guesses."

Xenophanes
(Diels-Kranz, Fragmente der Vorsokratiker, B 34)

"A nice adaptation of conditions will make almost any hypothesis agree with the phenomena."

(Black 1803)

4.1 A Philosophical View: Pragmatic Instrumentalism

It now behooves us to establish, with respect to the problem of uncertainty, a viewpoint, a perspective, a method of approach, that has hitherto received its principal development and application outside the boundaries of environmental modeling. Such prior development and application, however extraneous to our chief line of interest here, may very well be in a position to profit by the precedents established in methods, in conclusions, and, most particularly, in habit of thought*.

Uncertainty may have two basically different sources: mere "ignorance" in one of its numerous manifestations, or true indeterminism of the system under observation. The sources of uncertainty in environmental sciences and modeling are certainly of both kinds.

Macroscopic forms of elementary indeterminism are quite obvious, from the genetic variability within a species to the vagaries of the weather. Ignorance is usually blamed on the lack of sufficient and adequate data. This chronic lack of sufficient and adequate data, however, can only to a limited degree be blamed on the logistics of data collection or experimentation. It seems worth while to examine, therefore, the relationship between data and modeling. This relationship obviously suffers from a two-sided shortcoming, in that, on one hand, measurements are usually precedent and independent of the modeling efforts (and therefore more often than not turn out to be inadequate), and on the other, models are rarely formulated in terms of the measurements available.

Pure operationalism would hold the doctrine that theoretical constructs, i.e. the models, have to be formulated in terms of "measuring operations." (An elaborate discussion of all the "-isms" involved can be found, for example, in Popper (1953), and some critical discussion of Popper's ideas in Feyerabend (1975) and Lakatos (1978).) In other words, each of the elements described in a model would have to be directly measurable or experimentally accessible. However, it is obvious that measurements presuppose theories, or models for that matter. New theories may therefore well clash with the evidence, collected within the framework of an older theory. Lakatos (1978) put it in this way: "the contemporary observational theories, in the light of which the truth values of the basic statements of the (model) theory have to be established (i.e. by comparison), are false." In a similar line of argument, Feyerabend (1975) notes that "...observational reports, experimental results, 'factual' statements, either contain theoretical assumptions or assert them by the manner in which they are used...as a result, a theory may clash with the evidence not because it is not correct, but because the evidence is contaminated" – or, as I would say, perhaps even incommensurable.

The theory, or model, will in turn have to be responsive to the *problems*. Science, and most obviously applied science, starts from the problems, not from the observations or measurements. Yet observations may give rise to

*The well read will recognize these sentences, which I could not resist adapting from p.41 of the 1956 Dover edition of A.J. Lotka's *Elements of Mathematical (Physical) Biology*.

problems, especially if they were *unexpected*, contradictory to our expectations, that is, previous theories. The (scientific) problem is then solved through the construction of a theory that explains the unexpected and hitherto unexplained. And "... every worthwhile new theory raises new problems, ... problems of how to conduct new and previously unthought-of observational tests" (Popper 1972). As long as the model is testable in principle, its predictions (in terms of the corresponding measurements to be carried out, if not invented) could as well be taken as a challenge for field research and the experimenter. In fact, although stimulated by quite different motives, new measurement and observational techniques (in particular, by-products of space technology: satellite reconnaissance, aerial photography, radar-based weather observation, etc.; see, e.g., Lillesand and Kiefer 1979, Salomonson and Bhavsar 1980, Deutsch *et al.* 1981, and the remarks in Klemes 1983) are beginning to be recognized as an invaluable testing ground for hitherto next-to-untestable models.

Still, however, most of the measurements available are uncertain, or, at best, "statistical in nature," that is, based on samples. Uncertainty of a measure, however, also considerably depends on the yardstick used for the measurement. "All clocks are clouds, to some considerable degree – even the most precise of clocks" (Popper 1979), if one just takes a close enough view, that is, employs a sufficiently microscopic yardstick. The yardstick, in turn, is determined by the theoretical basis and conceptualizations used, and, from a more pragmatic point of view, by the problem (which, in turn, will influence the conceptualizations of the system, constrained by theory). Dealing with environmental systems we ought to use a "macroscope" for the yardstick (e.g. H.T. Odum 1971, de Rosnay 1975). Simulation models, in fact, are usually based on macroscopic conceptualizations. The basic problem, as discussed above, is in their reliance on data usually collected on a microscopic scale as a testing ground.

Scale is a key issue related to uncertainty. It has a bearing on the length of observation – a beautiful and apt example from statistical mechanics is given by Lotka (1924): "... for a long stretch of time the wholly determinate periodicity (with a period of 7,385 years, however) of the motion of the system of (26) pendulums (with periods of 0.5, 0.6, ..., 3.0 seconds, started simultaneously) is very effectively masked under the aspects of 'chance'." The importance of sampling frequency is highlighted, for example, by Kelley (1976) and Mejer and Jørgensen (1983). The role of scale for conceptualization of processes, which will influence sampling rather directly, was only recently addressed by Klemes (1983).

Simulation models, viewed as universal statements, or theories, transform a set of singular statements (the initial conditions) into another set of singular statements (the predictions). Since the first are uncertain, the latter have to be uncertain too. This, in light of an objective theory of truth as correspondence to the facts (what are the facts, then?), is certainly annoying as it very much affects the testability of these theories or models and, with it, their credibility. However, by humbly taking recourse to pragmatic instrumentalism, we may claim that our models are mere instruments for prediction. Models are sets of instructions to derive predictions, they are

technological computation rules and, in fact, algorithms and computer programs. We do not use them in the search for objective truth, but rather to make sufficiently useful predictions. "A theory is a tool we test by applying it, and which we judge as to its fitness by the results of its applications." This "Darwinian" instrumentalism, cited from Popper (1959, p.108), is criticized by Popper himself only a few pages later. It is, however, in keeping with the best tradition of the Vienna Circle of Mach, Wittgenstein, and Schlick.

This, however, will lead to some more problems for the credibility and applicability of models, and a slightly different interpretation of the testing process as compared with "pure" scientific theories. Theories are tested by attempts to refute them. For models as instruments, we can usually always find a "test to destruction" (see the application example in Section 3.1). In terms of the above approach, we will almost always be able, for any even only moderately complex nonlinear simulation model, to find an "allowable" input combination that results in unacceptable model response (at least from a very critical purist's point of view). Consequently, from the point of "pure" hypothesis testing, the model as a theory ought to be rejected. We still use it – within the limits of its applicability. What the severe and critical test can establish, under all the uncertainty associated with inputs and test conditions, is the range of applicability of the model. This range is an essential property of each model, and therefore it has to be well explored. Find the boundaries within which your model behaves properly (Meadows 1979), which is easy enough by means of numerical experimentation on the computer, e.g. Monte-Carlo-based trial and error. It should be made clear that Popper's falsificationism (from naive to sophisticated, as labeled and criticized by Lakatos (1978, p.93ff.)) can only strictly be applied to the lowest operational level of theoretical constructs, i.e. the individually testable hypothesis. In environmental applications, which, as a rule, are at the intersections of ecology, technology, and socioeconomic and political problems, models are of a more complex, composed nature. They may be closer to Lakatos's research programs than to Popper's theories. And, in fact, they may be best approached in terms of Feyerabend's (1975) creative anarchism of "anything goes."

Since it is obvious, fully intentional, and also inevitable that all models, and numerical simulation models in particular (and even such supposedly elementary precise and well established models as Schrödinger's equation describing the hydrogen atom), are (pragmatic) simplifications, and thus include uncertainty, we have to be aware of the implications and consequences *in relation to the problem to be solved.*

4.2 Uncertainty Analysis: Alternative Approaches

Uncertainty inherent in environmental modeling is inevitable – stochastic variability, heterogeneity, rich behavioral repertoires, and time-varying structural and functional attributes are all basic features of environmental systems. Thus, it seems unlikely that any moderately complex environmental system can be well defined in the traditional physicochemical

sense (Hornberger and Spear 1981). In fact, environmental systems have been described as being "poorly" or "badly defined" (Young 1978, 1983).

For a considerable time, this uncertainty and its inevitable consequences have been ignored altogether, resulting in a most misleading pseudo-precision in the results – and consequently overly optimistic and unrealistic expectations, failures in applications or rather, to be applied, disappointments, and finally a bad reputation for modelers in the scientific community (e.g. Biswas 1975, Watt 1977, Hedgpeth 1977, Hilborn 1979, Fedra 1980).

Uncertain systems require a different approach, profoundly different from the regular, orderly, and highly predictable "clocks" – they are "clouds," highly irregular, disorderly, and more or less unpredictable (Popper 1979). Rather than treat the variability of ecosystems as an annoying smokescreen that obscures their "true" behavior, we may treat this variability as a basic characteristic of such systems (Silvert 1983a, b).

As one consequence of all the uncertainty in environmental modeling, model development, parameter estimation, and "prediction" have to be understood as inseparably linked parts of one and the same process, i.e. modeling (which, as a matter of fact, is just one formalized way of doing scientific reasearch). The approach described above, linking a formal test of model structure, parameter estimation, and estimation of prediction uncertainty, is one approach to rational modeling under uncertainty.

Alternative approaches involve the direct *a priori* use of the probability density functions (Silvert 1983a,b), which are arrived at in the Monte Carlo approach *a posteriori*. Although this alternative approach is much more elegant in its direct way of treating variability or uncertainty, there are a few restrictions. The method requires the representation of the elements or features of a system in terms of (sampling) distributions, their mean values, and their moments. However, in practice, the "sample" may consist of only one measurement, making the estimation of the moments more or less impossible. And although many natural distributions are found to be skewed to a considerable degree, ease of mathematical treatment may lead to the assumption of normality or log-normality of the variables describing an ecosystem. Nevertheless, practical implementation of this type of approach requires a fair amount of mathematical sophistication and a willingness to be quite ruthless about model simplification (Silvert 1983b). Although "probabilistic model structures" make it possible to carry out stochastic modeling without extensive Monte Carlo simulations, the covariance calculations add a substantial computational burden. For a certain class of problems, where sufficient data are available to define probability density distributions, and where a structurally simple model will suffice, as for example in population dynamics, the approach adopted by Silvert (1983a,b) is certainly a complementary alternative.

Another set of alternatives, or rather complementary methods and approaches, can be subsumed under the term error analysis or uncertainty analysis (e.g. Reckhow 1979, 1981, Di Toro and van Straten 1979, O'Neill and Gardner 1979, O'Neill and Rust 1979, Gardner *et al.* 1980, 1981, Scavia *et al.* 1981, Gardner and O'Neill 1982). These methods use maximum-likelihood

techniques and first-order variance propagation to estimate overall model variance (or uncertainty) originating from uncertain initial conditions, parameters, or driving variables. They also require a fair degree of mathematical and statistical sophistication, and may involve considerable computational burden. And most important of all, they require several assumptions to be made about the model as well as about the set of data used for comparison. First-order error propagation employs a first-order linearization of the model, that is, the original nonlinear model is linearized and replaced by its first-order Taylor series approximation. This may eventually turn out to be inadequate. Since the second-order propagation equation involves second partial derivative matrices, "which are somewhat cumbersome to handle" (Scavia *et al.* 1981), a significant simplification of the model might have to be made to make the computational burden feasible. Also, variance around a mean behavioral value can eventually be a rather misleading measure of uncertainty, if the underlying population is strongly skewed (Section 3.2 and Scavia *et al.* 1981). In fact, Monte Carlo simulations – which, by comparison with the above-mentioned methods of error analysis, entail the complete nonlinear simulation model – may produce even bimodal distributions for certain state variables, indicating bifurcations (Section 3.3). Clearly, means and variance estimates are of little significance then.

Finally, for determining the deviation of state variables from the "true" values, usually by employing least squares, the error analysis has to make the assumption that either the observations are without error, i.e. they represent the "true" status of the system, or the error around them is of a known kind, usually Gaussian white noise, or the "true" values are represented by the deterministic model solution (Scavia *et al.* 1981). Clearly, in light of the above sections, these assumptions may seem unrealistic, and at best, are untested.

Recently, formalized parameter calibration routines have begun to be applied in the field of modeling complex aquatic ecosystems, for example by Lewis and Nir (1978), Jørgensen *et al.* (1978), Di Toro and van Straten (1979), and Benson (1979). In these methods a loss function is defined, usually in a squared-error form, and subsequently a parameter vector is sought that minimizes this loss function. This procedure thus avoids the analyst's subjective perception of which parameter ought to be adjusted to improve the fit. Also, the equally subjective judgment of agreement between simulation and observation is replaced by a more formal quantitative notion. However, although frequently called "objective function," this does not imply that the criterion chosen is free from subjective elements. For example, in problems with state variables with different physical dimensions, some (subjective) form of weighting is required in the formulation of a single-valued loss function. Furthermore, it is not easy to account for uncertainty in the field data, although methods to do this have been attempted (Beck and Young 1976, Lewis and Nir 1978, Jolankai and Szöllösi-Nagy 1978, Beck 1979, Di Toro and van Straten 1979).

Finally, however, it has to be recognized that the assumption that a single 'best' parameter vector exists is at least questionable, especially if data uncertainty is considered, and, in any case, experience shows that it is

extremely difficult to find such a unique vector if the number of parameters to be estimated is larger than, say, six to ten. If, however, such a best parameter vector exists – by definition – and can be identified by whatever method, its meaning and interpretation would still be problematic.

One way of comparing such approaches looking for a "best" solution with the methods described above is the following: if a "best" (by whatever set of criteria) parameter vector is sought, the "target" of the estimation procedure is a point, and the measure of success is some measure of the distance of the model response from this point. In the examples presented in this report, this point is extended to a region – acknowledging the uncertainty in the observations that define this point – and the measure of success is whether or not the model response is within this region. Instead of a continuous measure of distance, a discrete classification into "inside" and "outside" is used. Of course, any combination of the methods could be imagined and, in fact, the gradual shifting of the target region in the analysis process as described in Sections 3.1 and 3.2 is one such possible extension of the basic procedure.

As stated in the introduction, the Monte Carlo method is nothing more than computerized trial and error. As such, the method would be very inefficient for the calibration of complex simulation models and their repeated testing, since it is, principally, blind and unintelligent. This problem, however, can be overcome by a more structured design of the estimation scheme, with iterative cycles of estimation and analysis. The main justification for the use of Monte Carlo methods, however, is in their conceptual simplicity. This simplicity, and the resulting flexibility, allows for the accommodation of uncertainty, and at the same time it permits a very problem-specific exploitation of all the available information.

The method requires the formal definition of an acceptable model response *a priori*. In this definition, arbitrary classifications and subjective judgments cannot always be avoided. Although based on the available field data, the definition has to be formulated on the level of abstraction of the model. This involves subjective interpretation of the raw data, and consequently introduces some further uncertainty. This uncertainty is a problem common to any modeling approach. However, this inevitable subjective element has to be made explicit, open to criticism, and ready for easy revision on the basis of further experience (Figure 1). On the other hand, the approach allows for the easy inclusion of all kinds of additional information, not usually included in a "data set," such as some time series of observations on state variables. Much of the information available on environmental systems, however, is of this more general, semiquantitative type, resulting from many qualitative observations rather than quantitative measurements. Nevertheless, this information is most valuable, as the specific data available are usually scarce, scattered, error-corrupted, and typically on the wrong items.

Any model response generated can be classified as either "acceptable" or "not acceptable." The classification is discrete, and once the constraint conditions are formulated there is no more ambiguity, no gradual or partial agreement or disagreement between the model response and the observations, calling for arbitrary judgments. How small would the sum of squared errors have

to be for a given state variable to make a model acceptable? Although a least-squares criterion may be helpful in finding a "best" parameter set (according to the least-squares criterion with its implicit bias and problems) for a given model structure, it does not allow one to conclude whether or not the model structure is adequate. Subjective judgment *a posteriori* has to be used. Examples abound where only partial agreement of model output and observations is described as "acceptable or of reasonably good fit," ignoring the fact that severe discrepancies exist between parts of the model response (e.g. for some of the state variables) and the observations (Reckhow 1981). This is most obvious in the case of the introduction of unmeasured (and consequently unconstrained) state variables into a model (compare the application example in Section 3.3) – bacteria are an almost classical example in water quality modeling.

4.3 Consequences and Implications: Uncertainty and Forecasting

Uncertainty in ecological modeling is certainly an inevitable element in the method as well as in the object of study, which is most obvious when one tries to predict the future on the basis of a fuzzy present. The analysis of model uncertainty together with appropriate methods for model calibration under uncertainty, and of its consequences, i.e. its "inverse," prediction accuracy, is certainly at an early stage of development. However, being aware of model and especially prediction uncertainty and the thus obvious limits of predictability, i.e. the range within which a given model may reasonably be applied, might well help to avoid too naive a trust in numerical models. Analysis of the various sources of model uncertainty and their relations and interdependences will be necessary to improve model applicability. And the least impact from model error analysis on model application should be a critical reevaluation of the questions that can reasonably be addressed and answered by means of numerical models.

The implications of uncertainty are many: there are implications for the testability of hypotheses, which, in terms of simulation modeling, is primarily on model development. This may cast new light on the principle of parsimony adopted in Section 3.1. Citing Popper (1972) again, "...it can be shown that what is usually called the simplicity of a theory is associated with its logical improbability, and not with its probability, as has often been supposed. This indeed, allows us to deduce ... why it is always advantageous to try the simplest theories first. They are those which offer us the best chance to submit them to severe tests: the simpler theory has always a higher degree of testability than the more complicated one."

In addition, there are consequences for prediction accuracy, which largely influences model interpretation and, consequently, applications. However, since the uncertainty is a basic characteristic of the systems dealt with, we have to live with it, and exploit it wherever possible (Holling 1978). One possibility, as demonstrated above, is to estimate over which time span and over which range of conditions useful predictions – in terms of the problem to be solved – can be made. A major result of the prediction is thus in the

determination of its reliability and applicability. Clearly, this calls for an appropriate set of methods in planning, decision making, and management, where the uncertainty inherent in model-based forecasts needs to be fully acknowledged. Simulation models rarely attempt to predict the "future" in an absolute sense. They are designed to address questions of the "what if?" kind, which are explicitly based on (additional) *assumptions* or more or less speculative scenarios about the future. Complex environmental models should probably be understood as educational tools rather than engineering tools: they do not provide solutions to be readily implemented, but rather clues as to how a system might evolve if certain actions are taken, which should help to shape policies and assist decisions.

After all, much of the uncertainty associated with large-scale modeling, and environmental modeling in particular, is a necessary and direct consequence of the same causes that create the need for these models: these models are built exactly because the systems modeled are no longer directly experimentally accessible; and this for good reasons.

ACKNOWLEDGMENT

This research was supported by the Austrian Research Foundation (grant no. 3905).

REFERENCES

Barnes, R.S.K., and K.H. Mann (eds.) (1980) Fundamentals of Aquatic Ecosystems (Oxford: Blackwell).

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

Beck, M.B., and P.C. Young (1976) Systematic identification of DO–BOD model structure. Proceedings of American Society of Civil Engineers, Journal of Environmental Engineering Division 102:909–927.

Benson, M. (1979) Parameter fitting in dynamic models. Ecological Modelling 6:97–115.

Biswas, A.K. (1975) Mathematical modelling and environmental decision making. Ecological Modelling 1:31–48.

Black, J. (1803) Lectures on the Elements of Chemistry vol. I. Edinburgh.

Chen, C.W., and G.T. Orlob (1975) Ecological simulation for aquatic environments. In: B.C. Patten (ed), Systems Analysis and Simulation in Ecology vol. III (New York, NY: Academic Press), pp. 475–588.

Chorley, R.J. (1978) The hillslope hydrological cycle. In: M.J. Kirkby (ed.), Hillslope Hydrology (Chichester: Wiley), pp. 1–42.

Cushing, D.H., and J.J. Walsh (eds.) (1976) The Ecology of the Seas (Oxford: Blackwell).

Deutsch, M., D.R. Wiesnet, and A. Rango (eds.) (1981) Satellite Hydrology (Minneapolis, MN: American Water Resources Association).

Di Toro, D.M., D.J. O'Connor, and R.V. Thomann (1971) A dynamic model of the phytoplankton population in the Sacramento–San Joaquin Delta. American Chemical Society, Advances in Chemistry Series 106:131–180.

Di Toro, D.M., and G. van Straten (1979) Uncertainty in the parameters and predictions of phytoplankton models. Working Paper WP-79-27 (Laxenburg, Austria: International Institute for Applied Systems Analysis).

Dubois, D. (1976) Modeling and simulation of the mesoscale mosaic structure of the lower marine trophic levels. In: Lecture Notes in Computer Sciences vol.40 (Berlin: Springer), pp. 407–418.

Dubois, D., and P.L. Closset (1976) Patchiness in primary and secondary production in the Southern Bight: a mathematical theory. In: G. Persoone and E. Jaspers (eds.), Proceedings of 10th European Symposium on Marine Biology vol. 2 (Wetteren, Belgium: University Press), pp. 211–229.

Eigen, M., and R. Winkler (1975) Das Spiel. Naturgesetze Steuern den Zufall (Munich: Piper).

Fedra, K. (1980) Mathematical modelling – a management tool for aquatic ecosystems? Helgoländer Meeresuntersuchungen 34:221–235 (also reprinted as IIASA Research Report RR-81-2).

Fedra, K. (1981a) Estimating model prediction accuracy: a stochastic approach to ecosystem modelling. In: D.M. Dubois (ed.), Progress in Ecological Engineering and Management by Mathematical Modelling (Liège: Cebedoc).

Fedra, K. (1981b) Pelagic foodweb analysis: hypothesis testing by simulation. Kieler Meeresforschungen 5:240–258.

Fedra, K. (1981c) Hypothesis testing by simulation: an environmental example. Working Paper WP-81-74 (Laxenburg, Austria: International Institute for Applied Systems Analysis).

Fedra, K. (1983) A Monte Carlo approach to estimation and prediction. In: M.B. Beck and G. van Straten (eds.), Uncertainty and Forecasting of Water Quality (Berlin: Springer), pp. 259–291.

Fedra, K., G. van Straten, and M.B. Beck (1981) Uncertainty and arbitrariness in ecosystems modelling: a lake modelling example. Ecological Modelling 13:87–110 (also reprinted as IIASA Research Report RR-81-26).

Feyerabend, P.K. (1975) Against Method. Outline of an Anarchistic Theory of Knowledge (London: Verso).

Finch, P.D. (1981) On the role of description in statistical inquiry. British Journal for the Philosophy of Science 32(2):127–144.

Flögl, H. (1974) Die Reinhaltung der Salzkammergutseen. Österreichisch Wasserwirtschaft 26(1/2):5–11.

Flögl, H. (1976a) Die dreistufige Kläranlage Mondsee. Österreichische Wasserwirtschaft 28(1/2):17–24.

Flögl, H. (1976b) Die Ringkanalisation am Attersee. Österreichische Wasserwirtschaft 28(9/10):186–193.

Gardner, R.H., R.V. O'Neill, J.B. Mankin, and J.H. Carney (1980) Comparative error analysis of six predator–prey models. Ecology 61:323–332.

Gardner, R.H., R.V. O'Neill, J.B. Mankin, and J.H. Carney (1981) A comparison of sensitivity analysis and error analysis based on a stream ecosystem model. Ecological Modelling 12:173–190.

Goldberg, E.D., I.N. McCave, J.J. O'Brien, and J.M. Steele (eds.) (1977) Marine Modeling (New York, NY: Wiley–Interscience).

Greve, W. (1981) Invertebrate predator control in a coastal marine ecosystem: the significance of *Beroe gracilis* (Ctenophora). Proceedings of 15th European Marine Biology Symposium. Kieler Meeresforschungen, Sonderheft 5:211–217.

Greve, W., and F. Reiners (1981) The impact of prey–predator waves from estuaries on the planktonic marine ecosystem. In: V.S. Kennedy (ed.), Estuarine Perspectives (London: Academic Press).

Haeckel, E. (1870) Über Entwicklungsgang und Aufgabe der Zoologie. Jenaer Zeitschrift für Naturwissenschaft 5:353–370.

Hedgpeth, J.W. (1977) Models and muddles. Some philosophical observations. Helgoländer Meeresuntersuchungen 30:92–104.

Hilborn, R. (1979) Some failures and successes in applying systems analysis to ecological systems. Journal of Applied Systems Analysis 6:25–31.

Holling, C.S. (ed.) (1978) Adaptive Environmental Assessment and Management. IIASA International Series on Applied Systems Analysis vol. 3 (Chichester: Wiley).

Hornberger, G.M., and R.C. Spear (1980) Eutrophication in Peel Inlet – I. Problem-defining behavior and a mathematical model for the phosphorus scenario. Water Research 14:29–42.

Hornberger, G.M., and R.C. Spear (1981) An approach to the preliminary analysis of environmental systems. Journal of Environmental Management 12(1):7–18.

Imboden, D., and R. Gächter (1978) A dynamic lake model for trophic state prediction. Ecological Modelling 4:77–98

Jolankai, G., and A. Szöllösy-Nagy (1978) A simple eutrophication model for the bay of Keszthely, Lake Balaton. In: Modeling the Water Quality of the Hydrological Cycle, Proceedings of Baden Symposium, September, International Association of Hydrological Sciences Publication 125, pp. 137–150.

Jørgensen, S.E. (1980a) Lake Management. Water Development, Supply and Management vol. 14 (Oxford: Pergamon).

Jørgensen, S.E. (1980b) The application of intensive measurement for the calibration of eutrophication models. Journal of Ecological Modelling 8.

Jørgensen, S.E., H. Mejer, and M. Friis (1978) Examination of a lake model. Ecological Modelling 4:253–278.

Kelley, J.C. (1976) Sampling the Sea. In: D.H. Cushing and J.J. Walsh (eds.), The Ecology of the Seas (Oxford: Blackwell), pp. 361–387.

Klemes, V. (1983) Conceptualization and scale in hydrology. International Symposium on the Understanding of Processes on Basin Scale, Universidad Simon Bolivar, Caracas, Venezuela, January 11–14. To appear in Journal of Hydrology.

Kremer, J.N., and S.W. Nixon (1978) A Coastal Marine Ecosystem. Ecological Studies vol. 24 (New York, NY: Springer).

Lakatos, I. (1978) The methodology of scientific research programmes. Philosophical Papers vol. 1 (Cambridge: Cambridge University Press).

Lassen, H., and P.B. Nielsen (1972) Simple mathematical model for the primary production as a function of the phosphate concentration and the incoming solar energy applied to the North Sea. International Council for the Exploration of the Sea, Plankton Committee Report CM 1972/L:6.

Lewis, S., and A. Nir (1978) A study of parameter estimation procedures of a model of lake phosphorus dynamics. Ecological Modelling 4:99–118.

Lillesand, T.M., and R.W. Kiefer (1979) Remote Sensing and Image Interpretation (New York, NY: Wiley).

Lotka, J.A. (1924) Two models in statistical mechanics. American Mathematical Monthly 31:121–125.

Meadows, D.H. (1979) Social system modeling: the state of the science. Lecture presented at the International Institute for Applied Systems Analysis, Laxenburg, Austria, June 1979 (adapted from article in The Electronic Oracle, ed. D.H. Meadows (New York, NY: Wiley), to be published).

Mejer, H., and L. Jørgensen (1983) Model identification methods applied to two Danish Lakes. In: M.B. Beck and G. van Straten (eds.), Uncertainty and Forecasting of Water Quality (Berlin: Springer), pp. 173–182.

Monod, J. (1970) Le Hasard et la Nécessité (Paris: Editions du Seuil).

Morris, I. (ed.) (1980) The Physiological Ecology of Phytoplankton, Studies in Ecology vol.7 (Oxford: Blackwell).

Müller, G. (ed.) (1979) Attersee: Jahresbericht 1978. Arbeiten aus dem Labor, Weyregg, Report 3/1979.

Nihoul, J.C.J. (1975) Modeling of Marine Systems. Elsevier Oceanography Series vol. 10 (Amsterdam: Elsevier).

Odum, E.P. (1971) Fundamentals of Ecology (Philadelphia, PA: Saunders).

Odum, H.T. (1971) Environment, Power, and Society (New York, NY: Wiley).

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

O'Neill, R.V., and B. Rust (1979) Aggregation error in ecological models. Ecological Modelling 7:91–105.

Park, R., *et al.* (1974) A generalized model for simulating lake ecosystems. Simulation 23:33–56.

Parsons, T.R., M. Takahashi, and B. Hargrave (1977) Biological Oceanographic Processes (Oxford: Pergamon).

Pichot, G., and Y. Runfola (1974) Building of simulation models of ecosystems: nonlinear interaction parameter estimation. In: C.M. Ices (ed.), Mathematical Models of Continental Seas: Dynamic Processes in the Southern Bight, pp. 258–272.

Pichot, G,. and Y. Runfola (1975) Mathematical model of the nitrogen cycle in the Southern Bight of the North Sea. In: G. Persoone and E. Jaspers (eds.), Proceedings of 10th European Symposium on Marine Biology vol. 2 (Wetteren, Belgium: University Press), pp. 467–476.

Popper, K.R. (1953) Some problems in the philosophy of science. Appendix to Science: Conjectures and Refutations, lecture given in Cambridge, 1953. Reprinted in: Conjectures and Refutations, the Growth of Scientific Knowledge (London: Routledge and Keegan Paul, revised edn. 1972), pp. 59–65.

Popper, K.R. (1959) The Logic of Scientific Discovery (London: Hutchinson).

Popper, K.R. (1972) Truth, rationality, and the growth of scientific knowledge. In: Conjectures and Refutations, the Growth of Scientific Knowledge (London: Routledge and Keegan Paul, revised edn. 1972), pp. 215–250.

Popper, K.R. (1979) Of clouds and clocks. In: Objective Knowledge, an Evolutionary Approach (Oxford: Oxford University Press, revised edn.), pp. 206–255.

Radach, G. (1980) Preliminary simulations of the phytoplankton and phosphate dynamics during FLEX'76 with a simple two-component model. Meteorologische Forschungsergebnisse A, No. 22:151–163.

Radach, G., and E. Maier-Reimer (1975) The vertical structure of phytoplankton growth dynamics. A mathematical model. Mémoires de la Societé Royale de Sciences de Liège, Série 6, 7:113–146.

Reckhow, K.H. (1979) The use of a simple model and uncertainty analysis in lake management. Water Resources Bulletin 15:601–611.

Reckhow, K.H. (1981) Model verification: a philosophical and statistical analysis. Paper presented at Conference of American Water Resources Association, Atlanta, GA.

de Rosnay, J. (1975) The Macroscope: A New World Scientific System (New York, NY: Harper and Row).

Salomonson, V.V., and P.D. Bhavsar (1980) The Contribution of Space Observations to Water Resources Management. Advances in Space Exploration, COSPAR Symposium Series (Oxford: Pergamon).

Scavia, D., W.F. Powers, R.P. Canale, and J.L. Moody (1981) Comparison of first-order error analysis and Monte Carlo simulation in time-dependent lake eutrophication model. Water Resources Research 174:1051–1059.

Silvert, W. (1983a) The stochastic dynamics of ecosystems. Unpublished report.

Silvert, W. (1983b) Adaptive design of marine ecosystem research. Unpublished report.

Simons, T.J., and D.C.L. Lam (1980) Some limitations of water quality models for large lakes: a case study of Lake Ontario. Water Resources Research 16(1):105–116.

Steele, J.H. (1962) Environmental control of photosynthesis in the sea. Limnology and Oceanography 7:137–150.
Steele, J.H. (1974) The Structure of Marine Ecosystems (Cambridge, MA: Harvard University Press).
Steele, J.H. (1976) Patchiness. In: D.H. Cushing and J.J. Walsh (eds.), The Ecology of the Seas (Oxford: Blackwell), pp. 98–115.
Steele, J.H. (ed.) (1978) Spatial Patterns in Plankton Communities. NATO Conference Series 4: Marine Sciences vol. 3 (New York, NY: Plenum).
Steele, J.H., and B.W. Frost (1977) The structure of plankton communities. Philosophical Transactions of Royal Society, London, Series B, 280:485–534.
Steele, J.H., and M.M. Mullin (1977) Zooplankton dynamics. In: E.D. Goldberg *et al.* (eds.), The Sea vol. 6 (New York, NY: Wiley), pp. 857–890.
Vollenweider, R., and J.J. Kerekes (1980) OECD Cooperative Programme on Monitoring of Inland Waters (Eutrophication), Synthesis Report (draft), April.
Walsh, J.J. (1975) A spatial simulation model of the Peru upwelling ecosystem. Deep-Sea Research 22:201–236.
Watt, K.E.F. (1977) Why won't anyone believe us? Simulation 28:1–3.
Wheeler, J.A. (1956) A septet of Sibylls. Aid in the search for truth. American Scientist 4:360–377.
Winter, T.C. (1981) Uncertainties in estimating the water balance of lakes. Water Resources Bulletin 17(1):82–115.
Young, P.C. (1978) General theory of modelling for badly defined systems. In: G.C. Vansteenkiste (ed.), Modelling, Identification, and Control in Environmental Systems (Amsterdam: North-Holland), pp. 103–135.
Young, P.C (1983) The validity and credibility of models for badly defined systems. In: M.B. Beck and G. van Straten (eds.), Uncertainty and Forecasting of Water Quality (Berlin: Springer), pp. 69–98.

BIBLIOGRAPHY

This bibliography comprises references to reports on the application of Monte Carlo methods related to problems of uncertainty and error in environmental modeling.

Argentesi, F., and L. Olivi (1976) Statistical sensitivity analyses of a simulation model for the biomass–nutrient dynamics in aquatic ecosystems. In: Proceedings of 4th Summer Computer Simulation Conference, Simulation Councils, La Jolla, CA, pp. 389–393.
Beck, M.B. (1981) Hard or soft environmental systems? Ecological Modelling 11:233–251 (also reprinted as IIASA Research Report RR-81-4).
Beck, M.B., E. Halfon, and G. van Straten (1979) The propagation of errors and uncertainty in forecasting water quality. Working Paper WP-79-100 (Laxenburg, Austria: International Institute for Applied Systems Analysis).
Burges, S.J., and D.P. Lettenmaier (1975) Probabilistic methods in stream quality management. Water Resources Bulletin 11:115–130.
Carney, J.H., R.H. Gardner, J.B. Mankin, and R.V. O'Neill (1981) SEAP: A computer program for the error analysis of a stream model, Monte Carlo methodology and program documentation. Oak Ridge National Laboratory Technical Memorandum ORNL/TM-7654.
Clark, W.C., D.D. Jones, and C.S. Holling (1979) Lessons for ecological policy design: a case study of ecosystem management. Ecological Modelling 7:1–53.
Crawford, N.H., and R.K. Linsley (1966) Digital simulation in hydrology: Stanford Watershed Model IV. Department of Civil Engineering, Stanford University, Stanford, CA Technical Report 39.

Fedra, K. (1979a) Modeling biological processes in the aquatic environment, with special reference to adaptation. Working Paper WP-79-20 (Laxenburg, Austria: International Institute for Applied Systems Analysis).

Fedra, K. (1979b) A stochastic approach to model uncertainty: a lake modelling example. Working Paper WP-79-63 (Laxenburg, Austria: International Institute for Applied Systems Analysis).

Ford, D.E., K.W. Thornton, A.S. Lessen, and C. Stirgers (1978) Water quality evaluation of proposed Trexler Lake, Jordan Creek, Pennsylvania. US Corps of Engineers, Vicksburg, MS Report Y-78-10.

James, L.D. (1970) An evaluation of the relationship between streamflow patterns and watershed characteristics through the use of OPSET: A self-calibrating version of the Stanford Watershed model. University of Kentucky, Water Resources Institute, Lexington, KY Research Report 36.

McGrath, E.J., and D.C. Irving (1973) Techniques for efficient Monte Carlo simulation. Office of Naval Research, National Technical Information Service, Springfield, VA Report AD 762:721–723 (3 vols.).

Montgomery, R.H., V.D. Lee, and K.H. Reckhow (1980) A comparison of uncertainty analysis techniques: First-order analysis versus Monte Carlo simulation. Paper presented at Conference on Great Lakes Research, International Association for Great Lakes Research, Queens College, Kingston, Ontario.

O'Neill, R.V. (1971) A stochastic model of energy flow in predator compartment of an ecosystem. In: G.P. Patil, E.C. Pielou, and W.E. Walters (eds.), Many-Species Populations, Ecosystems, and Systems Analysis, Statistical Ecology vol. 3 (University Park, PA: Pennsylvania State University Press), pp. 107–121.

O'Neill, R.V. (1973) Error analysis of ecological models. In: D.J.Nelson (ed.), Radionuclides in Ecosystems (Springfield, VA: National Technical Information Service), pp. 898–908 (CONF-710501).

O'Neill, R.V., R.H. Gardner, and J.B. Mankin (1980) Analysis of parameter error in a nonlinear model. Ecological Modelling 8:297–312.

Scavia, D. (1979) The need for innovative verification of eutrophication models. In: J. Muller (ed.), Proceedings of National Workshop on Verification of Water Quality Models. US Environmental Protection Agency, Athens, GA.

Scavia, D. (1981) Uncertainty analysis of a lake eutrophication model. Ph.D. Dissertation, University of Michigan.

Spear, R.C., and G.M. Hornberger (1980) Eutrophication in Peel Inlet – II. Identification of critical uncertainties via generalized sensitivity analysis. Water Research 14:43–49.

van Straten, G. (1980) Analysis of model parameter uncertainty in simple phytoplankton models for Lake Balaton. Working Paper WP-80-139 (Laxenburg, Austria: International Institute for Applied Systems Analysis).

van Straten, G., and B. de Boer (1979) Sensitivity to uncertainty in a phytoplankton–oxygen model for lowland streams. Working Paper WP-79-28 (Laxenburg, Austria: International Institute for Applied Systems Analysis).

Tiwari, J.L., and J.E. Hobbie (1976) Random differential equations as models of ecosystems: Monte Carlo simulation approach. Mathematical Biosciences 28:25–44.

Tiwari, J.L., and J.E. Hobbie (1976) Random differential equations as models of ecosystems – II. Initial conditions and parameter specifications in terms of maximum entropy distributions. Mathematical Biosciences 31: 37–53.

Tiwari, J.L., J.E. Hobbie, J.P. Reed, D.W. Stanley, and M.C. Miller (1978) Some stochastic differential equation models for an aquatic ecosystem. Ecological Modelling 4:3–27.

Whitehead, P., and P. Young (1975) Water quality in river systems – Monte Carlo analysis and planning. Centre for Resource and Environmental Studies. Australian National University Report AS/R5.

Young, P.C., G.M. Hornberger, and R.C. Spear (1978) Modeling badly defined systems – some further thoughts. In: Proceedings of SIMSIG Simulation Conference, Australian National University, Canberra, pp. 24–32.

THE AUTHOR

Kurt Fedra was with IIASA between December 1977 and August 1982. He joined the Resources and Environment Area in August 1978 to work on the modeling of biological processes in the aquatic environment, mainly within the framework of a joint study with Austrian institutions on lake eutrophication control in Austria.

Dr. Fedra received his Ph.D. in zoology from the University of Vienna and was a research assistant at the Zoological Institute, working in international research programs on the ecology of the northern Adriatic Sea.

Dr. Fedra's interests are in environmental sciences, particularly systems ecology, and in computer-assisted methods of environmental planning and management. He is a recipient of the Theodor Körner prize and a member of the *Societas Internationalis Limnologiae*. During 1982–83, he was a Fellow at the Massachusetts Institute of Technology, Department of Civil Engineering.

RELATED IIASA PUBLICATIONS

Books

Askew, A.J., F. Greco, and J. Kindler (eds.) (1981) *Logistics and Benefits of Using Mathematical Models of Hydrologic and Water Resource Systems.* IIASA Proceedings Series vol. 13 (available from Pergamon Press, Oxford).

Beck, M.B., and G. van Straten (eds.) (1983) *Uncertainty and Forecasting of Water Quality* (available from Springer-Verlag, Berlin).

Orlob, G.T. (ed.) (1983) *Mathematical Modeling of Water Quality: Streams, Lakes, and Reservoirs.* International Series on Applied Systems Analysis vol. 12 (available from John Wiley and Sons, Chichester).

Wood, E.F., and A. Szöllösi-Nagy (eds.) (1980) *Real-Time Forecasting/Control of Water Resource Systems.* IIASA Proceedings Series vol. 8 (available from Pergamon Press, Oxford).

Research Reports

Beck, M.B. (1978) A comparative case study of dynamic models for DO–BOD–algae interaction in a freshwater river. RR-78-19 ($7.00).
Reprinted from *Theoretical Systems Ecology: Advances and Case Studies* (E. Halfon, Editor).

Beck, M.B. (1980) Model structure identification from experimental data. RR-80-4 (available for a handling charge of $1.00).
Reprinted from *Theoretical Systems Ecology: Advances and Case Studies* (E. Halfon, editor).

Beck, M.B. (1981) Hard or soft environmental systems? RR-81-4 (available for a handling charge of $1.00).
Reprinted from *Ecological Modelling* vol. 11 (1981).

Fedra, K. (1981) Mathematical modelling – a management tool for aquatic ecosystems? RR-81-2 (available for a handling charge of $1.00).
Reprinted from *Helgoländer Meeresuntersuchungen* vol. 34 (1980).

Fedra, K., G. van Straten, and M.B. Beck (1981) Uncertainty and arbitrariness in ecosystems modelling: a lake modelling example. RR-81-26 (available for a handling charge of $1.00).
Reprinted from *Ecological Modelling* vol. 13 (1981).

Levien, R.E., A.P. Wierzbicki, and B. Arthur (eds.) (1983) Mathematical modelling at IIASA. RR-83-7 (available for a handling charge of $1.00).
Reprinted from *Mathematical Modelling* vol. 3 (1982).

Lewandowski, A. (1982) Issues in model validation. RR-82-37 (available for a handling charge of $1.00).
Reprinted from *Angewandte Systemanalyse* vol. 3 (1982).

Rinaldi, S., and R. Soncini-Sessa (1977) Sensitivity analysis of Streeter–Phelps models. RR-77-1 ($4.00).

Wood, E.F. (1974) A Bayesian approach to analysing uncertainty among stochastic models. RR-74-16 ($4.00).

Wood, E.F. (1974) Analysis of uncertainty in deterministic rainfall runoff models. RR-74-18 ($5.00).