



International Institute for  
Applied Systems Analysis  
[www.iiasa.ac.at](http://www.iiasa.ac.at)

# Hypothesis Testing by Simulation: An Environmental Example

**Fedra, K.**

**IIASA Working Paper**

**WP-81-074**

**June 1981**



Fedra, K. (1981) Hypothesis Testing by Simulation: An Environmental Example. IIASA Working Paper. WP-81-074  
Copyright © 1981 by the author(s). <http://pure.iiasa.ac.at/1687/>

**Working Papers** on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work. All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage. All copies must bear this notice and the full citation on the first page. For other purposes, to republish, to post on servers or to redistribute to lists, permission must be sought by contacting [repository@iiasa.ac.at](mailto:repository@iiasa.ac.at)

NOT FOR QUOTATION  
WITHOUT PERMISSION  
OF THE AUTHOR

HYPOTHESIS TESTING BY SIMULATION:  
AN ENVIRONMENTAL EXAMPLE

Kurt Fedra

June 1981  
WP-81-74

*Working Papers* are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS  
A-2361 Laxenburg, Austria

## PREFACE

In recent years there has been considerable interest in developing models for environmental systems, much of it directed toward progressively larger and more complex simulation models. However, such a trend causes concern about several important issues. In particular, relatively little attention has been paid to the problems of errors and uncertainty in the field data, of inadequate amounts of field data, and of uncertainty about estimates of the model's parameters and about relations among the system's important variables.

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

This paper deals with the formalization of a procedure for hypothesis testing when the system under study is especially complex and observations of its behavior are subject to high levels of uncertainty. The procedure, based on the use of Monte Carlo simulation and extending previous work on model estimation (IIASA WP-80-87, for example), has several parallels with procedures for model structure identification using recursive estimation algorithms (see RR-80-4). The discussion of the paper therefore adds breadth and variety (of approach and method) to the development of a framework for modeling poorly defined environmental systems (see also RR-81-2, RR-81-4).

### ACKNOWLEDGEMENTS

The research described in this paper was supported by the research grant No.3905 of the Austrian "Fonds zur Förderung der wissenschaftlichen Forschung". The author gratefully acknowledges data and information made available by colleagues from the Biologische Anstalt Helgoland, and W.Greve's help in particular. I am indebted to A.John for her linguistic help.

A preliminary version of this paper was presented at the 15<sup>th</sup> European Symposium on Marine Biology, Kiel, October 1980.

## ABSTRACT

The study of environmental systems as ecological, physico-chemical as well as socio-economic 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 when 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 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, which explicitly considers the above constraints, is proposed.

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, temperature, and advection/diffusion, are described in systems models ranging in complexity from two compartments to many compartments and species groups. With each of the proposed models, yearly cycles of the systems behavior are simulated. A comparative analysis of the response of each of the models allows conclusions to be drawn on the adequacy of the alternative hypotheses. 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.

## TABLE OF CONTENTS

INTRODUCTION	1
Hypothesis Testing and Simulation Modeling	
THE ENVIRONMENTAL BACKGROUND	8
Describing the Environmental System	
HYPOTHESIS GENERATION AND TESTING	13
Designing alternative models	
Hypothesis No.1:	16
two compartments in a simple physical framework	
testing hypothesis No.1	17
Hypothesis No.2:	20
a four compartment web	
Hypothesis No.3:	25
one more trophic level	
Hypothesis No.4:	28
more biological detail	
a comparison with Steele's size-dependent formulations	
DISCUSSION	33
The Generalizable Lesson	
REFERENCES	38
APPENDIX	42
Model equations	

HYPOTHESIS TESTING BY SIMULATION:  
AN ENVIRONMENTAL EXAMPLE

Kurt Fedra

INTRODUCTION:

Hypothesis Testing and Simulation Modeling

Environmental systems are generally large, diverse, and complex. Variability in space and time and an extremely high number of interacting components which themselves are subject to changes in time, make their observation and understanding an extremely difficult, though challenging, scientific task of increasing socio-economic importance. This is especially true for marine systems, where in addition to all other problems the logistics of the research pose major difficulties.

Trophical relationships, i.e. the interrelations of eater and eaten, predator and prey, decomposers and substrate, play an essential role in environmental systems dynamics. Inseparably linked with the physical processes of, e.g., transport and mixing, energy flows, material cycles, or foodwebs are all conceptualizations of basic processes in environmental systems. In the aquatic environment, food webs link energy flows and mineral cycles in the most obvious way: primary production links the physical (energy) environment with the biological, organic elements of the environment, whereas the last link of decomposers connects the organic matter built and transformed, once again with the mineral cycles. At the same time, food webs describing



the production, transformation, distribution and finally, decomposition of organic material, are major elements in most problems of immediate practical interest to man, such as fisheries (c.f. Andersen and Ursin 1977), the bio-degradation and -accumulation of toxic substances (e.g. Thomann 1978), eutrophication, or problems of water quality in general.

The very high number of interactions between the numerous elements of ecological systems requires conceptual simplifications, aggregation, and abstraction of the systems under study, so as to make the theories one can formulate about the structural properties and the function of a system traceable. At the same time however, this introduces certain tradeoffs between precision or detail, and generality. Scientific research tries to establish theories, or general, universal statements, from which, in turn, singular statements or specific events can be deduced or predicted. A certain general applicability of a conceptualization is therefore an essential element of the scientific method itself. Only a purely descriptive approach requires a maximum of precise detail. In contrast, for an explanatory approach the principal features of any system have to be filtered from the site-specific ones as well as from random disturbance- -thus by necessity detail has to be ignored in order to make the general patterns visible.

Universal statements, describing those properties of a system which 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, that is to say, 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, cf. Fedra et al. 1980, Fedra, in press a) it must be possible to deduce or predict testable singular statements (observations or experimental results). Disagreement between the prediction deduced from the hypothesis or model and the available observations would then require rejection of the given hy-

pothesis, 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), 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 example given in Popper (1959), where he refers to a mechanical experiment (the breaking of a certain piece of thread), the initial conditions to be specified are simple enough: a weight and the characteristics of a thread (e.g. material, diameter etc.), which are, in relation to the expected outcome of the experiment, measurable without considerable error. This however, is no longer the case 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 results of two basic shortcomings, one in the measurement techniques available, another one 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.

The same line of argument can be followed with regard to the observation used for model-output comparison 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 examples however, we have to compare predictions with measurements from the system, which always include some measurement error, that is to say, they are ranges. Also, in environmental systems the degree of abstraction and aggregation is quite different for the measurements and for the model conceptualization. Therefore, the observations and measurements can only serve 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: it is possible, for the deterministic case referred to above, to describe the initial conditions or inputs as a set of numbers (forming a vector, determining a point in a  $n$ -dimensional input-hyperspace), and to do the same for the expected result of the experiment (the observed behavior of the system), resulting again in a point in a  $m$ -dimensional output-behavior-space. In the presence of uncertainty, i.e. considerable measurement and sampling errors, in part due to lumping and aggregation, the two points will have to be extended to regions in their respective hyperspaces. Instead of the two vectors we have to deal with classes of vectors with certain statistical properties and probability structures.

To test any specific hypothesis, we now examine if for a class of admissible initial conditions predictions (which are 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 class, is infinite, and only samples can be examined. Also, the class 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 class of admissible sets, however, is also characterized by the correlation structure, which determines the "shape" of the admissible input region in the hyperspace.

Figure 1 summarizes the approach outlined above in a flow chart, indicating the major steps in the procedure as well as their recursive relationships, and Figure 2 shows an example of a model response-space projection on a plane defined by two output-constraint variables, also indicating the defined empirical range, forming a rectangle in that plane.

This method of testing a given hypothesis does not indicate how to 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 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 to start with the simplest possible conceptualization, or the least complex model one can formulate. If this simple version fails to give an acceptable behavior over the allowable parameter ranges, complexity is increased by adding elements and more complex process descriptions to the model, until a satisfactory behavior can be achieved. However, there is in any case more than one way to increase a models complexity. 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, since I am considering conceptual models only (as opposed to statistical models, they are based on physical processes and include only terms directly interpretable in the real world system), additional observations can be exploited in many cases. Knowledge accumulated from the study of similar systems may also be helpful in increasing a given model's complexity.

Building up complexity and subjecting each version or level of the model to extensive tests, should allow one to learn about the way structural changes influence a models response. At the same time, the intricate connection of structure and inputs (the parameters) has to be emphasized, since the model's behavior is certainly responsive to both. Since 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 process seems preferable to a purely arbitrary and subjective procedure.

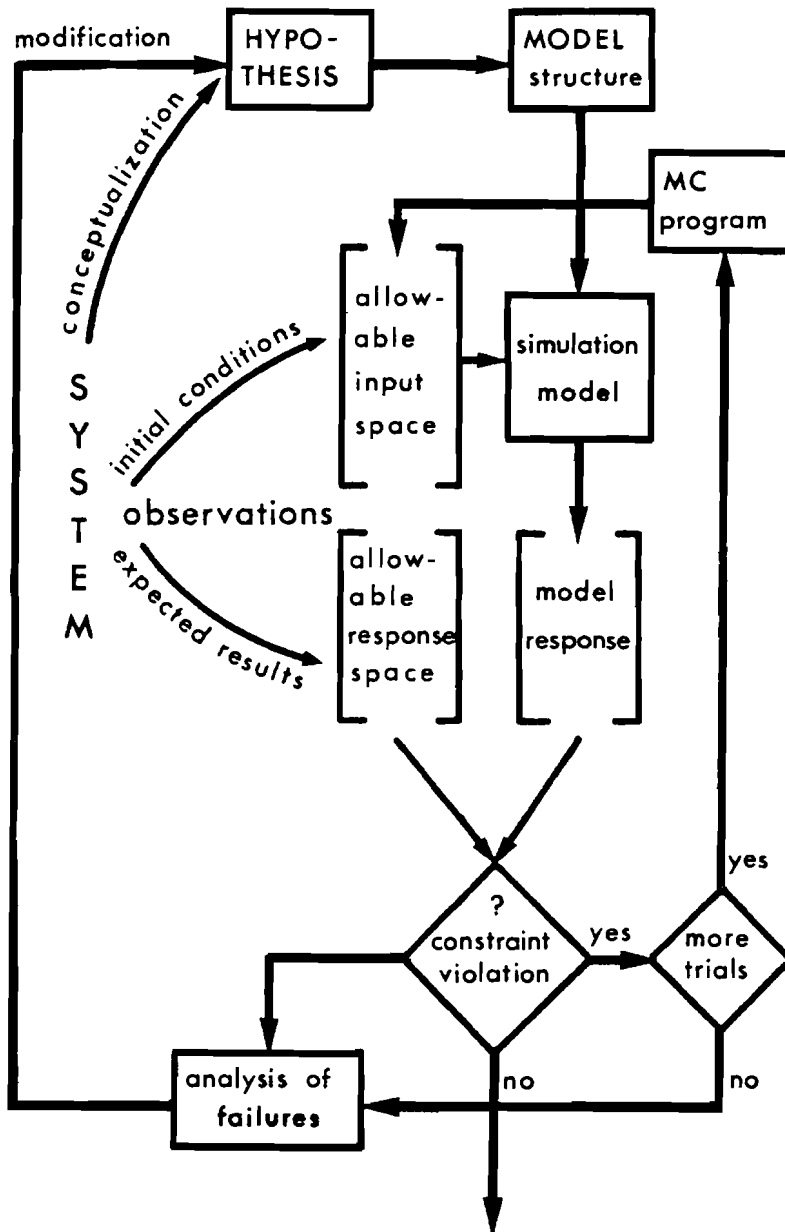
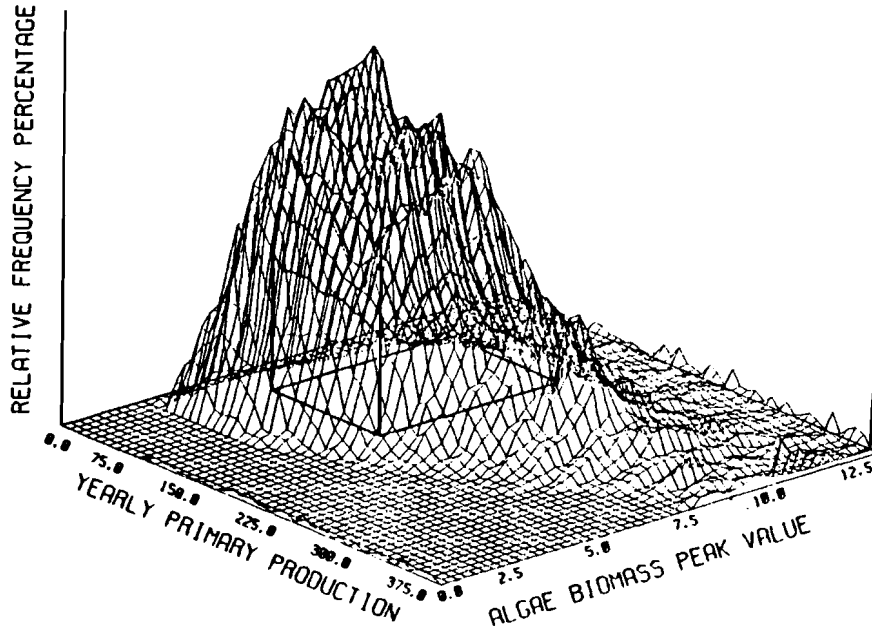


Figure 1: Flow chart of the approach

# MODEL RESPONSE-SPACE PROJECTION

ATTERSEE PHOSPHORUS MODEL: STANDARD INPUT RANGE



# INPUT DATA-SPACE PROJECTION

DATA-SPACE STRUCTURE FOR EMPIRICAL RESPONSE RANGE

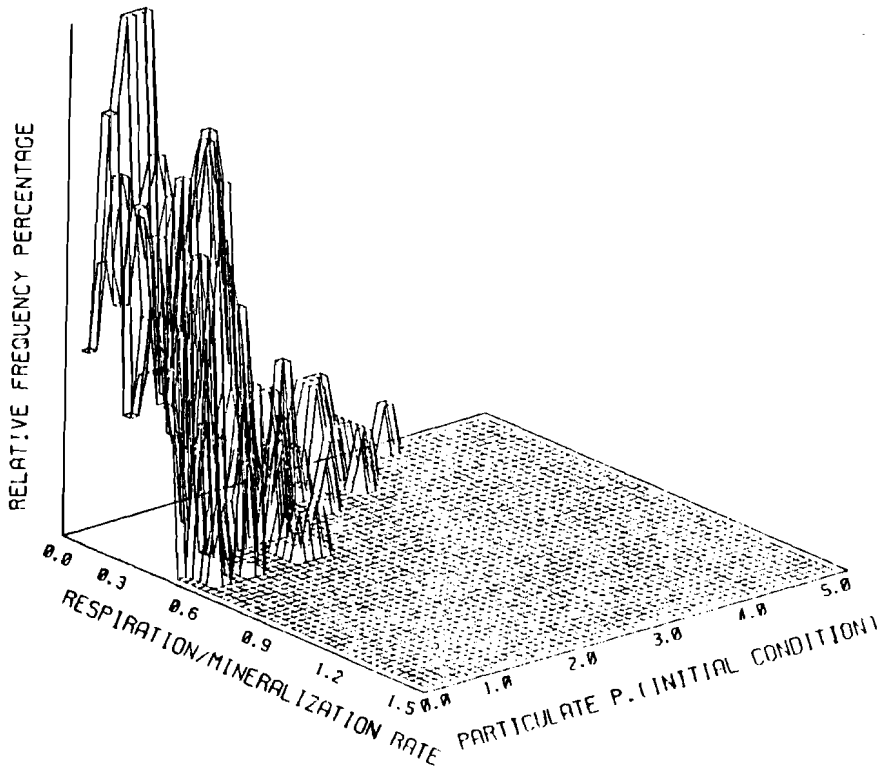


Figure 2: Model response-space projection on a plane of two constraint variables, indicating the defined empirical range of systems behavior (top) and projection from the 22-dimensional input-space region corresponding to the empirical behavior range (from Fedra, 1980).

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 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 an appropriate level of abstraction and aggregation. This description, which already has to be formulated in the 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.

As an example 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 physico-chemical as well as biological variables at the sampling station "Helgoland-Reede" for the period 1964 - 1979 (Hagmeier 1978, Lucht and Gillbricht 1978, Biologische Anstalt Helgoland, yearly reports 1964 - 1979, including unpublished data of Hagmeier, Hickel, Mangelsdorf, Treutner, Gassmann, Gillbricht). However, various other sources have been used for additional information (e.g. Steele, 1974, Nihoul, 1975) to compile a data set typical for an arbitrary location representative of the German Bight, Southern North Sea. For the sake of comparison, and to demonstrate how general some of the observed features are, data from several other parts of the North Sea are included in Figure 9.

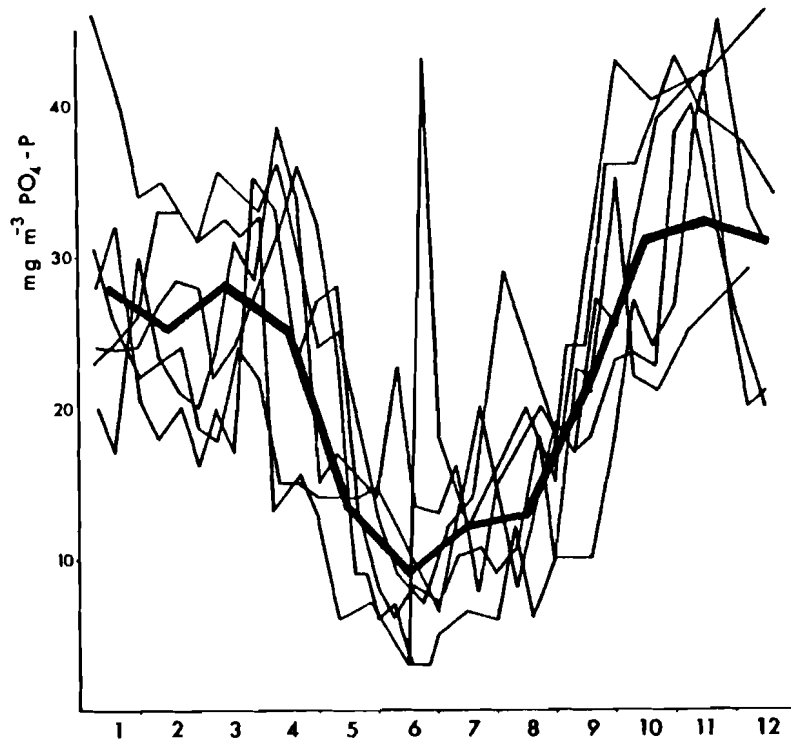


Figure 3: Phosphorus dynamics ( $P-PO_4$ ) for selected years from 1964 to 1979; thick line indicates monthly averages for the years 1965 to 1975; after unpublished data from Weigel und Mangelsdorf; Harms; Harms und Hagmeier; Harms, Mangelsdorf und Hagmeier; Mangelsdorf.

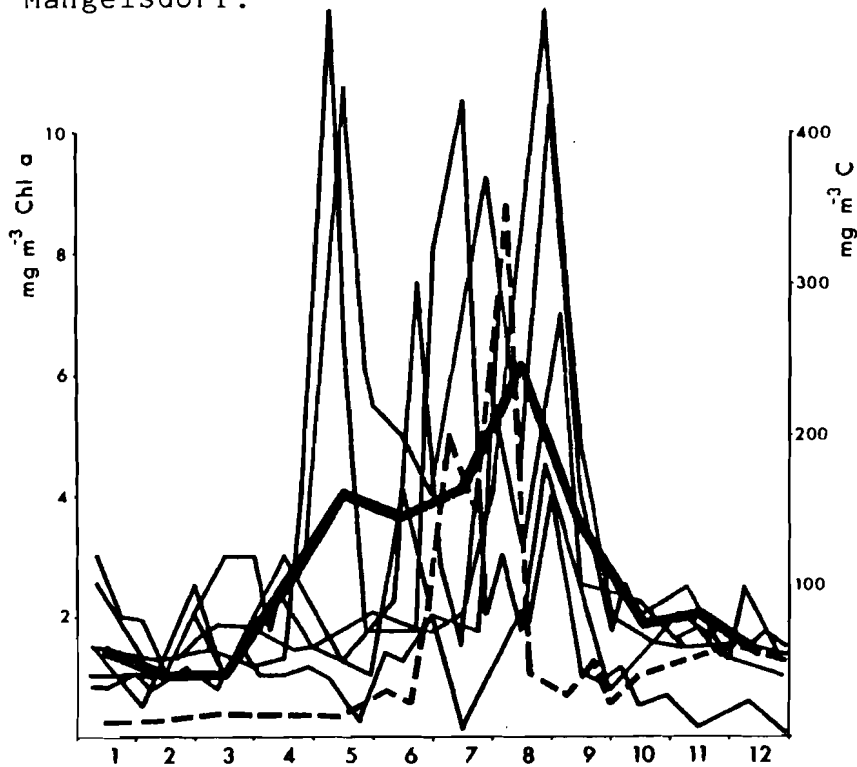


Figure 4: Chlorophyll dynamics for selected years from 1964 to 1979; thick line indicates monthly averages for the years 1965 to 1975; broken line shows micro-zooplankton carbon for the year 1975; after unpublished data from Weigel, Hagmeier and Treutner; Hagmeier, Kanje and Treutner.



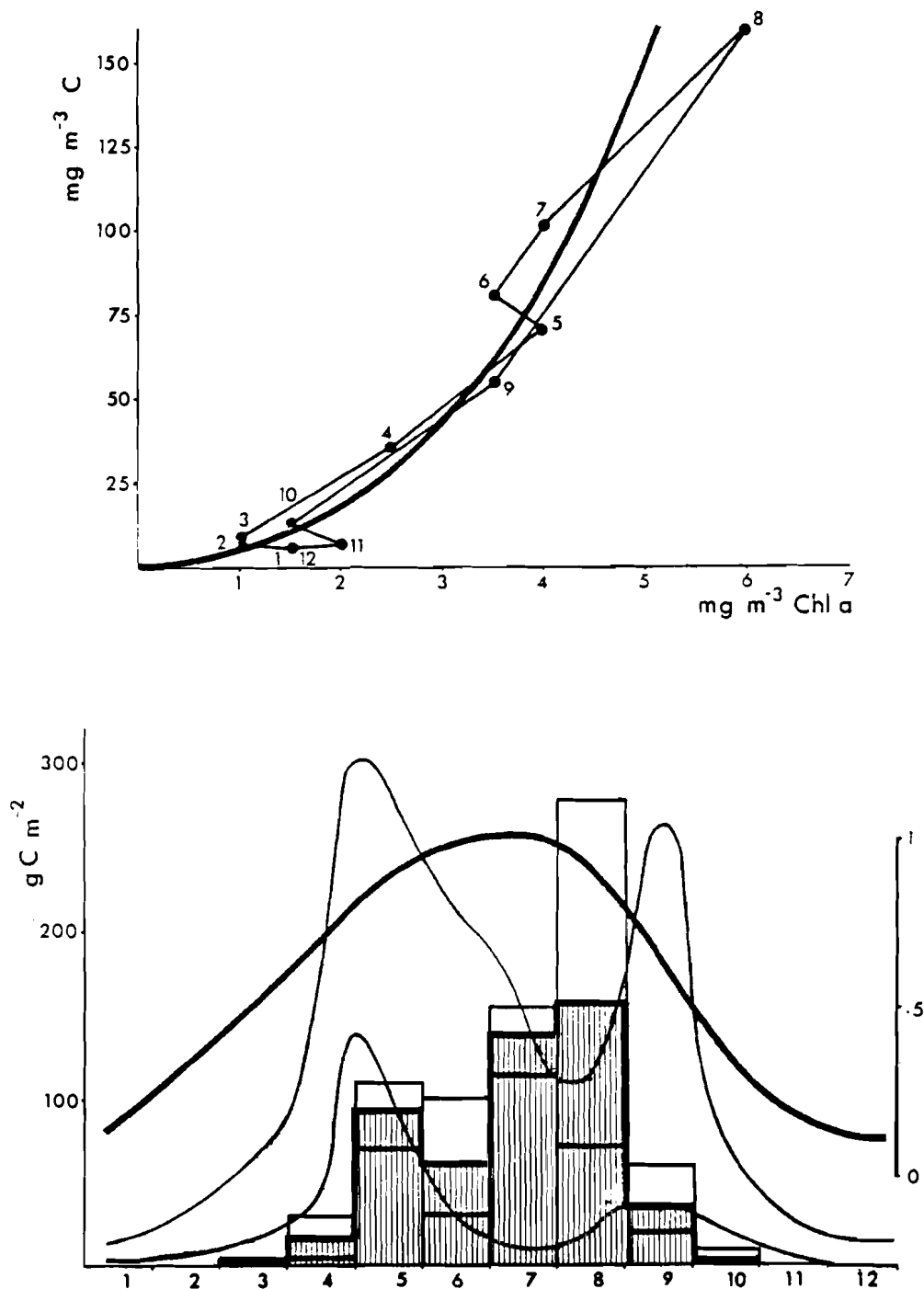


Figure 5a: Relationship between (monthly averages for 11 years) phytoplankton carbon and chlorophyll, approximated by an exponential curve; numbered dots indicate data points of individual months.

5b: estimates of monthly sums of primary production for the years 1966 - 1968, after Hagmeier; smooth curve shows the light- and temperature dependent relative productivity used in the models; envelope of thin lines for daily values after Mommaerts, 1975 (inner left scale).

Figures 3,4 and 5 summarize the data used. The driving environmental variables water temperature and radiation were found smooth enough and well behaved for a direct utilization of the longterm averages, approximated by simple sine waves. Data for nutrients ( $P-PO_4$ ) and algae (measured as chlorophyll as well as in terms of carbon, recalculated from counts) showed consistent yearly patterns. However, when including the year to year variations (as well as the implicit sampling errors), 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, which, due to their variable timing are averaged out when looking at the longterm mean (Fig.4). Also, the longterm mean is about one order of magnitude below the spiky peaks of the individual year's data. Little information was available on zooplankton biomass values. However, some additional information from independent experimentation, mainly on primary production, was also found. For example, estimates of monthly primary production for three years are shown in Figure 5b. Also, the (time-variable) ratio of phytoplankton carbon to chlorophyll was used for the models described below, approximated by a simple exponential curve (Figure 5a).

Among the invariable generalizable features derived from the observations are the following:

- ( 1) Primary producers are below a level of  $4. \text{ mg m}^{-3}$  chlorophyll during the first three months of the year;
- ( 2) between Julian day 120 and day 270 there is at least a twofold increase in biomass;
- ( 3) there have to be at least two peaks within that period, with a more than 25% reduction of the first peak value in between the two peaks;
- ( 4) after day 270, biomass must be below  $4. \text{ mg m}^{-3}$  chlorophyll again;
- ( 5) the higher of the two peak values must not exceed  $25 \text{ mg m}^{-3}$  chlorophyll;

- ( 6) yearly primary production must be above 300 and below 700 g C m<sup>-2</sup>;
- ( 7) herbivorous consumers (zooplankton) reach their first biomass peak value (defined at least as a two-fold increase of their initial biomass before a consecutive decline) after the phytoplankton;
- ( 8) the maximum density of herbivorous consumers must not exceed 1000 mg C m<sup>-3</sup>;
- ( 9) PO<sub>4</sub>-P has to be above 20 mg m<sup>-3</sup> between day 1 and 90;
- (10) the average between day 120 and 240 has to be below 20 mg m<sup>-3</sup>; (11) P-PO<sub>4</sub> has to be above 20 mg m<sup>-3</sup> after day 270;
- (12) it must never exceed 50 mg m<sup>-3</sup>; and it must never be below 2 mg m<sup>-3</sup>;
- (13-17) all state variables must be cyclically stable (+/- 25% tolerance level).

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

The description so far is little more than a summary of the more persistent patterns in the data. To make that a description of the system we want to study, we have to define what that system should include, its elements and its boundary conditions. This, however, is already part of the hypothesis generation, as the assumptions used here are no longer directly deduced from the data. Also, to make this potentially misleading point more explicit, the kind of data collected and the way they are collected is of course already part or rather consequence of a (generally implicit) conceptualization or model of the system. Measuring the few selected variables out of the very large

number of potentially measurable items already requires an implicit definition of the system under study and the assumption that whatever is measured, is an important and meaningful attribute of the system, in light of the objective of the respective study. Also, measurement strategies, i.e. the distribution of samples in time and space, imply numerous assumptions (as a rule untested assumptions) on the spatio-temporal behavior of the system.

In the example described, tidal patterns have been ignored for the timing of the sampling. From some measurements series, covering full tidal cycles with numerous measurements however, it is obvious that e.g. algae biomass (measured as cell numbers) can vary for more than an order of magnitude with the tides. Ignoring this short-term dynamic feature of the system in the sampling strategy considerably adds to the scatter of the observations.

#### HYPOTHESES GENERATION AND TESTING:

##### Designing Alternative Models

These relations might well lead to a circular argument: given a certain (implicit) hypothesis about the systems structure or function, information is collected according to this hypothesis, which is then in turn used to prove or--horrible dictu--"verify" the initial hypothesis, now made explicit. It therefore seems to be very important to make all these implicit assumptions based on a priori information explicit, to make all of them subject to critical, independent tests and to explore the consequences in terms of future testability of any complex hypothesis. Also, there are several implicit assumptions hidden in the way the data are interpreted and the description is derived. Ignoring the short-term spatio-temporal variations (e.g. caused by the tides) and looking at average features instead, implies that we are considering a hypothetical body of water, not absolutely fixed in space. The horizontal extension of this waterbody is rather arbitrarily limited by the require-

ment of homogeneity within this spatial element. In the vertical, the waterbody considered is defined by the extent of the measurements used, but again homogeneity has to be assumed.

Another crucial step to be made is the specification of boundary conditions: we assume the system as it will be described in the subsequent models to have no material exchange at its upper boundary, that is to say, with the atmosphere; we also assume that there are no lateral flows, which implies that for any element under study the environment is big enough and homogeneous to make flows due to advection/diffusion negligible; and finally we have to specify the conditions at the lower boundary, where we will assume an "endless sink" of constant chemical properties, which is very large when compared to the productive upper layer we are studying; the exchange between the upper layer and this sink will be controlled by eddy diffusivity.

All these assumptions are more or less unrealistic, whenever we think in terms of specific physical units in time and space; however, this is not what we attempt to model, and the basic idea behind all these assumptions is that the simplified process largely dominates the behavior of the conceptual system as compared to the processes ignored.

In the literature, one can find numerous conceptualizations or models of aquatic ecosystems, and 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 and Takahashi, 1977; Kremer and Nixon, 1978; Barnes and Mann, 1980). Numerous contributions to the literature appeared, some of them dealing 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 was covered with respect to biological and physiological factors (e.g. Steele and Frost 1977; Steele and Mullin, 1977; Morris, 1980), or with the emphasis on the physical and spatial aspects (e.g. Walsh, 1974; 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.

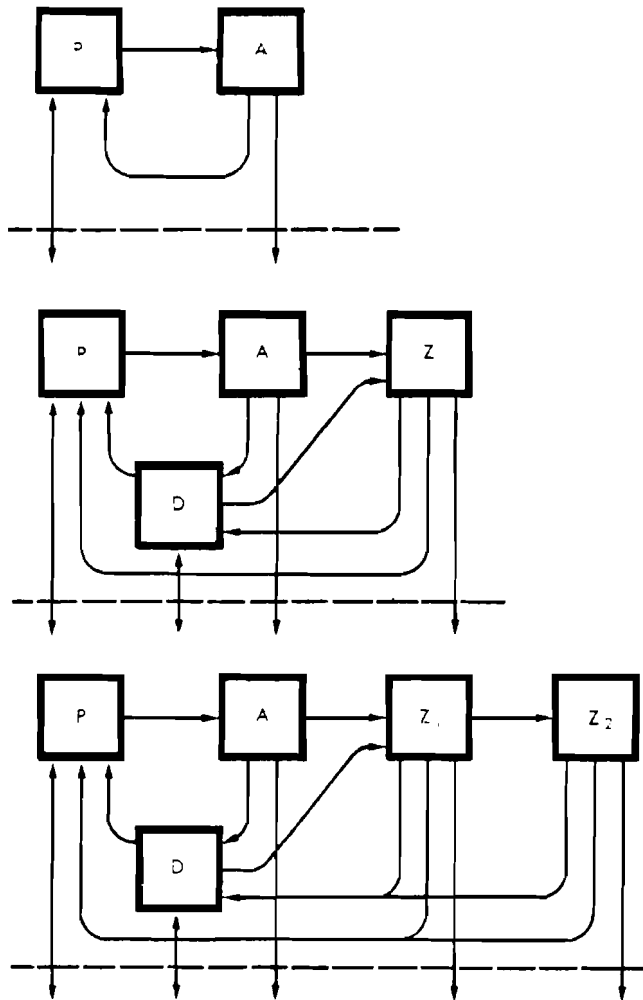


Figure 6: Flow diagrams for the models compared; P: phosphate; A: phytoplankton; D: detritus; Z: zooplankton; Z<sub>1</sub>: herbivores; Z<sub>2</sub>: carnivores.

Hypothesis No.1:

two compartments in a simple physical framework.

Let me now try to formulate one very simple hypothesis about the pelagic food web described in the data set above. Again it should be stressed that the model described below is not proposed 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 temperature, 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. Fig. 6 gives a diagrammatic representation of this system.

The model description uses Monod-kinetics to describe nutrient limitation of primary production, using 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 DiToro et al., (1971) of Steele's (1962) equation; (for a discussion of the implications of this formulation see Kremer and Nixon, 1978). Mortality is described as a nonlinear, concentration-dependent function of algae 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 upper layer's (10 m) volume, where the  $P-PO_4$  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 varia-

tion 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.

The governing model equations are summarized in the Appendix. This model requires only six parameters to be estimated, given the initial conditions and the driving variables are "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 run for 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 input-search-space) was followed with all the models described below. The model is formulated in terms of phosphorus, with constant stoichiometric conversions to carbon and a time-variable carbon-chlorophyll ratio (compare Figure 5). A discussion of the description of the major biological processes can be found in Fedra 1979.

#### Testing hypothesis number one.

To test the hypothesis formulated in Model 1, the model was incorporated into a Monte Carlo framework, which randomly sampled a set of model parameters from the allowable ranges (see Table 1), run the model for 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. This process was repeated for a sufficiently high number of trials (in fact, more than 100,000 model runs were performed with each of the models). Since 100,000 runs of even a comparatively simple simulation model produces a large amount of almost in-



comprehensible information, several auxiliary programs for the automatic analysis of the simulation results were used. Table 1 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 7.

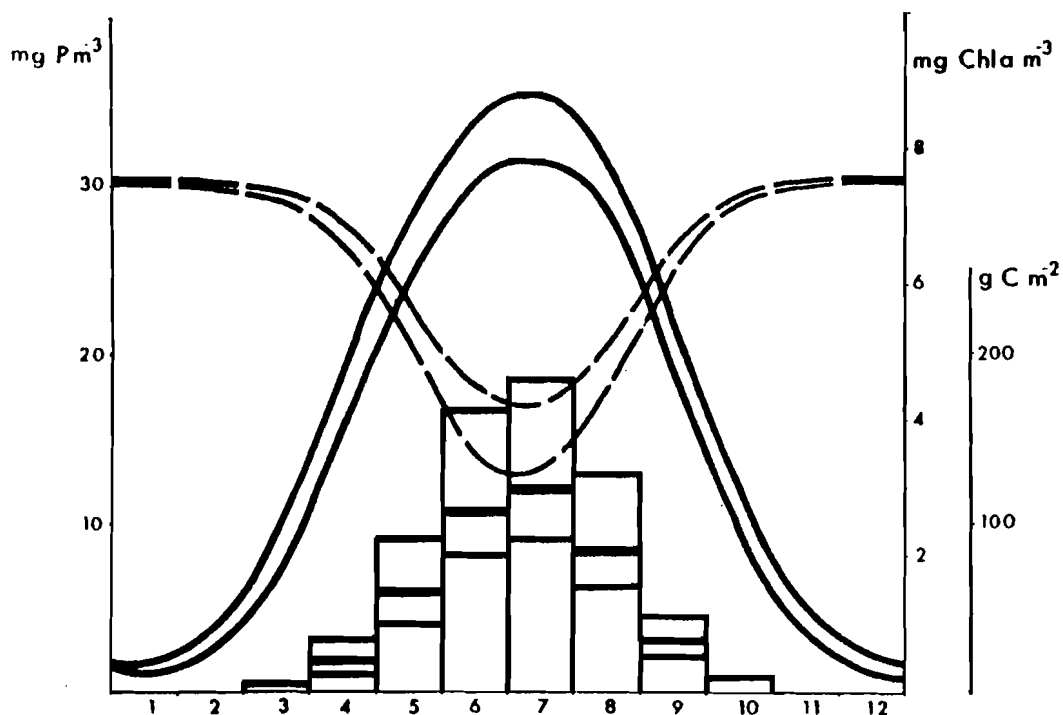


Figure 7: Sample output from Model 1, showing an envelope for the state variables phosphate and phytoplankton for a set of runs which only violate the condition of two phytoplankton peaks; histogram shows monthly sums (minimum, mean, maximum) of primary production.

TABLE 1: Automatic Parameter Estimation Analysis Program

Parameter Statistics for MOD1.mc - output

run selection: violation of condition 3 only

	31 runs evaluated	Mean	Minimum	Maximum	S.D.	range sampled
Parameter values:						
1 Michaelis const.	6.08	2.23	13.84	3.33	2.00	15.00
2 Phytoplankton mort.	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 coeff.	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: (all concentrations in mg m<sup>-3</sup>)

7 Chl. high 1-90	0.	0.	0.	0.	(not violated)	
8 Chl. summer peak	8.16	7.84	8.72	0.24		
9 Chl. first	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 270	0.	0.	0.	0.	(not violated)	
15 Chl. maximum	8.16	7.84	8.72	0.24		
16 PO <sub>4</sub> maximum	30.00	30.00	30.00	0.		
17 PO <sub>4</sub> minimum	15.47	12.86	16.87	1.07		
18 PO <sub>4</sub> low before day 90	29.92	29.67	29.99	0.092		
19 PO <sub>4</sub> low after day 270	27.84	27.56	28.36	0.18		
20 PO <sub>4</sub> average 120-240	19.02	17.27	19.93	0.73		
21 Primary prod. g C m <sup>-2</sup>	406.20	303.05	625.44	86.58		

Correlation matrix of Parameters

	1	2	3	4	5
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	0.3

Summarizing, 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 7 shows a sample output from Model 1.

Hypothesis number one 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 algae peaks) were analyzed. The technical details of this kind of analysis are described elsewhere (Fedra et al., 1980). However, the analysis indicates that phytoplankton mortality is a critical process, and consequently deserves refinement. This can be deduced from the significant correlations between the mortality rate coefficient and the other parameters as well as different output variables in groups of simulations violating different constraint conditions.

Hypothesis No. 2: a four-compartment web.

As a slightly more realistic alternative to Model 1, a second version was formulated which incorporates detritus and omnivorous zooplankton. The description of primary production as well as the physical framework are essentially the same as in the first version. Model two, however, splits the phytoplankton mortality in a natural background-mortality, which is described as concentration dependent, and losses due to grazing. Background mortality as well as zooplankton mortality now feed 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 in the nutrient pool. Figure 6b shows the flowchart for this model. Grazing was described based on a simple encounter theory. With this inclusion of a herbivorous zooplankton compartment, a choice had to be made on how to describe grazing. Numerous, different, formulations abound in the literature, and to give

one single example, Jørgensen (1980, Table 3.9) lists 14 different formulations of zooplankton grazing rates. Given 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 this 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 to grow in the spring, with a consequent collapse of the zooplankton population itself. This however, 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 contributes to them.

However, after "rejecting" 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 the Appendix.

Again this version was subjected to the above described simulation procedure, the resulting response was analyzed (see Table 2a). 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 fulfillment of condition 2, requiring two phytoplankton peaks. However, this version was incapable of producing enough algae carbon over the year, thus violating condition 6 (see Table 2b). This is simply due to the fact that only at comparatively low primary productivity levels the system was 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 Tables 2) and zooplankton respiration (parameter 7), which is a major source of nutrient recycling. This directly points at the positive feedback loop in these processes, and the resulting stability

problems in this version of the model.

TABLE 2a: Automatic Parameter Estimation Analysis Program,

Parameter Statistics for MOD2.mc - output

run selection: 3123 runs numerically stable over 3 years of simulation

3123 runs evaluated	Mean	Minimum	Maximum	S.D.	range sampled	
Parameter values:						
1 Michaelis constant	9.17	6.00	11.99	1.72	6.00	12.00
2 Phytoplankton mort.	0.047	0.	0.10	0.025	0.00	0.10
3 Light optimum	386.07	300.05	499.84	57.03	300.00	500.00
4 Mixing coefficient	0.07	0.05	0.13	0.01	0.05	0.15
5 Max. growth coefficient	1.88	0.50	2.50	0.44	0.50	2.50
6 Zooplankton grazing	0.38	0.05	1.00	0.20	0.05	1.00
7 Zoopl. detritus uptk.	0.07	0.	0.15	0.04	0.00	0.15
8 Zoopl. respiration	0.06	0.01	0.15	0.04	0.01	0.15
9 Zoopl. mortality	0.27	0.05	0.50	0.13	0.05	0.50
10 Remineralization	0.14	0.05	0.25	0.06	0.05	0.25
11 Temperature trigger	2.95	2.50	3.50	0.29	2.50	3.50
12 Grazing half-sat.	13.02	0.55	20.00	4.98	0.50	20.00

Output-constraint variables: (all concentrations in mg m<sup>-3</sup>)

13 Chl. high day 1-90	7.73	0.	9.65	0.93		
14 Chl. summer peak	7.33	3.36	10.08	1.16		
15 Chl. first peak	7.10	0.94	10.08	1.34		
16 Chl. low between peaks	5.29	0.03	9.52	1.62		
17 Day of first peak	122.74	120.00	250.00	10.58		
18 Chl. second peak	5.59	0.	9.61	1.62		
20 Chl. maximum	8.19	6.29	10.08	0.62		
21 PO <sub>4</sub> maximum	22.23	8.41	25.36	2.32		
22 PO <sub>4</sub> minimum	1.17	0.001	7.14	0.87		
23 PO <sub>4</sub> low until day 90	3.26	0.66	24.36	1.74		
24 PO <sub>4</sub> low after day 270	6.29	0.32	21.86	2.85		
25 PO <sub>4</sub> average 120-240	3.26	0.12	15.39	2.02		
26 Primary production	240.92	41.51	442.56	66.92		
27 Day of zoopl. peak	180.29	110.00	325.00	32.33		
28 Zoopl. at algae peak	5.76	0.	19.89	4.43		
29 Zoopl. peak value	9.29	0.006	21.37	4.04		

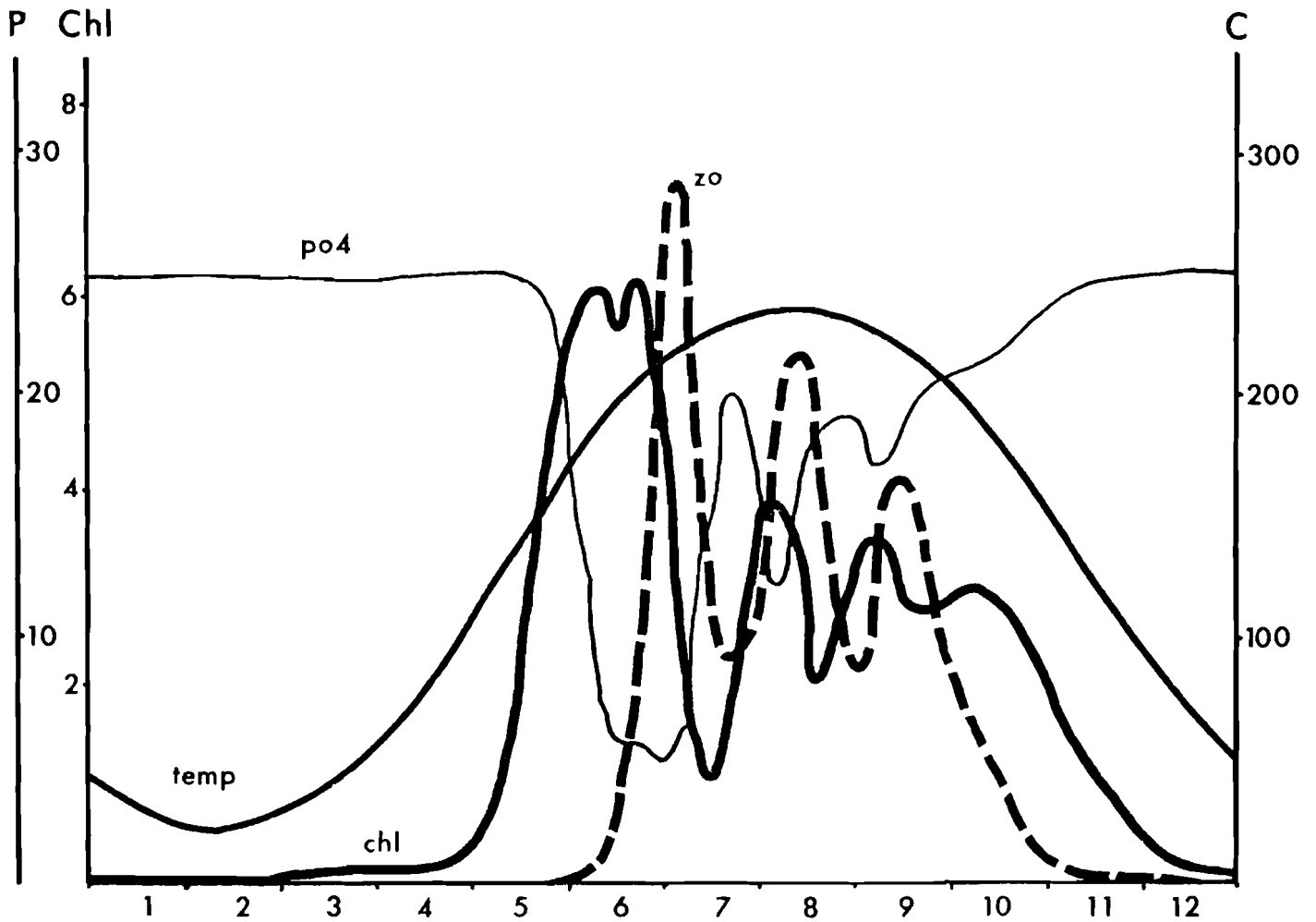


Figure 8: Sample output from Model 3, run violating only the condition of yearly primary production above  $300 \text{ g C m}^{-2}$ . Thick line: phytoplankton (chlorophyll a in  $\text{mg m}^{-3}$ ); broken line: herbivorous zooplankton (carbon in  $\text{mg m}^{-3}$ ); thin line: phosphate ( $\text{P-PO}_4$  in  $\text{mg m}^{-3}$ ); smooth curve: surface temperature (in degree centigrade).

TABLE 2b: Automatic Parameter Estimation Analysis Program,

Parameter Statistics for MOD2.mc - output

run selection: 43 runs violating condition 6 only (primary production)

	43 runs evaluated	Mean	Minimum	Maximum	S.D.	range	sampled
Parameter values:							
1 Michaelis constant	9.88	5.06	14.88	3.17	5.00	15.00	
2 Phytoplankton mort.	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 Zoopl. detritus uptk.	0.10	0.005	0.20	0.06	0.01	0.50	
8 Zoopl. respiration	0.09	0.02	0.28	0.06	0.01	0.25	
9 Zoopl. 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-sat.	13.80	3.94	23.44	5.07	0.00	25.00	
Output-constraint variables: (all concentrations in mg m <sup>-3</sup> )							
13 Chl. high day 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 <sub>4</sub> maximum	25.76	24.76	26.72	0.433			
22 PO <sub>4</sub> minimum	4.83	2.16	11.97	2.19			
23 PO <sub>4</sub> low until day 90	25.36	24.73	25.91	0.23			
24 PO <sub>4</sub> low after day 270	22.19	20.06	25.11	1.12			
25 PO <sub>4</sub> average 120-240	17.11	13.70	19.86	1.67			
26 Primary production	40.60	19.08	75.76	14.08			
27 Day of zoopl. peak	165.63	145.00	220.00	16.20			
28 Zoopl. at algae peak	0.69	0.	3.48	0.92			
29 Zoopl. peak value	14.33	7.66	18.74	3.00			

Hypothesis No.3: one more trophic level

Consequently, 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 and Reiners, in press). A sample output of this version 3 is shown in Figure 8, and the equations are given in the Appendix. Another 5 additional 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 runtime aborted runs (due to the violation of some runtime-checks on the state variables, confining them within certain plausible ranges or numerical instabilities in solving the system of differential equations) grew dramatically to almost 99.9 % of the trial runs when sampling the broad initial parameter intervals given in Table 3.

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. Due to its higher complexity, Model 3 was able to generate a broad spectrum of behavioral features (compare Table 3); 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 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 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. Or, alternatively, if all plankton groups survive the winter well, the onset of high primary productivity will quickly lead to explosive growth and consequent collapse.



TABLE 3: Parameter statistics for Model 3 - output:  
 250000 trial runs made  
 219 runs evaluated

	mean	minimum	maximum	S.D.
1 Michaelis constant	11.131	2.306	19.816	4.818
2 Phytoplankton mortality	0.192	0.	0.496	0.136
3 Light optimum	419.037	300.647	548.945	74.599
4 Mixing coefficient	0.050	0.001	0.197	0.045
5 Maximum growth coeff.	7.097	0.764	9.977	2.110
6 Zooplankton grazing	1.012	0.003	1.995	0.573
7 Zooplankton detritus uptk	0.508	0.044	0.799	0.191
8 Zooplankton respiration	0.180	0.	0.722	0.169
9 Zooplankton mortality	0.206	0.	0.779	0.184
10 Remineralization	0.205	0.	0.497	0.143
11 Temperature trigger	2.943	2.502	3.493	0.289
12 Grazing rate carnivores	1.152	0.038	1.989	0.532
13 Mortality rate carnivores	0.146	0.001	0.787	0.111
14 Respiration carnivores	0.067	0.	0.620	0.085
15 MM constant algae	16.417	0.431	29.995	7.902
16 MM constant detritus	10.770	0.002	29.803	8.802
17 MM constant herbivores	15.066	0.109	29.653	8.821

parameter ranges sampled:

parameter No.:	1	2.000	20.000
parameter No.:	2	0.	0.500
parameter No.:	3	300.000	550.000
parameter No.:	4	0.	0.200
parameter No.:	5	0.500	10.000
parameter No.:	6	0.	2.000
parameter No.:	7	0.	0.800
parameter No.:	8	0.	0.800
parameter No.:	9	0.	0.800
parameter No.:	10	0.	0.500
parameter No.:	11	2.500	3.500
parameter No.:	12	0.020	2.000
parameter No.:	13	0.	0.800
parameter No.:	14	0.	0.800
parameter No.:	15	0.	30.000
parameter No.:	16	0.	30.000
parameter No.:	17	0.	30.000

Output constraint variables: (all concentrations in  $\text{mg m}^{-3}$ ,  
production values in  $\text{gC m}^{-2}$ )

chl high 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 270	4.678	1.039	9.112	1.689
chl maximum	6.423	2.385	10.111	1.488
po4 maximum	23.616	6.541	27.994	3.250
po4 minimum	1.975	0.062	18.797	2.098
po4 low until 90	15.387	1.351	25.170	8.266
po4 low after 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
po4 average 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
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
zoopl. 2 end	1.139	0.	41.086	3.933
zoopl. 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

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  $300 \text{ g C m}^{-2}$  could seem unrealistically high. Another comparable estimate, given in Pichot and Runfola (1975) for the Southern Bight off the Belgian coast, is given with 17.5

g N m<sup>-2</sup>, which amounts to less than half the estimate of Hagmeier (BAH reports 1976-1969), when converted to carbon units. Reducing the constraint of minimum yearly primary production to, say 100 g C m<sup>-2</sup>, would make the model "acceptable".

This points at 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 system's conceptualization. The problem is not a purely statistical one, which--in principle--could be overcome by changes in the sampling and measurement strategies, e.g. by larger sample sizes or increasing the number of sampling units. More often than not, micro-scale measurements and macro-scale concepts are also qualitatively different. Quite obviously, the definition of the constraint conditions can be critical, and thus indicates where further effort in data analysis (or collection) would be worthwhile.

Hypothesis No.4: more biological detail

Since zooplankton dynamics were found to play a critical role in the system's dynamics, a more detailed and biologically "realistic" representation of grazing and zooplankton population dynamics was attempted. To test the importance of the (size- and age-class dependent) internal control mechanisms in the zooplankton compartment, Steele's (1974) model was incorporated into the same Monte Carlo framework as the above models. Several modifications had to be made, since the model describes the pelagic system in terms of nitrogen and zooplankton numbers and individual size. Also, the model which was originally developed for the northern North Sea was designed for the stratified summer period only, and ignores the effects of temperature as a controlling variable for the biological rate constants. The model incorporates nutrients (nitrogen), phytoplankton, herbivorous zooplankton numbers, individual size, and eggs produced

(which translates into number of juveniles released through the specification of an initial size). Egg production starts, as soon as a certain individual size is reached. Zooplankton dynamics are all related to individual size by a simple allometric relation. The simplest possible version with only one zooplankton age-group or cohort was used. The basic structure of the model (and the philosophy behind it) are fully explained in Steele (1974), and further extensions and application examples are reported in Steele and Mullin (1977), Steele and Frost (1977), and Frost (1980). Since the model is formulated in terms of nitrogen, the respective nitrogen data set from the system was used (Figure 9). To illustrate the general patterns, some more comparable data sets from the North Sea are also included in the graph.

To summarize, the model which was originally designed for the period of summer stratification only, does well during this period (Figure 10). When extending over a whole yearly cycle however, it suffers from the same (generic ?) shortcomings as the other models discussed. Either the zooplankton population collapsed over the winter, or explosive growth and consequent collapse was observed during the summer. This is at least in part, attributable to the fact that the model does not consider temperature dependency of the biological processes (since it was designed for the summer period in the northern part of the North Sea only); also, phytoplankton production is described as a function of nutrient concentrations only, and the model does not include detritus. However, a more likely conclusion seems to be that there are qualitative differences between summer and winter periods in the dynamics of the planktonic systems. These might involve more complex, adaptive control mechanisms. One could think of thresholds in environmental conditions, triggering different behavior, representing e.g. resting stages, or seasonal changes in species composition. Also, time variable material inputs to the (almost estuarine) system could be necessary to describe the full range of a yearly cycle of behavior more realistically. This would of course require yet another reformulation of the models.

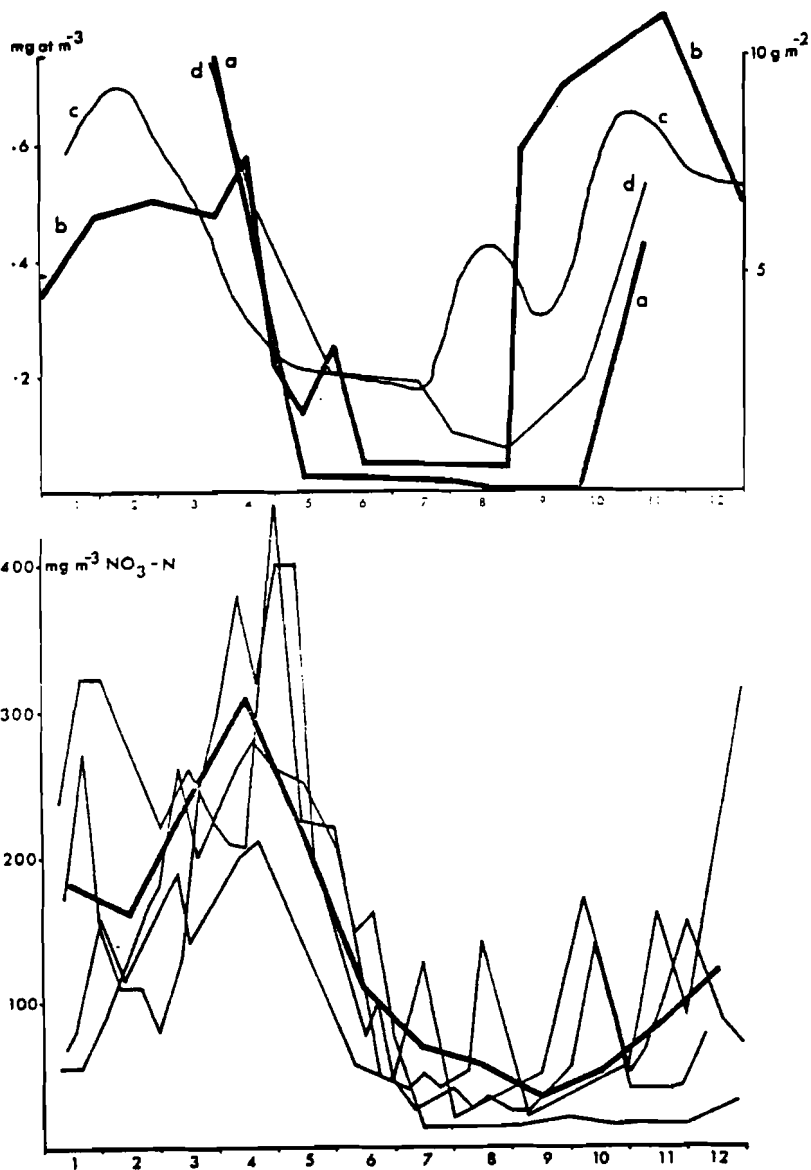


Figure 9: Comparable data sets from the North Sea: (a) and (d) Nitrate and phosphate concentrations, center of northern North Sea after Steele, 1974 (inner and outer left scale); (b) dissolved Nitrogen after Janssen and Meuris 1975, (right scale); (c) phosphate after Steele, 1962, from Nihoul 1975 (outer left scale); lower part: Selected Nitrate data from the BAH Reports, compare Figs. 3 and 4.

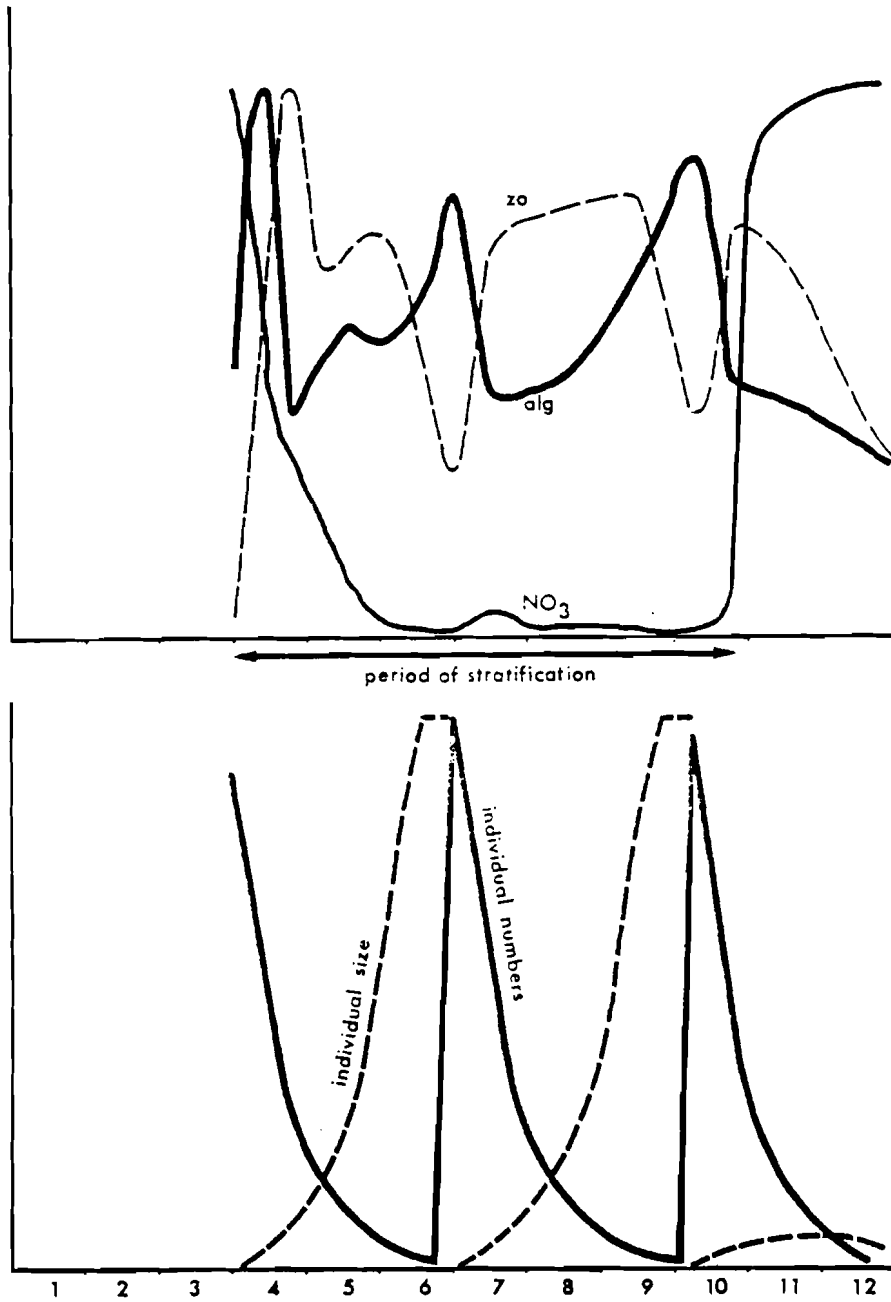


Figure 10: Model output from Steele's model: nutrients, phytoplankton, and zooplankton biomass (top); zooplankton numbers and individual size (bottom).

Some of the models found in the literature, which describe the pelagic food web of the North Sea (or parts of it), are designed for certain periods of the year only (e.g. Steele 1974; Greve 1981). In case of Steele's model, it could be shown that the model cannot be extended over a full yearly cycle or even a sequence of several years without major modifications. As another example, Greve's (1981) model--being rather sophisticated in the description of species interactions in the higher trophic levels--is designed for a short period of zooplankton development only. Others, like the model of Radach and Maier-Reimer (1977) are more theoretically oriented, and do not attempt to describe yearly cycles of the natural systems at all. 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 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 used, although seemingly liberal, are quite demanding when compared with many examples of arbitrary judgement, 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 paper 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 approach to model or hypothesis testing.

DISCUSSION:

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 however, 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 formulates 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 (compare Figure 1).

The method requires the formal definition of an acceptable model response a priori. In this definition, arbitrary classifications and subjective judgements cannot always be avoided. Although being based on the available field data, the definition has to be formulated on the model's level of abstraction. 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 (compare Figure 1).

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 judgements. How small would the sum of squared errors have to be for a given state variable to make a model acceptable



? Although a least square criterion may be helpful in finding a "best" parameter set (according to that least square 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 judgement a posteriori has to be used. Examples abound where only partial agreement of model output and observations are described as "acceptable or reasonably good fit", ignoring the fact that severe discrepancies between parts of the model response (e.g. for some of the state variables) and the observations exist. This is most obvious in case of the introduction of unmeasured (and consequently unconstrained) state variables in a model--bacteria are an almost classical example in water quality modeling.

One basic idea of the approach is to use the available information according to its relevance to the models' (this is the theory's) level of abstraction or aggregation. Assuming a given model structure, this information is grouped into a set of singular statements, the initial conditions, to be substituted for the variables of the universal statement (the theory or model), and--since we are doing the analysis 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 possible e.g. 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 (Fedra et al., 1981). 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--mass and energy conservation, or more empirical rules like maximum efficiencies or process rates.

Obviously, the description of the states of a system can be accomplished much more easily on the appropriate level than the description of 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 response-hyperspace of a model back into the input-hyperspace.

The test is then as follows: whether or not this region in the input space exists within the specified possible or plausible bounds. In addition, several other features of the inputspace can be used as a basis for either rejecting or corroborating a given hypothesis, for example, the uniqueness of the inputspace region, whether it is closed or not, and its structure, which is determined by the interdependencies of the individual input values. In addition, all these features, including the relationship or correlation of input- and outputspace, allow us to learn something of the way the proposed systems' structure functions. The method facilitates an understanding of the systems behavior at the appropriate level of abstraction, which is the input and output of the model, and it also provides diagnostic information for hypothesis generation.

Hypothesis generation, that is the conjecture of the initial or an alternative hypothesis after the failure of a previous one, is a crucial step: the hypotheses we are using in environmental systems analysis are fairly complex, or rather composite, that is to say, they are built from numerous individual constructs, each of them being a hypothesis in itself. Their complex, dynamic and nonlinear interactions makes 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 versus observations -- as has been proposed by some authors--as soon as feedbacks between the isolated process and the remainder of the system exist. In complex environmental examples, this will almost always be the case.

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. Although the two or more hypotheses then 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, i.e. 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 field work as well, allowing for a more precise formulation of questions to be addressed in the expensive field observations.

The approach described above, although largely based on trial and error and the extensive (ab)use of computers, can be exploited in more than one way. It can provide a rational and formal framework for the analysis of complex systems, help in model selection, be used for model calibration, and finally for the probabilistic interpretation of model predictions (Fedra, in press a). But above all, the approach emphasizes testability. Any rigorous scientific approach to the study and analysis of complex, hard to handle systems which are no longer easily understandable and traceable, requires that all the individual elements of the systems' conceptualization, all the assumptions that are necessary, are made explicit - and thus testable.

REFERENCES

Andersen, K.P. and Ursin, A. 1977 A multispecies extension of the Beverton and Holt theory of fishing, with accounts of phosphorus circulation and primary production. Meddr Danm. Fisk.-og Havunders. N.S.7, 319-435.

Barnes, R.S.K. and Mann, K.H. (Eds.) 1980 Fundamentals of Aquatic Ecosystems. Blackwell Scientific Publication, Oxford; 229 pp.

Biologische Anstalt Helgoland, Jahresberichte 1964 - 1979.

Cushing, D.H. and Walsh, J.J. (Eds.) 1976 The Ecology of the Seas. Blackwell Scientific Publications, Oxford; 467 pp.

DiToro, D.M., O'Connor, D.J., and Thomann, R.V. 1971 A dynamic model of phytoplankton populations in the Sacramento-San Joaquin Delta. Advan. in Chem. Series 106, 131-180.

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

Dubois, D. and Closset, P.L. 1976. Patchiness in primary and secondary production in the Southern Bight: a mathematical theory. In: G. Persoone and E. Jaspers [eds.], Proceedings of the 10<sup>th</sup> European Symposium on Marine Biology, University Press, Wetteren, 2:211-229.

Fedra, K. 1979 Modeling biological processes in the aquatic environment; with special reference to adaptation. WP-79-20, 56pp, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Fedra, K. 1980 Mathematical Modeling - A Management Tool for Aquatic Ecosystems? Helgoländer Meeresunters. 34:221-235.

Fedra, K., van Straten, G. and M.B. Beck 1981 Uncertainty and arbitrariness in ecosystems modeling: a lake modeling example. *Ecological Modeling*, 13: 187-210.

Fedra, K. (in press a) Estimating model prediction accuracy: a stochastic approach to ecosystems modeling. Proc. of the IInd State-of-the-Art Conference in Ecological modeling, Liege, Belgium, 1980.

Fedra K. (1980) A Monte Carlo approach to estimation and prediction. of Water Quality. Proceeding of the IIASA workshop on Uncertainty and Forecasting IIASA Conference Paper (M.B. Beck and G. van Straten eds.) (forthcoming)

Frost, B.W. Grazing. In: Morris, I. (Ed.) 1980 The Physiological Ecology of Phytoplankton. *Studies in Ecology*, Vol. 7, Blackwell, Oxford; 465 - 491.

Greve, W. and Reiners, F. (in press) The impact of prey-predator waves from estuaries on the planktonic marine ecosystem. *Estuarine Perspectives*.

Greve, W. (in press) Invertebrate predator control in a coastal marine ecosystem: the significance of *Beroe gracilis* (Ctenophora). Proc. 15<sup>th</sup> Europ. Mar. Biol. Symp., Kieler Meeresforsch., Sonderh. 5, 1981.

Hagmeier, E. 1978 Variations in phytoplankton near Helgoland. *Rapp. P.-v. Reun. Cons. int. Explor. Mer*, 172, 361-363.

Janssen, D. and Meuris, A. 1975 Dynamic behavior of nutrients. Tech. Rep. 75/Chim. Synth. 01. Progm. natan Envir. phys. biol., Projet Mer, Dep. Sci. Policy, Belgium, 10pp.

Jørgensen, S.E. 1980 Lake Management. *Water Development, Supply and Management*, Vol. 14 Pergamon, Oxford. 167 pp.

Kremer, J.N. and Nixon, S.W. 1978 A Coastal Marine Ecosystem.

Ecological Studies 24, 217pp, Springer New York.

Lassen, H. and Nielsen, P.B. 1972 Simple Mathematical Model for the Primary Production. ICES, Plankton Committee CM 1972/L:6

Lucht, F. and Gillbricht, M. 1978 Long-term observations on nutrient contents near Helgoland in relation to nutrient input of the river Elbe. Rapp.P.-v. Reun.Cons.int.Explor.Mer, 172: 358-360.

Mommaerts, J.P. 1975 Four years of primary production measurements. A partial synthesis. Tech. Rep. Biol. 75/synth. 01 Progm natn Envir. phys. biol., Projet Mer, Dep. Science Policy, Belgium, 9pp.

Morris, I. (Ed.) 1980 The Physiological Ecology of Phytoplankton. Studies in Ecology, Vol.7, Blackwell, Oxford; 625pp.

Nihoul, J.C.J. 1975 Modeling of Marine Systems. Elsevier Oceanography Series, 10, 272pp, Elsevier, Amsterdam.

Pichot, G. and Runfola, Y. 1975. Mathematical model of the nitrogen cycle in the Southern Bight of the North Sea. In: G.Persoone and E.Jaspers [eds.], Proceedings of the 10<sup>th</sup> European Symposium on Marine Biology, University Press, Wetteren, 2:467-476.

Popper, K.R. 1959 The Logic of Scientific Discovery. 480 pp, Hutchinson, London.

Radach, G. 1980. Preliminary simulations of the phytoplankton and phosphate dynamics during FLEX'76 with a simple two-component model. Meteor Forsch. Ergebnisee A, No.22:151-163

Radach, G. and Maier-Reimer, E. 1975. The vertical structure of phytoplankton growth dynamics. A mathematical model. Mem.Soc.Roy.Sci. de Liege, Ser.6,7 113-146.

Steele, J.H. 1962 Environmental control of photosynthesis in the

sea. *Limnol. Oceanogr.* 7, 137-150.

Steele, J.H. 1974 *The Structure of Marine Ecosystems*. 128pp, Harvard Univ. Press. Cambridge, Massachusetts.

Steele, J.H. 1976 Patchiness. In: Cushing, D.H. and Walsh, J.J. (Eds.) 1976 *The Ecology of the Seas*. Blackwell Scientific Publications, Oxford; 98-115.

Steele, J.H. (Ed.) 1978 *Spatial patterns in plankton communities*. Plenum Press, New York 470pp.; NATO Conference Series, Ser.:4 Marine Sciences, Vol.3.

Steele, J.H. and Frost, B.W. 1977. The structure of plankton communities. *Phil. Trans. Royal Soc. London. Series B*, Vol.280, 485-534.

Steele, J.H. and Mullin, M.M. 1977 Zooplankton dynamics. In: E.D. Goldberg et al. (Eds.), *The Sea*. Vol.6, 857-890, John Wiley, New York.

Thomann, R.V. 1978 Size-dependent model of hazardous substances in aquatic food chains, *Ecological Research Series*, EPA-600/3-78-036.

Walsh, J.J. 1975 A spatial simulation model of the Peru upwelling ecosystem. *Deep-Sea Research*, Vol.22: 201-236.



APPENDIX : model equations

parameter numbers (pmetr(i)) correspond to the numbers used in the Tables above. The equations are given in FORTRAN code.

\*\*\*\*\*

variable names and meanings:

alg	algae biomass, expressed in nutrient units
ass	zooplankton assimilation
chl	chlorophyll a, $\text{mg m}^{-3}$
death	mortality/mineralization of phytoplankton
depth	depth of productive layer in meters
dgraz	zooplankton grazing (detritus)
dmix	detritus exchange with deep water
Dt*	time derivatives
eps	extinction coefficient in $\text{m}^{-1}$
fotop	photoperiod in fraction of a day (24 hrs)
gmax	maximum growth rate of algae
graz	zooplankton grazing (phytoplankton)
growth	actual growth rate of algae
plim	nutrient limitation factor
po4	limiting nutrient (phosphorus)
rad	radiation in $\text{cal cm}^{-2}$
radlim	light limitation factor
rem	mineralization of detritus
remo	phytoplankton removal by zooplankton (not assimilated)
resp	zooplankton respiration
sink	algae losses due to sinking/mixing
temp	water temperature in Centigrade
tfact	temperature factor
vmix	nutrient exchange with deep water
xmix	ratio of volume exchanged
zgraz	carnivorous zooplankton grazing (zooplankton)
zloss	zooplankton losses due to mixing
zmort	zooplankton mortality
zo	omnivorous zooplankton
zo2	carnivorous zooplankton
zresp	carnivorous zooplankton respiration
z2mort	carnivorous zooplankton mortality

```
*****
c
c food-chain simulation model:
c two compartment versions 1:
c for parameter ranges compare Table 1
c
c primary production:
c
c self shading:
c
c     chl = .64 * (40.* alg)** .4
c     eps = .04 + .054 * chl ** .67 + .009*chl
c     zk  = eps * depth
c
c light limitation:
c
c     ratio  = 0.9 * rad / pmeter(3)
c     x1     = ratio * exp(-zk)
c
c     part1  = exp(    -x1)
c     part2  = exp( -ratio)
c     part3  = (part1 - part2) / zk
c
c     radlim = part3 * fotop * 2.7
c
c nutrient limitation:
c
c     plim = po4/(po4 + pmeter(1))
c
c temperature dependent max. growth rate:
c
c     gmax = pmeter(5) * exp(.065*temp)
c
c     growth = alg * gmax * radlim * plim
c     sink   = alg * xmix
c     death  = alg **1.25 *pmeter(2)
c
c     Dtalg = growth -sink -death
c
c nutrients:
c
c     vmix = (po40-po4) * xmix
c
c     Dtpo4 = -growth + death + vmix
c
*****
```

```
*****
c
c MOD2 : four compartment web
c for parameter ranges compare Table 2
c
c herbivorous grazing:
c
c   temperature factor:
c     tfact = exp(-2.3 * abs(temp-16.5)/15.)
c
c   saturation and threshold:
c     fph = max(0.,(alg-alg0)/(alg+pmeter(12)))
c     rate = pmeter(6) * fph * tfact
c
c     graz = alg * zo * rate
c
c algae loss due to sinking/mixing:
c     sink = alg * xmix
c
c detritus remineralization:
c     rem = det * pmeter(10) * exp(.065*temp)
c
c zooplankton detritus uptake
c     dgraz = det * zo * pmeter(7)
c
c zooplankton respiration
c     resp = zo * pmeter(8) * exp(.06*temp)
c
c zooplankton mortality
c     zmort = zo**1.2 * pmeter(9)
c
c losses from upper layer due to mixing
c     zloss = zo * xmix
c
c detritus exchange with deep water
c     dmix = (detlow-det) * xmix
c
c     Dtalg = growth - death - graz - sink
c     Dtpo4 = - growth + rem + resp + vmix
c     Dtzo  = graz + dgraz - resp - zmort - zloss
c     Dtdet = death + zmort - dgraz + dmix - rem
c
*****
```

```
*****
c
c MOD3: five compartment foodweb
c
c for parameter ranges compare Table 3
c
c herbivorous grazing
  graz = (alg/(alg+pmeter(15))) * pmeter(6) * tfact * zo
c
c assimilation
  ass = min(graz,zo)
c
c phytoplankton removal
  remo = graz - ass
c
c detritus uptake
  dgraz = (det/(det+pmeter(16))) * zo * pmeter(7)
c
c respiration, mortality and losses due to mixing
  resp = zo * pmeter(8) * tfact
  zmort = zo**1.2 * pmeter(9)
  zloss = zo * xmix/4.
c carnivorous grazing, mortality and respiration
  zgraz = (zo/(zo+pmeter(17))) * zo2 * pmeter(12) *tfact
  z2mort = zo2**1.2 * pmeter(13)
  zresp = zo2 * pmeter(14) * tfact
c
c algae
  Dtalg = growth -death -graz -aloss
c nutrients (phosphorus)
  Dtpo4 = -growth +rem +resp +zresp +vmix
c omnivorous zooplankton
  Dtzo = ass +dgraz -zgraz -resp -zmort -zloss
c carnivorous zooplankton
  Dtzo2 = zgraz -zresp -z2mort
c organic detritus
  Dtdet = death +remo +zmort2 +zmort -dgraz +dmix -rem
c
*****
*****
```