Ecological Modelling, 4 (1978) 29—40

THE EMPIRICAL MODELING OF AN ECOSYSTEM

ROBERT E. ULANOWICZ, DAVID A. FLEMER *, DONALD R. HEINLE and ROGERS T. HUFF

Center for Environmental and Estuarine Studies, University of Maryland, P.O. Box 38, Solomons, Md. 20688 (U.S.A.)

* *Present address:* Office of Biological Services, Fish and Wildlife Service, U.S. Department of the Interior, Washington, D.C. 20240 (U.S.A.)

(Received 6 November 1976)

ABSTRACT

Ulanowicz, R.E., Flemer, D.A., Heinle, D.R. and Huff, R.T., 1978. The empirical modeling of an ecosystem. *Ecol. Modelling*, 4: 29—40.

The authors have endeavored to create a verified a-posteriori model of a planktonic ecosystem. Verification of an empirically derived set of first-order, quadratic differential equations proved elusive due to the sensitivity of the model system to changes in initial conditions. Efforts to verify a similarly derived set of linear differential equations were more encouraging, yielding reasonable behavior for half of the ten ecosystem compartments modeled. The well-behaved species models gave indications as to the rate-controlling processes in the ecosystem.

INTRODUCTION

Most modelers are acutely aware of the necessity to relate their constructs to quantifiable observables. This tendency in extremum has led to a school of modeling which stresses the a-posteriori aspects of systems analysis. The approach holds that one should minimize and generalize the a-priori assumptions which go into a given model. Insofar as is possible, antecedent data should determine the structural and parametric elements of the model.

One possible methodology to accomplish this end has been described in detail by Mobley (1973) and will be briefly discussed later in this paper. This method has been applied to several compartments of a planktonic system by Ulanowicz et al. (1975).

To date, however, verification of a-posteriori type models is notably lacking. Until such time as an a-posteriori model is applied with some degree of success to independent data, the models will be regarded by many as mere descriptors of data, i.e., complicated curve-fitting.

The goal of the authors is to achieve a verified a-posteriori model of a total ecosystem. The work described herein is toward this end. While a fully

verified ecosystems model cannot be claimed, the results are most encouraging. Even the difficulties incurred and the partial nature of the final model reveal interesting aspects of this particular ecosystem and, possibly, systems in general.

EXPERIMENTAL

The observation of all the major elements in a functioning ecosystem is generally an awesome task. To keep the temporal and spatial scales manageable and to insure that the same populations are sampled each time, many investigators have turned to studying a microcosm of the total system (Levandowsky, 1977).

The authors had been engaged in using the microcosm as an empirical tool to assess the effects of treated sewage upon estuarine planktonic systems. This necessitated following the time evolution of the various nutrient and planktonic species. In principle, that is the type of data required by the Mobley algorithm and a preliminary run provided the data for the earlier application (Ulanowicz et al., 1975).

The run reported upon here consisted of three replicate pairs of microcosms. One pair served as a control and the other two pairs were enriched to 0.1% and 1.0% secondarily treated waste water from an Annapolis, Md., wastewater facility. All external environmental variables were kept the same among the six microcosms. Thus, if the same biological mechanisms were operating in all six tanks, a modeler might expect a single system of equations to describe the evolution under each treatment — only the initial conditions need be changed to reflect the different nutrient allocations.

One may regard the data for the three conditions as independent. An a-posteriori model can be derived from one condition and applied to the other sets with only a change in initial conditions. The subsequent ability of the model to track the latter conditions would provide some measure of verification of the model.

The technical details of the experiment have been set forth by Ulanowicz et al. (1975). Only the pertinent elements are repeated here for completeness.

Six, 757-l polyethylene cylinders mounted in the laboratory compartment of the Research Barge ORCA contained the microcosms. A 500 W General Electric quartz-iodide wide flood lamp provided approximately 0.18 cal/m^2/min of synthetic sunlight at the surface of each tank. The tubs were lighted for 16 h of each day and were automatically stirred gently four times a day for 1-h periods. Temperature varied by about 2° C around a mean of 25° C. The salinity remained near $8^0/_{00}$ throughout the experiment.

The tanks were filled with water taken from a depth of 0.5 m below the surface and filtered through a 500-μm mesh. Natural populations of copepods, rotifers, protozoans, and algae were thereby introduced. Just prior to taking the first sample, treated sewage was added to four of the tanks so that one pair received a 1% addition and the remaining two a 0.1% addition.

The systems were sampled for the next 14 days at the same time each day. Those variables recorded on a daily basis and germane to this study included particulate carbon (PC), active chlorophyll *a* (CHLA), herbivore carbon (HERB), particulate nitrogen (PN), dissolved organic carbon (DOC), ammonia (NH_3), nitrates (NO_3), nitrites (NO_2), total phosphorus (TP), total dissolved phosphorus (TDP), and dissolved inorganic phosphorus (DIP). (DIP).

DATA TRANSFORMATION PRIOR TO MODELING

One step which necessarily preceeds a-priori is the lexical, i.e., the identification of the compartments (or state variables) to be modeled. The foregoing list of variables is centered around the elements carbon, nitrogen, and phosphorus. The resultant model should describe the cycling of these elements through the abiotic, primary producer and herbivorous trophic levels. The measured variables are not immediately suitable for such description. With a few assumptions, however, they can be recast into the desired compartments. Strictly speaking, these assumptions are an a-priori part of the final model and are described below for completeness.

In the first instance, the only clue to phytoplankton biomass is contained in the active chlorophyll *a* measurements. To arrive at a value for the carbon content of the primary producers (PP), it is necessary, in the absence of more detailed information, to make the assumption that the chemical composition of the total phytoplankton biomass does not change throughout the experiment. Thus, the primary producer carbon will be at all times proportional to the measured active chlorophyll *a* (CHLA). The constant of propor-

TABLE I

Data transformations

	Values used in model = *f*[values recorded]
(1)	PP (μg atom C/l) = 2.5 * CHLA (μg/l)
(2)	DN (μg atom N/l) = PN (μg atom N/l) $-$ 0.1132 * PP (μg atom C/l) $-$ 0.1983 * HERB (μg atom C/l)
(3)	DP (μg atom P/l) = TP (μg atom P/l) $-$ TDP (μg atom P/l) $-$ 9.43×10^{-3} * HERB (μg atom C/l)
(4)	DC (μg atom C/l) = 83.33 * PC (μg C/l) $-$ PP (μg atom C/l) $-$ HERB (μg atom C/l)
(5)	DOP (μg atom P/l) = TDP (μg atom P/l) $-$ DIP (μg atom P/l)
(6)	NO (μg atom N/l) = NO_3 (μg atom N/l) + NO_2 (μg atom N/l)

HERB, NH_3, DON, DIP used as recorded in μg atom/l.
See text for description of variables.

tionality is assumed to be 2.5 μg at C for each μg of chlorophyll a, after Strickland (1965).

In like manner, the carbon : nitrogen : phosphorous atomic ratios of the phytoplankton and herbivores are assumed to be 106 : 12 : 1 and 100 : 19.83 : 0.74 (Redfield et al., 1963; Ryther and Dunstan, 1971), respectively. These ratios then allow the estimation of the concentrations of any two of the elements in a living compartment once data on the remaining element is available. Since data are available on the carbon content of the living compartments (PP and HERB) as well as the total particulate accumulations of the three elements, the detrital compartments (DC, DN, DP) may be calculated by difference (see eqs. (2), (3), and (4) in Table I).

To complete the representation of compartments, the dissolved organic phosphorous (DOP) is estimated by the difference between the total dissolved phosphorus (TDP) and the dissolved inorganic phosphorus (DIP). Also, the nitrate (NO_3) and nitrite (NO_2) measurements are pooled as a single species (NO).

The set of transformation equations is summarized in Table I. The units of each variable are included for clarity.

THE MODELING EXERCISE

The present description of the a-posteriori method as found in the literature cited above can be paraphrased in four steps:

(1) *The assumption of what the state variables (compartments) will be.* This is usually a prerequisite for data acquisition.

(2) *The choice of a general mathematical description of all possible interactions.* If one considers the time rate of change of a given state variable to be an autonomous function of all of the k state variables,

$$\frac{dN_i}{dt} = f_i(N_1, N_2, N_3, \ldots, N_k) , \tag{1}$$

then Mobley suggests that the unspecified f_i be expanded as an algebraic series of the N's and terms greater than the quadratic be neglected, e.g.,

$$\frac{dN_i}{dt} = a_i + \sum_{j=1}^{k} b_{ij}N_j + \sum_{j=1}^{k}\sum_{m=1}^{k} c_{ijm}N_jN_m . \tag{2}$$

This compromise has the advantage of being quite general in nature, limiting the number of possible coefficients, and still including the rudiments of nonlinear behavior.

(3) *The regression of the more important terms of the model to the time series data.* This step is sometimes misunderstood. It should be emphasized that the iterative algorithm used (Greenberger and Ward, 1956) is simultaneously a sensitivity analysis and a regression scheme. This iterative method starts with all constants set equal to zero and amends the value of a single

constant each iteration, beginning with the most sensitive and continuing until the criterion of fit is achieved. It is possible to limit the regression to the n most sensitive coefficients (as determined during the course of the iteration). This allows one to begin with a field of coefficients whose number exceeds that of the data points taken, so long as regression is confined to a number of coefficients not exceeding the number of data points. This exclusion of certain terms is part of the model parsing.

(4) *Testing the significance of the remaining terms to arrive at a final reduced model.* Each term, and then groups of terms, is dropped in turn from the model. The change thereby induced in the fit is used in statistical hypothesis testing techniques to evaluate whether the term in question can be dropped from the model. The complexity of the final model is, hence, determined by the confidence levels specified in the hypotheses testing.

Two important points not fully discussed by Mobley include:

(1) *The ability of the derived model to reproduce, upon integration, the data used to fix the coefficients.* The criterion-of-fit and the hypotheses testing are both based upon the errors between the observed derivatives and their predictor values determined by prescribed combinations of coefficients and state variables. There is no guarantee that the model, when integrated, will reasonably reproduce the data from which it was derived.

(2) *The ability of the model to simulate independent data.*

Problems immediately arose with regard to the first point. Ecological data, even from controlled microcosms, tend to be noisy, and this noise is exaggerated in the derivative (the criterion vector, in this case). It was obvious that some form of data smoothing would be necessary if the higher frequency noise in the derivative were not to dominate the regression scheme. To this end, the authors employed orthogonal polynomials (Forsythe, 1957) to represent the data and its derivatives over the given time domain. Experimentation showed that fourth degree polynomials were quite adequate in representing the 15 values of the state variables and their derivatives. Integration of models derived from the smoothed data proved to be quite successful in reconstructing the input data. (It should be pointed out that the 15 data points for each species are now being represented by five polynomial coefficients.)

The authors also felt that the hypothesis testing would be more meaningful if the errors between the original data and the integrated model were used for comparison instead of those recommended by Mobley. The algorithm was accordingly rewritten.

The amended methodology yielded systems of equations which were quite efficient in simulating the generative data.

When the data (15 time series points) for each species were regressed to the generalized form (66 terms to choose from), the terms which most efficiently fit the observations were almost exclusively bilinear. This indicated that the mechanisms of interaction between the components were nonlinear in nature.

Difficulties arose, however, when the initial conditions applied to these

resultant quadratic models were changed to reflect conditions in the other microcosms. Universally, the models went unstable shortly after integration commenced. Numerous attempts at dropping very sensitive quadratic terms resulted only in new unstable configurations.

In desperation, the authors removed the nonlinear terms from consideration. That the resultant linear models fit the data with accuracy was not surprising, considering that one had as many as ten terms to choose from in fitting data summarized by five polynomial coefficients. The encouraging aspect was that, when run with different initial conditions, several of the compartment models remained stable and positive.

Further selection of terms to limit the number of parameters in each species model ensued. The primary criterion used was the relative sensitivity of the constant in the Greenberger—Ward algorithm, although in a few instances less sensitive terms were retained, preferentially to conform with biological intuition. This particular step in the modeling process needs to be made less subjective. The final reduced model described the generating data as well as the full model with at least 95% confidence in all cases, save on (NH_3). The reduced model is presented in Table II.

TABLE II

The reduced model

(1) $\frac{dPP}{dt} = -0.00279 * PP - 0.0543 * HERB + 0.1193 * NO$

(2) $\frac{dHERB}{dt} = 0.00121 * PP - 0.004459 * DN$

(3) $\frac{dDC}{dt} = 0.009364 * PP - 0.02316 * DON$

(4) $\frac{dNH_3}{dt} = -0.000199 * PP + 0.02282 * NH_3 - 0.1576 * DOP + 0.0002304 * DC$

(5) $\frac{dDON}{dt} = -0.05702 * NH_3 - 0.01045 * NO + 0.5474 * DIP + 0.00180 * DC$

(6) $\frac{dDN}{dt} = 0.001353 * PP - 0.01088 * HERB$

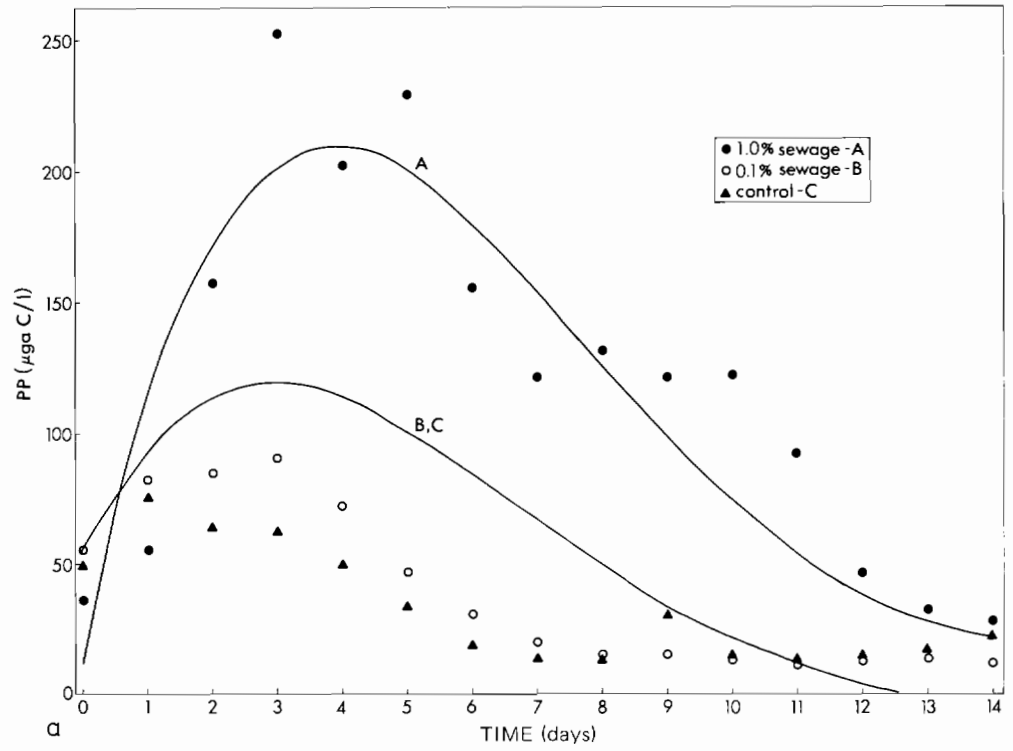
(7) $\frac{dNO}{dt} = -0.0003882 * PP - 0.01370 * NO + 0.06898 * DOP$

(8) $\frac{dDIP}{dt} = 0.002750 * NH_3 - 0.0003603 * DON - 0.0001958 * NO - 0.02340 * DIP$

(9) $\frac{dDOP}{dt} = 0.00004611 * PP - 0.0008026 * NO$

(10) $\frac{dDP}{dt} = -0.00006686 * PP + 0.0002551 * HERB + 0.0006665 * NO$

All constants in h^{-1}.
All concentrations in μg atom/l (of C, N, or P).


● 1.0% sewage - A
○ 0.1% sewage - B
▲ control - C
A
B,C
PP (µga C/l)
TIME (days)
a

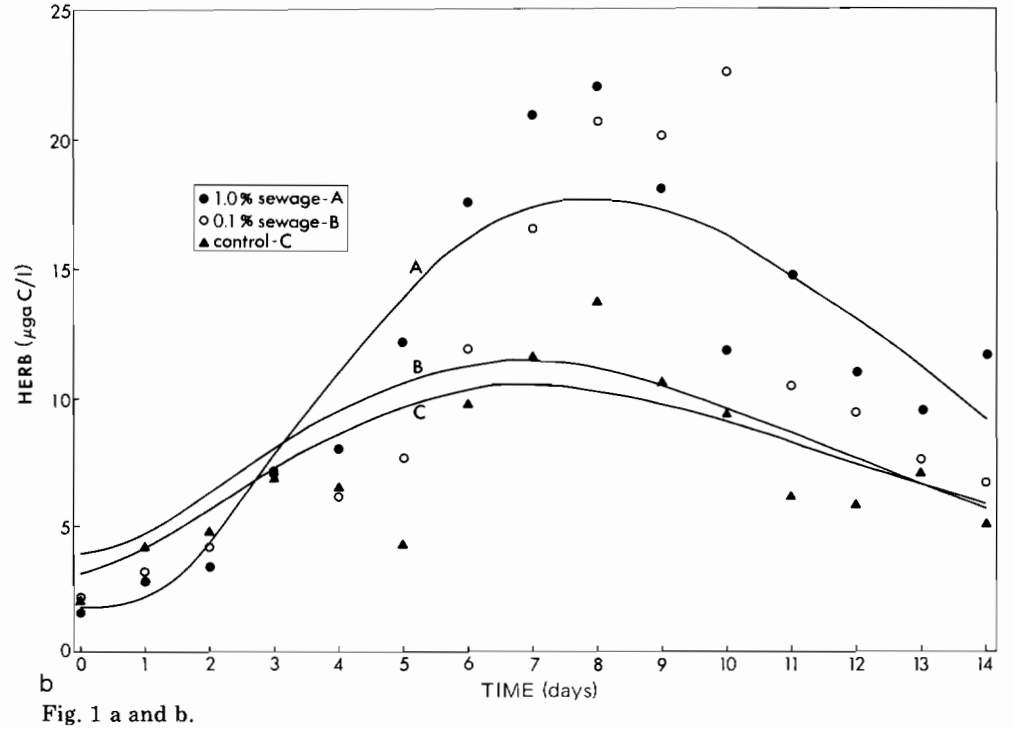
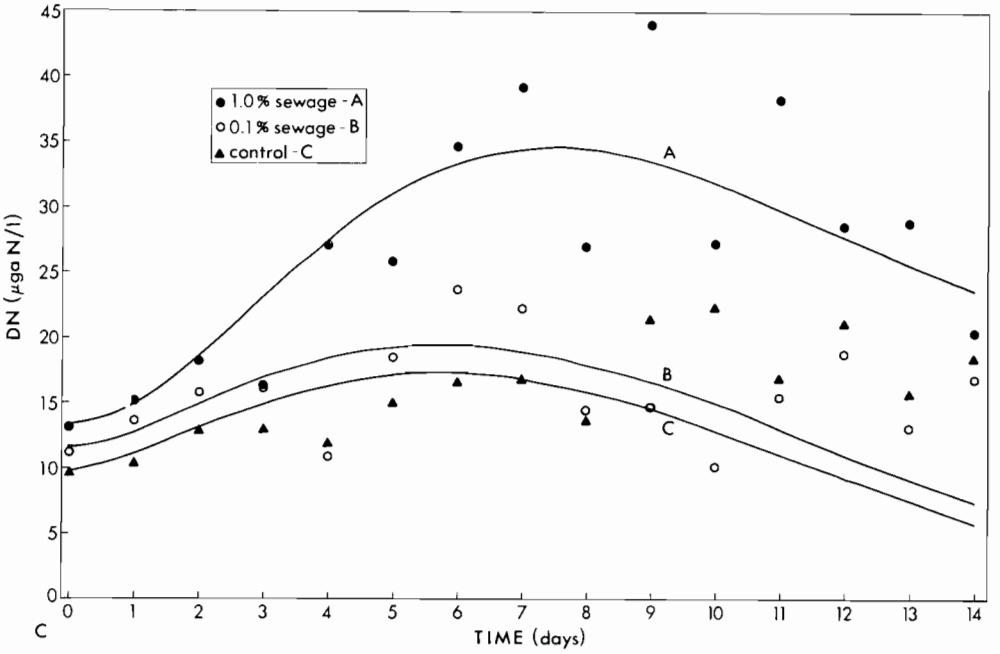
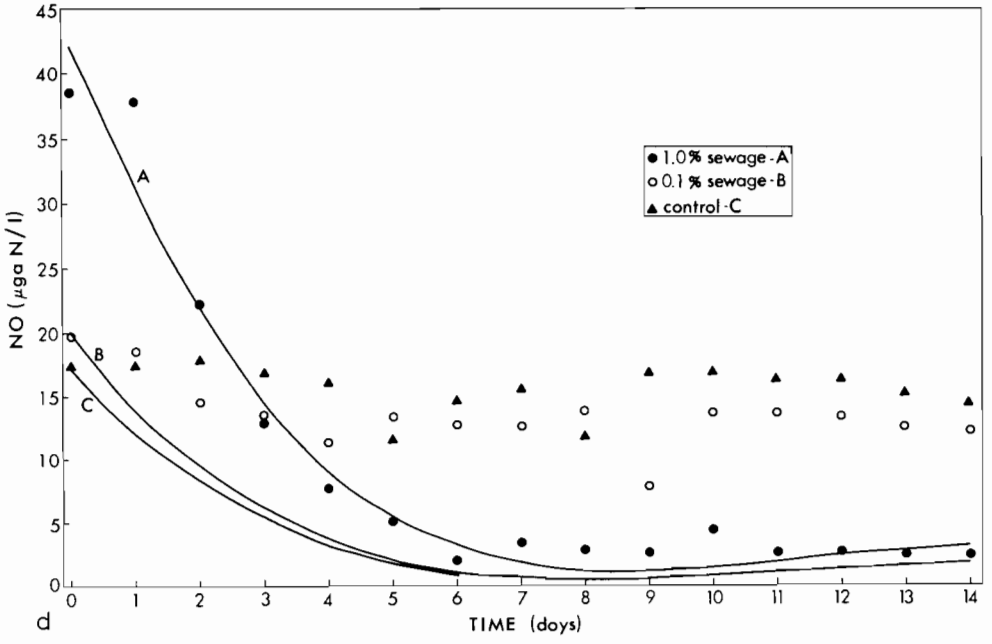

● 1.0% sewage - A
○ 0.1 % sewage - B
▲ control - C
A
B
C
HERB (µga C/l)
TIME (days)
b

Fig. 1 a and b.



c



d

Fig. 1 c and d.

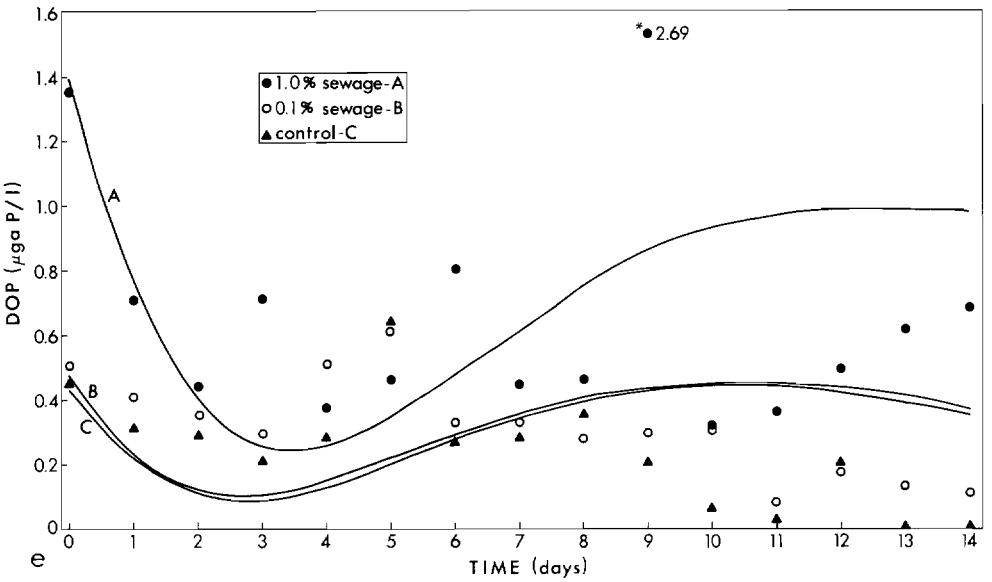


Fig. 1 (pp. 35—37). Behavior of the empirical model for various initial conditions.

Five of the eigenvalues of the reduced model possessed positive real parts, thereby introducing instability into the model. However, five of the species (PP, HERB, DN, NO, DOP) were not tightly coupled to the dominant unstable eigenvalues. In fact, perusal of Table II will reveal that these species are actually decoupled from the rest of the model, i.e. they form a self contained sub-model. The sub-model remains well-behaved over the 15-day sampling period for all three sets of initial conditions.

The predictions of the five cited species are graphed in Fig. 1. In each figure the curve A represents the integration of the model derived from the data of the 1% sewage microcosma. Curves B and C represent the behavior of the model under the initial conditions representative of 0.1% sewage and the control, respectively. The actual data from the three circumstances are pictured for comparison.

DISCUSSION

Probably the most interesting result of this endeavor is the failure of the quadratic differential equations to give reasonable results when applied to independent data. This is in contrast to the qualified success of the linear equations to do the same. B.C. Patten (personal communication, 1976) has frequently met with some success in employing linear ecological models in instances where nonlinear models have failed. These empirical observations, which parallel the independent results presented here, have led him (Patten,

1975) to hypothesize that ecological interactions on the macroscopic scale are intrinsically linear in nature.

The above results are necessary, but certainly not sufficient, to provide a proper test of Patten's hypothesis. It is still possible that other, less general, nonlinear forms would not exhibit the unstable behavior observed here. It is also possible, though not too probable, that noise in the data has precluded the selection of a stable set of bilinear terms.

There are problems, however, with pursuing other nonlinear interactions in the a-posteriori context. Primary among these is the attendant loss in generality that accompanies the choice of most nonlinear forms. The a-posteriori approach should be viewed less in the strict sense of a model and more as a tool with which data can be analyzed in the hope that repeated application to different circumstances and systems will uncover macroscopic patterns that lead to phenomenological laws. In such a search all possibilities must be kept open. Hence, the decomposition of the functions describing the interactions (right-hand side of eq. (1)) should be into terms capable of synthesizing very general behavior.

The problem incurred with the quadratic terms also points up a difficulty in using the Greenberger–Ward algorithm, or any equivalent combination of sensitivity analysis and regression. Namely, the most sensitive terms with respect to effecting a fit of the data are likely to also be the sensitive terms which induce instability in a model. This quandary does not make empirical modeling impossible, but does underscore the importance of attempting some form of verification.

The separation of the ten species into two sub-models is striking. It could be fortuitous. On the other hand, it could be symptomatic of a mismatch between the fundamental kinetics and the daily sampling interval. The species involved primarily in rapid transactions could describe spurious trends when sampled daily. These trends would give rise in turn to false and possibly destabilizing terms in their models. This process is likely related to aliasing in time series analysis and is in evidence in the ammonia model.

The derived models did not track the independent data with precision. In fact, one should not attach too much significance to the numerical magnitude of the rate constants. One may expect, however, that the qualitative behavior of the system is mirrored in the final set of constants.

To investigate these interactions it is useful to regard eqs. (1), (2), (6), (7) and (9) of Table II as a set of five, first-order linear differential equations,

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = \sum_{j=1}^{5} b_{ij} N_j \ . \tag{3}$$

The set of coefficients, b_{ij}, can be decomposed into symmetric and anti-symmetric parts according to the identity

$$b_{ij} = s_{ij} + a_{ij}, \tag{4}$$

where

$$s_{ij} = \tfrac{1}{2}(b_{ij} + b_{ji}), \tag{5a}$$

$$a_{ij} = \tfrac{1}{2}(b_{ij} - b_{ji}). \tag{5b}$$

The advantage of this device is that the anti-symmetric components can be regarded as describing "flows of causality" in analogy to predator—prey interactions, whereas the symmetric parts described mutualistic (+) or competitive (—) interactions in accordance with the sign of the component.

Fig. 2 graphically displays the flow directions according to the signs of a_{ij}. The two interesting features of the diagram are the closed loop of "causality" connecting NO, PP and DOP and the apparent "sink" of causality from PP and DN into HERB.

Fig. 3 illustrates the mutualistic (+) and competitive (—) interactions between the five species. The minus signs above the NO and PP compartments indicate the self-limitation of these species.

Two features emerge from this analysis. The first is the cycle associated with primary productivity. This cycle appears to consist of stimulating transfers damped in part by the self-limitation of two of the species. Further management of the system is provided by the herbivore compartment which drains the phytoplankton and detrital nitrogen compartments.

Obviously this is only a partial picture of the microcosm dynamics. Presumably, more data taken at shorter intervals and extending over a longer time would allow for the inclusion of other species and cycles.

An analysis of the dominant modes of the system (performed on the symmetric matrix, s_{ij}) reveals little new information beyond the fact that the model system will eventually go unstable because of the presence of two positive eigenvalues. An analysis of the eigenvectors reveals that one pair of eigenvalues (nearly equal in magnitude, but opposite in sign) is primarily associated with the production cycle. A second pair (again nearly equal in magnitude, but opposite in sign) is dominated by the herbivore compartment. The remaining eigenvalue is small in comparison with the others and is dominated by the detrital compartments.

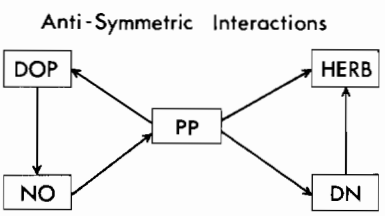


Fig. 2 (*left*). Anti-symmetric behavior of the empirical model.

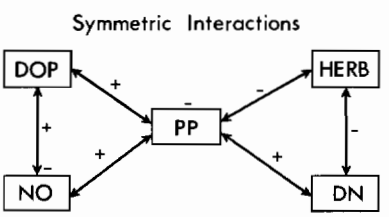


Fig. 3 (*right*). Mutualistic and competitive interactions in the empirical model.

FUTURE WORK

The results to date give encouragement for perfecting the linear modeling scheme. There are two points in the algorithm described above which unnecessarily require a-priori intervention.

The first of these involves specification of the degree of data-smoothing employed. It should be possible to use the same hypothesis testing routines to systematically determine the degree of polynomial smoothing appropriate to each time series.

Secondly, the determination of the reduced model should be automated according to objective criteria.

With these two improvements the algorithm would, for all practical purposes, be fully automated. All that would be necessary as input would be the time series data along with the desired confidence limit, and a suitably parsed linear model would result without further intervention. Such an objective process should provide a useful tool in the continuing search for phenomenological laws of ecosystem behavior.

ACKNOWLEDGEMENTS

The experimental work discussed herein was supported by the National Science Foundation, grant number GI-29906. Funds for the computational work came from the State of Maryland, Power Plant Siting Commission, project number 17-74-04. The authors would like to thank Mrs. Linton Beaven, Mrs. Shelly Sulkin, and Mr. Richard Murtagh for their assistance in chemical analyses of the microcosms and Mr. Curtis Mobley for his helpful review of the manuscript.

REFERENCES

Forsythe, G.E., 1957. Generation and use of orthogonal polynomials for data-fitting with a digital computer. J. SIAM, 5: 74—88.

Greenberger, M.H. and Ward Jr., J.H., 1956. An iterative technique for multiple correlation analysis. IBM Techn. Newsletter, No. 12, pp. 85—97.

Levandowsky, M., 1977. The use of microcosms in marine biology. In: O. Kinne (Editor), Marine Ecology, Vol. 3. Wiley, New York, N.Y., in press.

Mobley, C.D., 1973. A systematic approach to ecosystems analysis. J. Theor. Biol., 41: 119—136.

Patten, B.C., 1975. Ecosystem linearization: An evolutionary design problem. Am. Nat., 109(969): 529—539.

Redfield, A.C., Ketchum, B.H. and Richards, F.A., 1963. The influence of organisms on the composition of seawater. In: M.N. Hill (Editor), The Sea, Vol. II. Wiley-Interscience, New York, N.Y., pp. 26—77.

Ryther, J.H. and Dunstan, W.M., 1971. Nitrogen, phosphorous and eutrophication in the coastal marine environment. Science, 171: 1008—1113.

Strickland, J.D.H., 1965. Production of organic matter in the primary stage of marine food chains. In: J.P. Riley and G. Skirrow (Editors), Chemical Oceanography. Academic Press, New York, N.Y., pp. 478—610.

Ulanowicz, R.E., Flemer, D.A., Heinle, D.R. and Mobley, C.D., 1975. The *a-posteriori* aspects of estuarine modeling. In: L.E. Cronin (Editor), Estuarine Research, Vol. I. Academic Press, New York, N.Y., pp. 602—616.