View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by International Institute for Applied Systems Analysis (IIASA)

International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Uncertainty in the Parameters and Predictions of Phytoplankton Models

Di Toro, D.M. and Straten, G. van

IIASA Working Paper

WP-79-027

April 1979

Di Toro, D.M. and Straten, G. van (1979) Uncertainty in the Parameters and Predictions of Phytoplankton Models. IIASA Working Paper. WP-79-027 Copyright © 1979 by the author(s). http://pure.iiasa.ac.at/1156/

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

Working Paper

UNCERTAINTY IN THE PARAMETERS AND
PREDICTIONS OF PHYTOPLANKTON MODELS

Dominic M. Di Toro
Gerrit van Straten

April 1979
WP-79-27

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

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

UNCERTAINTY IN THE PARAMETERS AND PREDICTIONS OF PHYTOPLANKTON MODELS

Dominic M. Di Toro
Gerrit van Straten

April 1979
WP-79-27

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 the development of models for the description of river, lake, and reservoir water quality. Much of this interest has been directed towards the construction of progressively larger and more complex deterministic simulation models. Considerably less attention has been paid to the problems of uncertainty in field data, in model description and in parameter estimates and their consequences when forecasting the future behaviour of a water body. This report is the first in a series of forthcoming publications of IIASA's Resources and Environment Area task on Models dealing with these problems. It has been prepared during a three-month summer visit of the first author to the Institute. The discussions held during the work have contributed considerably to the recognition of the problem of how to cope with uncertainty in ecological modeling as an important and challenging issue for further development of the field. Should the reader have any remark, comment or criticism that could help to improve the methodology provided, or could lead to new directions, he should not hesitate in communicating his ideas to the authors. His contribution would be gratefully appreciated.

ABSTRACT

A methodology is developed to evaluate in quantitative terms the effect of uncertainty in the data and the model on the reliability of parameter estimates in phytoplankton models, and to assess the effect of the resulting parameter uncertainty on model predictions. The method of maximum likelihood is adopted as the basis of the analysis, resulting in a weighted least squares estimation problem. The analysis provides an estimate for both the weights and the model errors, where the weights appear to be determined by the data errors and the model errors simultaneously.

A preliminary application of the method is presented for a 16 state variable, 20 parameter phytoplankton model for Lake Ontario. Extensive data for 14 of the 16 state variables is used to calculate the parameter uncertainty covariance matrix and model error variances. The degree of uncertainty of parameters and their mutual cross-correlations are assessed in terms of the subjective options held by workers in the field. Also a preliminary estimate of the effects of the quantity of data available is presented. Finally, the consequences of parameter uncertainty on the prediction error are indicated. It follows that the presence of cross-correlation in the parameter set resulting from the calibration considerably mitigates the error of prediction.

UNCERTAINTY IN THE PARAMETERS AND PREDICTIONS OF PHYTOPLANKTON MODELS

D.M. Di Toro and G. van Straten

1. INTRODUCTION

The construction of phytoplankton models, or indeed any model, is aimed at increasing one's understanding of the particular system being considered and helping one to make predictions about the consequences of alterations to that system. Since most models of natural phenomena have little a priori content--hence the word "model" rather than "theory"--it is necessary that a detailed calibration be carried out in order to establish the degree of validity of the model computations. In the final analysis it is the comparison of computation to observed data that gives a model whatever credibility it has prior to validation on different data sets. Since the objectives of a modelling study, to increase understanding and allow predictions, both depend on the calibration (the former since the quantitative values of the parameters are an important part of the understanding and insight gained, and these are a direct result of the adjustments to fit data, and the latter since the credibility of the predictions is at least partly judged by the adequacy of the fit), it is clear that the quality of the calibration is an important issue.

The quantitative methods that are available for judging the quality of a model's fit to data tend to concentrate on

the statistical character of the residual sequences which are analyzed for randomness, constant variance, etc., in order to highlight any persistent model weakness. These types of investigations should be made in modelling studies for which the data is extensive enough to permit analysis. Such procedures have recently been applied to a number of different phytoplankton models of Lake Ontario [Thomann et al, 1979]. Less formal comparisons, e.g. a judgement of the adequacy of the fit, are always made, but their basis is uncertain.

It is common [e.g. Draper & Smith, 1966], in the case of linear regression models, to calculate the errors in the regression coefficients and the error of predictions for the regression in addition to the residuals' analysis. The purpose of this investigation is to adapt these methods, where possible, in order to make quantitative statements about the reliability of the model parameters and model errors in phytoplankton models.

The estimate of model error variance and parameter covariance matrix follows from an application of the method of maximum likelihood to an hypothesized error structure. The results are valid to the extent that the assumed error structure (Gaussian, independent, etc.) is correct and that the asymptotic properties of maximum likelihood estimators and their covariances apply for the number of observations available. For the application presented below, the quantity of available data is quite large and it is expected that the large sample results apply.

The estimation of model prediction error relies on a linearization of the model equations in order to calculate the model uncertainty due to parameter uncertainty. If the model response to parameter variation is reasonably linear, then the computation is reasonable. If jump discontinuities in model state variables are suspected, then these linearized methods are of little value. There is really no substitute for an intuitive understanding of the situation and the model in order to decide whether any set of methods is applicable. For phytoplankton models of the type considered in this investigation the states are certainly smooth functions of the parameters. Hence the methods to be discussed are believed to be applicable to the quantification of the parameter and prediction uncertainty.

2. FRAMEWORK

The phytoplankton model considered in this investigation is constructed from the basic mass balance equations for a natural water system. These have the form of classical continuity equations with source and sink terms that represent the physical-chemical-biological reactions that transform the materials of concern. If the water body is divided into volume elements, V'_k, then for each concentration c'_{kl} of material l, $l = 1,2,\ldots,N_c$ within that volume a mass balance equation of the form:

$$V'_k \frac{dc'_{kl}}{dt} = \sum_j E'_{jk}(c'_{jl} - c'_{kl}) + \sum_j Q'_{jk}c'_{jl} + S_l(c'_{k1},c'_{k2},\ldots;\phi'_k;\theta) + W'_{kl} \quad , \tag{1}$$

applies. The hydrodynamic transport is specified by the dispersion coefficients E'_{jk}, and the advective flows, Q'_{jk}, between adjacent volume segments. The kinetic sources and sinks, $S_l(c'_{k1},c'_{k2},\ldots;\phi'_k;\theta)$, are a function of the concentrations within the volume elements, the exogenous variables ϕ'_k that additionally characterize the elements, and the parameters of the kinetics, θ. The mass inputs, W'_{kl}, to that volume element complete the specification of the mass balance requirements. The transport coefficients are obtained either by hydrodynamic calculation or from the analysis of suitable tracer behaviour. The mass inputs and exogenous variables are measured. The values of the parameters, θ, are the object of the calibration analysis.

The solution of these equations, $c'_{kl}(t)$, is obtained by numerical integration so that the concentrations at any time t'_j are available. Let:

$$f_{kl}(t_j;\theta) = c'_{kl}(t_j) \quad , \tag{2}$$

denote a solution for any particular set of parameters, θ. The observations of variable ℓ are denoted by $c_{kl}(t_j,r_i)$ for time t_j at location r_i which is assumed to be within volume segment k. A formal statement of the calibration objective is the minimization of some measure of data and computed values' deviation by choosing parameter values, θ, within a set of accept-

able values. For example, a simple least squares criterion:

$$\min_{\text{acceptable}} \Theta \sum_{l=1}^{N_c} \sum_{k=1}^{N_V} \sum_{j=1}^{N_t} \sum_{i=1}^{N_{jkl}} [c_{kl}(t_j,r_i) - f_{kl}(t_j;\theta)]^2 \ , \qquad (3)$$

where N_c is the number of variables, N_V is the number of volume elements, N_t is the number of times at which data are available, and N_{jkl} is the number of observations within volume k at time j for concentration l. An immediate difficulty is apparent with this criterion: the results are influenced by the units chosen for the variables, since the sum is over all concentration variables as well. It is clear, therefore, that a weighting procedure is required. The method employed below supplies these weights.

3. METHODOLOGY

Since the observation $c_{kl}(t_j,r_i)$ is subject to errors, it is necessary to take these into account explicitly. That is, a statistical model of the errors is required in order to proceed with an analysis. Assume that the observed concentrations have the form:

$$c_{kl}(t_j,r_i) = f_{kl}(t_j;\theta) + \eta_{jkl} + \varepsilon_{ijkl} \ , \qquad (4)$$

where $f_{kl}(t_j;\theta)$ is the computed model mean of the l^{th} variable in segment k at time t_j; η_{jkl} is the random model error which accounts for the deviations between the computed and the true mean; ε_{ijkl} is the random measurement error and spatial deviations within the volume element k of the repeated measurements made within that volume element at t_j, which accounts for the difference between the observations and the true mean. The measurement error and the spatial heterogeneity within the volume segments are associated with this variable.

To simplify the notation, let k represent a variable l within a volume segment k', ordered in any convenient way so that k denotes a compartment, i.e., a concentration for a volume segment, and $k = 1,\ldots,N_cN_V$. Thus the concentration at position i,

at time j, for compartment k, is assumed to have the form:

$$c_{ijk} = f_k(t_j;\theta) + \eta_{jk} + \varepsilon_{ijk} \quad . \tag{5}$$

The random variables, η_{jk} and ε_{ijk} are assumed to be independent random variables with zero means and variances $\sigma^2_{\eta k}$, respectively $\sigma^2_{\varepsilon jk}$. That is:

$$\eta_{jk} = N(0,\sigma^2_{\eta k}) \quad , \tag{6}$$

$$\varepsilon_{ijk} = N(0,\sigma^2_{\varepsilon jk}) \quad . \tag{7}$$

The purpose of the estimation procedure is to make the "best" estimate of the unknowns: $\theta,\sigma^2_{\eta k},\sigma^2_{\varepsilon jk}$.

The assumption of Gaussian, independent random variables is the simplest and leads to familiar formulas. However, it is not necessary. The structure of the spatial variations, ε_{ijk}, can be analyzed directly. In particular, the covariance matrix, which is quite large, can have significant off-diagonal elements. The assumption that spatial heterogeneity has no structure across compartments assumes that:

$$\text{cov}\ (\varepsilon_{ijk},\varepsilon_{i'j'k'}) = \delta_{ii'}\delta_{jj'}\delta_{kk'}\sigma^2_{\varepsilon jk} \quad , \tag{8}$$

where $\delta_{ii'} = 0$ for $i \neq i'$ and $\delta_{ii} = 1$ for $i = i'$, etc; $\delta_{ii'}$ implies no structure within segments; $\delta_{jj'}$ implies no time correlations, and $\delta_{kk'}$ implies no compartment interaction, that is, no cross variable or cross segment interaction for the spatial heterogeneity.

Similarly, the model errors are assumed to be uncorrelated in time and across compartments, that is:

$$\text{cov}\ (\eta_{jk},\eta_{j'k'}) = \delta_{jj'}\delta_{kk'}\sigma^2_{\eta jk} \quad , \tag{9}$$

and, in addition, they are assumed to be stationary:

$$\sigma^2_{\eta jk} = \sigma^2_{\eta k} \quad , \tag{10}$$

and thereby independent of time, t_j. It should be pointed out that these assumptions can be checked either before the estimation is made, for the spatial heterogeneity variation, or after the estimation, by examining the residuals.

The estimation problem is nonlinear in the parameters and therefore quite difficult. It takes the form of a nonlinear analysis of variance of a mixed model (that is, both fixed and random effects) [Searle, 1971], which appears to be rather complex in general terms. However, in this application there is a reasonable approximation available. The number of observations, N_{jk}, for each variable, within each volume at time t_j is large relative to the number of times (ship cruises), N_t, and the number of compartments $N_c N_v$. Therefore, fixing j and k and analyzing c_{ijk} for the spatial variability independently of the model variability seems a reasonable approach.

Let $\hat{\bar{c}}_{jk}$ be the estimate of the true spatially averaged concentration, $\bar{c}_{jk}$, at time j for the compartment k and assume that $\hat{\bar{c}}_{jk}$ can be computed as the sample mean of the within segment observations:

$$\hat{\bar{c}}_{jk} = \frac{1}{N_{jk}} \sum_{i=1}^{N_{jk}} c_{ijk} \quad . \tag{11}$$

The true spatially averaged concentration is expressed by

$$\bar{c}_{jk} = f_k(t_j;0) + \eta_{jk} + \mu_{jk} \quad , \tag{12}$$

where μ_{jk} denotes the deviation from the true spatial average due to the finite number of within volume samples, and η_{jk} is the model error. The variance of μ_{jk} can be estimated from the sample variance of the individual c_{ijk}:

$$\sigma^2_{\varepsilon jk} = \frac{1}{N_{jk} - 1} \sum_{i=1}^{N_{jk}} (c_{ijk} - \hat{\bar{c}}_{jk})^2 \tag{13}$$

so that the variance of the spatial average is:

$$\sigma^2_{\mu jk} = \frac{1}{N_{jk}} \sigma^2_{\varepsilon jk} \quad , \tag{14}$$

from the properties of the sample mean. Consequently, the observed segment mean concentrations $\bar{c}_{jk}$ have mean $f_k(t_j;\Theta)$ and variance:

$$V\{\bar{c}_{jk}\} = V\{f_k(t_j;\Theta) + \eta_{jk} + \mu_{jk}\} = \sigma^2_{\eta k} + \frac{\sigma^2_{\varepsilon jk}}{N_{jk}} \quad , \qquad (15)$$

and, being the sum of independent Gaussian random variables, $\bar{c}_{jk}$ is also Gaussian:

$$\bar{c}_{jk} = N(f_k(t_j;\Theta), \sigma^2_{\eta k} + \sigma^2_{\varepsilon jk}/N_{jk}) \quad . \qquad (16)$$

The method of maximum likelihood provides the estimation equations for Θ and $\sigma^2_{\eta k}$. The log likelihood function is:

$$\log L = \sum_{jk} \{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2_{\eta k} + \sigma^2_{\varepsilon jk}/N_{jk}) - \frac{1}{2}\frac{[\bar{c}_{jk} - f_k(t_j;\Theta)]^2}{\sigma^2_{\eta k} + \sigma^2_{\varepsilon jk}/N_{jk}}\} \quad . \qquad (17)$$

To estimate Θ_α, set $\partial \log L/\partial\Theta_\alpha = 0$, i.e.:

$$\frac{\partial \log L}{\partial \Theta_\alpha} = \sum_{jk} \frac{[\bar{c}_{jk} - f_k(t_j;\Theta)]}{\sigma^2_{\eta k} + \sigma^2_{\varepsilon jk}/N_{jk}} \frac{\partial f_k(t_j;\Theta)}{\partial \Theta_\alpha} = 0 \quad , \qquad (18)$$

which is equivalent to a weighted nonlinear least squares problem [Draper & Smith, 1966] with weights:

$$w_{jk} = \frac{1}{\sqrt{\sigma^2_{\eta k} + \sigma^2_{\varepsilon jk}/N_{jk}}} \quad . \qquad (19)$$

Note that the weights are functions of the unknown model error variances, $\sigma^2_{\eta k}$. If there is no model error the weight is the reciprocal of the standard error of $\bar{c}_{jk}$. The model error variances are estimated by maximizing log L with respect to $\sigma^2_{\eta k}$, i.e.

$$\frac{\partial \log L}{\partial \sigma^2_{\eta k}} = \sum_j \{ -\frac{1}{2} \frac{1}{\sigma^2_{\eta k} + \sigma^2_{\varepsilon jk}/N_{jk}} + \frac{1}{2} \frac{[\bar{c}_{jk} - f_k(t_j;\theta)]^2}{[\sigma^2_{\eta k} + \sigma^2_{\varepsilon jk}/N_{jk}]^2} \} = 0 \ , \tag{20}$$

or

$$\sigma^2_{\eta k} = \sum_j w^4_{jk} \{ [\bar{c}_{jk} - f_k(t_j;\theta)]^2 - \sigma^2_{\varepsilon jk}/N_{jk} \} / \sum_j w^4_{jk} \ . \tag{21}$$

Since w_{jk} is a function of $\sigma^2_{\eta k}$, the equation is nonlinear. The above form is useful for a numerical solution as shown subsequently. Note, however, that if N_{jk} is large it is likely that $\sigma^2_{\eta k} >> \sigma^2_{\varepsilon jk}/N_{jk}$ so that w_{jk} is independent of j and eq (21) becomes:

$$\sigma^2_{\eta k} \simeq \frac{1}{N_t} \sum_{j=1}^{N_t} [\bar{c}_{jk} - f_k(t_j;\theta)]^2 - \frac{1}{N_t} \sum_{j=1}^{N_t} \sigma^2_{\varepsilon jk}/N_{jk} \ , \tag{22}$$

so that the model error variance is the residual variance less the average of the spatial heterogeneity and measurement error contributions. Also from eq (22) it is clear that if $\sigma^2_{\varepsilon jk}/N_{jk}$ is large then the equation for $\sigma^2_{\eta k}$ gives negative results. This is due to the assumption that $\sigma^2_{\varepsilon jk}$ can be computed independently of $\sigma^2_{\eta k}$. The role of large N_{jk} is clear in this context. Eqs. (18) and (21) form a set of nonlinear simultaneous equations for the estimation of θ and $\sigma^2_{\eta k}$.

The asymptotic covariance matrix of the maximum likelihood estimates is [Kendall & Stuart, 1963]:

$$n \ \mathrm{Cov} \ (\hat{\theta}_\alpha, \hat{\theta}_\beta) = \frac{\Delta_{\alpha\beta}}{\Delta} \ , \tag{23}$$

where:

$$\Delta = | \ E \ \{ \frac{\partial \log f}{\partial \theta_\alpha} \frac{\partial \log f}{\partial \theta_\beta} \} | \ , \tag{24}$$

and $\Delta_{\alpha\beta}$ is the $\alpha\beta$ minor determinant of the Fisher matrix. That is:

$$n \text{ cov } (\hat{\Theta}) = [E \{\frac{\partial \log f}{\partial \Theta_\alpha} \frac{\partial \log f}{\partial \Theta_\beta}\}]^{-1} \quad , \tag{25}$$

where f is the probability density function of the observations and n = the number of observations. For linear regression, the problem is cast as one observation of the vector $(\bar{c}_{11}, \bar{c}_{12}, \ldots)$ in which case the probability density function of the observations is equivalent to the likelihood function for $\bar{c}_{jk}$; that is f = L and:

$$E \{\frac{\partial \log f}{\partial \Theta_\alpha} \frac{\partial \log f}{\partial \Theta_\beta}\} = \sum_{jk} \sum_{j'k'} w_{jk}^2 \, w_{j'k'}^2 \, \frac{\partial f_k(t_j, \hat{\Theta})}{\partial \Theta_\alpha} \frac{\partial f_{k'}(t_{j'}, \hat{\Theta})}{\partial \Theta_\beta}$$

$$\cdot E \{(\bar{c}_{jk} - f_k(t_j, \hat{\Theta}))(\bar{c}_{j'k'} - f_{k'}(t_{j'}, \hat{\Theta}))\} \quad . \tag{26}$$

But $\bar{c}_{jk} - f_k(t_j; \hat{\Theta}) = N(0, \sigma_{\eta k}^2 + \sigma_{\varepsilon jk}^2 / N_{jk})$ and independent for all jk so that the expectation is just:

$$E \{(\bar{c}_{jk} - f_k(t_j, \hat{\Theta}))(\bar{c}_{j'k'} - f_{k'}(t_{j'}, \hat{\Theta}))\} = \delta_{jj'} \delta_{kk'} (\sigma_{\eta k}^2 + \sigma_{\varepsilon jk}^2 / N_{jk}) \quad , \tag{27}$$

so that:

$$E \{\frac{\partial \log f}{\partial \Theta_\alpha} \frac{\partial \log f}{\partial \Theta_\beta}\} = \sum_{jk} w_{jk}^2 \frac{\partial f_k(t_j, \hat{\Theta})}{\partial \Theta_\alpha} \frac{\partial f_k(t_j, \hat{\Theta})}{\partial \Theta_\beta} \quad , \tag{28}$$

which is the conventional nonlinear least squares expression for the variance-covariance matrix elements with the weights given by eq. (19) evaluated at the maximum likelihood estimates of Θ and $\sigma_{\eta k}^2$.

3.1. Comparison to Nonlinear Least Squares

The formulas which result from the maximum likelihood method for the estimation of Θ and its covariance matrix $V\{\Theta\}$ are:

$$\sum_{jk} w_{jk}[\bar{c}_{jk} - f_k(t_j;\Theta)] \frac{\partial f_k(t_j;\Theta)}{\partial \Theta_\alpha} = 0 \quad , \tag{29}$$

$$V\{\Theta\} = [\sum_{jk} w_{jk}^2 \frac{\partial f_k(t_j;\Theta)}{\partial \Theta_\alpha} \frac{\partial f_k(t_j;\Theta)}{\partial \Theta_\beta}]^{-1} \quad , \tag{30}$$

where the equation within the square bracket is for the α,β element of matrix. These are equivalent to equations obtained from conventional nonlinear least squares [Draper & Smith, 1966] with weighting coefficients:

$$w_{jk} = \frac{1}{\sqrt{\sigma_{\eta k}^2 + \sigma_{\varepsilon jk}^2/N_{jk}}} \quad . \tag{31}$$

Of interest is the fact that the maximum likelihood equations prescribe the proper weighting for the aggregations of differing concentration variables. The result of the proper relative weight for the N_c different concentrations is a direct consequence of the method of maximum likelihood. The weight is the reciprocal of a standard error, which includes both model error and spatial heterogeneity and measurement error variance. The result is quite appealing since large variances for a particular $\bar{c}_{jk}$, whether it is due to a large spatial heterogeneity measurement error or a large model error, causes that observation to be less heavily weighted in the computation.

The estimation formula [eq. (21)] for the model variances, $\sigma_{\eta k}^2$, is not classical in the sense of least squares since it balances the residual variance with the weights and the spatial heterogeneity and measurement error variance to make an estimate of the total variance due to the model error.

3.2. Numerical Solution Techniques

The simultaneous solution of eqs. (18) for Θ and eq. (21) for $\sigma_{\eta k}^2$ is a rather formidable numerical task. The approach

adopted is to make an initial estimate of the system variances at the starting values of the parameters, Θ_0, and then solve the nonlinear least squares problems with fixed weights using available numerical routines [Reid, 1972]. This procedure is iterated until convergence is achieved.

The model error variance eq. (21) is solved by successive substitution:

$$\sigma^2_{\eta k}(\ell + 1) = \sum_j w^4_{jk}(\ell)\{[\bar{c}_{jk} - f_k(t_j;\Theta)]^2 - \sigma^2_{\varepsilon jk}/N_{jk}\}/\sum_j w^4_{jk}(\ell) \quad , \tag{32}$$

with an initial value $\sigma^2_{\eta k}(0) = 0$. This procedure has been found to converge rapidly.

The solution of the nonlinear least squares problem requires both $f_k(t_j;\Theta)$ and $\partial f_k(t_j;\Theta)/\partial\Theta_\alpha$ at the observation times t_j. Numerical integration of the differential eqs. (1) yields the values of all concentrations in all volume elements. The Jacobians are calculated initially using a numerical approximation to the derivative:

$$\frac{\partial f_k(t_j;\Theta)}{\partial\Theta_\alpha} = \frac{f_k(t_j;\Theta + \delta\Theta_\alpha) - f_k(t_j;\Theta)}{\delta\theta_\alpha} \quad . \tag{33}$$

This requires N_Θ additional integrations, one with each parameter perturbed. Subsequent Jacobians are estimated using a rank one updating formula at each new set of parameter values during the search. Occasionally the routine will reevaluate the entire Jacobian using a central difference approximation if inaccuracies are suspected [Reid, 1972].

This method of using differences to evaluate the Jacobian, instead of an integration of the sensitivity equations, is essentially a trade-off between computer time and convenience associated with the numerical method using differences, and accuracy associated with the method of direct integration. The convenience of the numerical difference approximations was judged to be the overriding factor for these preliminary calculations.

Scaling for the problem is accomplished by defining a unitless parameter vector, ϕ, where the actual parameter value is a multiple of the nominal parameter values Θ_0. That is

$$\Theta_\alpha = \Theta_{\alpha 0} \cdot \phi_\alpha \quad , \tag{34}$$

and ϕ is the vector of parameters that is adjusted. This keeps the metric of the search space roughly the same in each parameter axis.

4. PRELIMINARY APPLICATION--LAKE ONTARIO

The seasonal phytoplankton model developed for Lake Ontario [Thomann et al, 1974] is one of a series of essentially similar models that have been applied to the three lower Great Lakes [Di Toro et al, 1977], estuaries, and streams [Di Toro et al, 1971]. These model characteristics are well understood and a large amount of experience has been accumulated during these applications. The availability of a compactly coded version [Simons, 1976] and a data set with computed spatial heterogeneity variances as well as the concentration means made Lake Ontario the ideal test case for the methodology.

The seasonal phytoplankton model state variables are: chlorophyll-a; herbivorous and carnivorous zooplankton; non-living organic nitrogen, ammonia, and nitrate; unavailable and soluble reactive phosphorus. These are computed for two layers, the epilimnion and hypolimnion, for a span of one year. The vertical transport has been calibrated using a temperature balance. The inputs of all variables have been estimated. The exogenous variables: temperature, solar radiation, photoperiod, and non-algae extinction coefficient have been specified. A complete and detailed description of the model is available [Thomann et al, 1975].

The data set used for the calculations described below are the 1972 International Field Year observations. They comprise eight survey cruises during which chlorophyll-a, herbivorous and carnivorous zooplankton biomass, total Kjeldahl nitrogen, ammonia, nitrate, total and soluble reactive phosphorus were measured in both layers. The eight variables in the two layers comprise the sixteen compartments considered. Suitable linear combinations of the computed model variables are compared to these observations.

A calibration based on an aggregate of data from previous surveys (1967-1970) has been made [Thomann et al, 1978] and this is used as the starting-point for the parameter estimate calculation. These calculations are in progress. The initial experience with the numerical solution method shows that it is not straightforward, even for small dimensional parameter

searches. However, since the principle interest is in the parameter uncertainty and prediction error evaluation, it is possible to proceed with these computations using the nominal parameter values rather than the maximum likelihood estimates. An effect of this approximation is investigated subsequently.

4.1 Estimation of Parameter Variance

The calculations of parameter covariances presented below are based on the nominal parameter values of the original calibration which are listed in Table 1. The exogenous variables are also representative of 1967-1970, rather than the 1972 data set being used. As a result, the model errors, as shown in Table 2, are larger than would be expected from a direct calibration to the data. This will inflate the parameter covariances somewhat but should not affect the general conclusions.

There are 20 parameters in the model that specify the kinetics of the phytoplankton, zooplankton, and nutrient cycles. Their nominal values and descriptions are listed in Table 1. Originally, because of computer storage restrictions, it was not possible to compute the full 20-by-20 dimensional covariance matrix at one time. Therefore, a series of ten or less parameter covariance estimates were computed and later completed with the full twenty parameter case. This can easily be done with the method employed in this investigation since the parameters which are fixed are simply left at their nominal values and treated as known constants rather than as estimated parameters which have an uncertainty.

Table 3 presents the diagonal elements of the parameter covariance matrix as coefficients of variation in percent, i.e. $100\sqrt{V(\hat{\theta}_\alpha)}/\hat{\theta}_\alpha$. The absent values in each column are held constant for that calculation and assumed to be exactly known as discussed above. As expected, comparison of the twenty parameter case with the various ten parameter cases shows that the percent of uncertainty is highest when all parameters are assumed to be estimated from the calibration, rather than known a priori.

Certain general conclusions can be drawn immediately: with the exception of the Michaelis constants, the phytoplankton

parameters are known within coefficient of variation ~ 50%. The large errors in Michaelis' constants are not unexpected in Lake Ontario since these coefficients only influence behaviour during a short period of time, and the precision with which they can be estimated from the calibration data is limited. Settling velocities for phytoplankton and the particulate fraction of nitrogen have uncertainties of ~ 80%, whereas phosphorus settling is uncertain to ~ 200%. The causes of this difference are unknown at present.

Perhaps the most striking result is the large uncertainty associated with the zooplankton kinetic coefficients. Zooplankton grazing rates are uncertain by approximately a factor of two, whereas herbivorous assimilation efficiency and respiration approach or exceed tenfold uncertainty. Although the magnitude of the uncertainty is somewhat misleading, the fact that zooplankton kinetics are the least certain, in terms of the certainty with which their kinetic parameters are known, is expected. The absolute magnitude of this zooplankton uncertainty should not be taken as having quantitative significance because of the large model error associated with the miscalibration and the substantially fewer zooplankton measurements that are available in the 1972 data set.

4.2. Parameter Covariance

The offdiagonal elements of the parameter uncertainty covariance matrix specify the degree to which uncertainty is correlated. Table 4 presents the results for the phytoplankton kinetics. Uncertainty in growth rate and respiration rate is strongly positively correlated. To understand these and other results, consider the simplified equation for phytoplankton chlorophyll, P, total inorganic nitrogen, N, and available phosphorus, p, in kinetic form:

$$\dot{P} = (K_1 r \frac{N}{K_{mN} + N} \frac{p}{K_{mp} + p} - K_2)P - \frac{w}{H}P + \cdots \quad , \qquad (35)$$

$$\dot{N} = -a_{NP}(K_1 r \frac{N}{K_{mN} + N} \frac{p}{K_{mp} + p})P \quad , \qquad (36)$$

$$P = -a_{pP}(K_1 r \frac{N}{K_{MN} + N} \frac{p}{K_{mp} + p})P \quad , \tag{37}$$

where K_1 is the growth rate, r is a light reduction factor (assumed to be known), K_2 is the respiration rate, w is the settling velocity and H is the layer depth. K_{mN} and K_{mp} are the Michaelis constants for inorganic nitrogen (N) and soluble reactive phsophorus (p). The nitrogen and phosphorus to chlorophyll ratios are a_{NP} and a_{pP}. Since the calculated quantity of chlorophyll depends on the difference between K_1 and K_2, if the actual K_1 is larger than estimated, it follows that the actual K_2 is also larger than estimated. The quantity within the parentheses can be estimated with a precision that exceeds the precision of estimation for each parameter within the expression. Therefore uncertainties are correlated in such a way as to keep the net growth rate within the range implied by the calibration data. A similar reasoning explains the positive correlation between growth rate and phosphorus Michaelis constant.

The strong negative correlations between growth rate and the nutrient stoichiometric coefficients a_{Np} and a_{pP}, follow from the nutrient equations (36) and (37). If growth rate is increased, then in order to keep nutrient uptake rate within observed bounds, the stoichiometric coefficient must decrease. The strong respiration rate-stoichiometric coefficients correlation follows from their correlations to growth rate. The strong inverse correlation between the inorganic nitrogen and phosphate Michaelis constants is a consequence of the dependence of growth and uptake on the product of Michaelis Menten expressions. In order for that term to remain constant the uncertainty must be negatively correlated.

The strong positive correlation between the nitrogen/chlorophyll and phosphorus/chlorophyll stoichiometric ratios suggests that their quotient is the quantity that is actually being estimated. That is, the nitrogen to phosphorus ratio of the population is more accurately estimated and the tendency to keep this quantity between narrow bounds explains the strong positive correlation in the uncertainties of both other ratios.

The importance of these cross correlations also becomes evident when the prediction error is considered, as shown sub-

sequently. Consider the phytoplankton growth equation. For any set of exogenous variables the population net growth rate is determined by the difference between the growth and respiration rate. The difference between two almost equal uncertain expressions is more uncertain than either expression considered individually. This observation appears to imply that the error in net growth rate will be so large that prediction becomes impossible. However, the strong correlation of uncertainties modifies this result since the presence of a positive covariance between the two terms lessens the uncertainty of the difference. That is, for only K_1 and K_2 uncertain and α_1,α_2 constants, the variance of the difference is:

$$V\{\alpha_1 K_1 - \alpha_2 K_2\} = \alpha_1^2 V\{K_1\} + \alpha_2^2 V\{K_2\} - 2\alpha_1\alpha_2 \text{ cov } (K_1,K_2). \quad (38)$$

The first two terms on the right-hand side of this equation represent the amplification of error since although the mean value of the difference is computed by a difference:

$$E\{\alpha_1 K_1 - \alpha_2 K_2\} = \alpha_1 E\{K_1\} - \alpha_2 E\{K_2\} \quad , \quad (39)$$

the first two terms of the variance are a sum of the variances. It is the third term in eq. (38) that mitigates this effect. A strong positive correlation implies that the uncertainty of the difference is not as large as it would be if the uncertainty of both terms were independent.

This tends to suggest that large parameter uncertainties per se are not a sufficient condition for the conclusion that prediction errors will be large, since the cancellation due to cross-correlated uncertainties may be quite important. However, although the prediction error for situations near that used for the calibration might not be excessive in the presence of large parameter uncertainty, as the predicted state moves into new regions of state space the fortunate cancellations may decrease with the consequent increase in prediction error.

4.3. Effect of Unobserved State Variables

The relationship between the quantity of data available and parameter uncertainty is illustrated in Table 5. Uncertainty estimates for the case of only epilimnion data available as well as both epilimnion and hypolimnion data are compared. These results are computed by withholding the hypolimnion compartments from the sums in the elements of the covariance matrix, eq. (30). The index k ranges only over the epilimnion concentrations. Parameters that affect phytoplankton growth are not greatly influenced by the lack of hypolimnion data. However, respiration uncertainty increases (62 to 110%) since it has a greater role in the hypolimnion which is below the euphotic zone. It might also be expected that phytoplankton settling velocity uncertainty would be greatly increased without hypolimnion observations whereas the results in Table 5 indicate that only a modest change, 66 to 74%, takes place. The probable reason is that at 0.1 m/day settling velocity and a 73 m depth of the hypolimnion, the effect of the settling velocity on hypolimnetic chlorophyll concentration is small during the period of stratification and, therefore, the observations do not contain much information about this parameter.

As a general rule, the magnitude of the cross-correlation coefficients decreases somewhat for only epilimnion data available. This might indicate that the prediction error will increase due to lack of cross-correlation as well as due to increased parameter uncertainty.

The effect of unobserved state variables is further illustrated in Table 6. The epilimnion observations only are used and uncertainties for phytoplankton, zooplankton and phosphorus parameters are calculated. The effect of withholding the chlorophyll and zooplankton data is shown in the second column. It appears that phytoplankton parameter uncertainties are not unduly affected. This result is rather surprising at first glance, since it suggests that it is possible to estimate phytoplankton kinetic coefficients without any direct observations of phytoplankton concentration. However, the result is due to an assumption implicit in this particular calculation. For this

example the nitrogen kinetic parameters are assumed to be known exactly; they were not included in the uncertain parameter set. Therefore using the certain nitrogen parameters and observations it appears that is is possible to deduce the phytoplankton parameters without a large increase in estimated uncertainty. It is, therefore, an artefact of this particular calculation. It is included in Table 6 as an illustration of the effect of assuming that some parameters are known exactly.

The effect of withholding phytoplankton and zooplankton observations on the zooplankton kinetics is striking. The uncertainties, which are large to begin with, increase almost tenfold without zooplankton observations.

Removing the nitrogen observation, as well as phytoplankton and zooplankton, increases the phytoplankton uncertainty by tenfold as well. Therefore it is reasonably clear that observations of these primary state variables are critical in order to reduce parameter uncertainty. It is likely that similar results would be obtained for other unobserved state variables. This suggests that model frameworks which include unobserved state variables are likely to have very large parameter uncertainties associated with these state variable kinetic coefficients.

4.4. Estimation of Model Prediction Uncertainty

The model prediction uncertainty due to parameter uncertainty is obtained from:

$$V\{f_k(t_j;\Theta)\} = \sum_{\alpha=1}^{N_\Theta} \sum_{\beta=1}^{N_\Theta} \left(\frac{\partial f_{kj}}{\partial \Theta_\alpha}\right)\left(\frac{\partial f_{kj}}{\partial \Theta_\beta}\right) \text{cov}\ (\Theta_\alpha,\Theta_\beta) \quad , \quad (40)$$

by making the assumption that the second and higher order terms in a Taylor series expansion for $f_k(t_j;\Theta + \delta\Theta)$ around the nominal point can be neglected. Therefore, the result obtained with Eq. (40) is only meaningful if the model responds linearly to changes in parameters. If this is not the case a Monte Carlo simulation, perturbing the parameters in accordance with the estimated variance-covariance structure, would be required.

Consider the results to be expected if the "prediction error" is calculated for the nominal parameter estimates and the exogenous variable values which characterize the calibration data set. Since, in this case, an estimate of the model error (which would be the prediction error if the calibration data were not available) is in fact available, a comparison of the "prediction error" for this case and the estimated model errors provides a check on both the validity of the covariance estimate and the prediction error equation. The results are shown in Table 7. The range in prediction error computed over the year (actually at the times for which there are data available) is compared to the estimated model error. In all cases the prediction error is less than the model error. For most of the variables, the maximum computed prediction error approaches the estimated model error. Since the exogenous variables and the time periods of calculation are the same for both the model error estimate and the prediction error estimate it is expected that these errors should be nearly equivalent. The lower prediction error estimates are probably due to either the fact that the nominal parameters are not the maximum likelihood estimates, or that the prediction error equation neglects higher order terms, or both. It is uncertain at present which of these two sources of errors is the most serious.

It is worthwhile noting that (40) can also be written as:

$$V\{f_k(t_j;\Theta)\} = \sum_{\alpha=1}^{N_\theta} \left(\frac{\partial f_{kj}}{\partial \Theta_\alpha}\right)^2 V\{\Theta_\alpha\} + \sum_{\substack{\alpha=1 \\ \alpha \neq \beta}}^{N_\theta} \sum_{\beta=1}^{N_\theta} \left(\frac{\partial f_{kj}}{\partial \Theta_\alpha}\right)\left(\frac{\partial f_{kj}}{\partial \Theta_\beta}\right) \text{cov}\,(\Theta_\alpha, \Theta_\beta) \quad . \quad (41)$$

Thus, the model prediction variability can be attributed to the parameter variance per se, and a contribution due to the parameter cross correlations. The latter term is negative, due to the sign relation between gradients and parameter covariances. Therefore, the total variance is lower than would be obtained if only the parameter variance were taken into account. It is of interest to investigate the extent to which the cross-correlation of the uncertainties reduces the predicted error for this case. The results are shown in Table 8, which presents the prediction errors at the times at which data are available

so that the seasonal variation can also be observed. These results clearly highlight the importance of the cross-correlations. If these are neglected and only the parameter variances are considered, the error of prediction is enormously inflated. This would lead to erroneously little confidence in the model predictions.

The seasonal variation of the prediction error is of interest. In general the relative uncertainties are highest during the growing season. This is particularly apparent for chlorophyll-a, which has the highest uncertainty in the period of zooplankton growth. Since the zooplankton growth is highly uncertain, the accuracy of the chlorophyll-a predictions drops accordingly. Also the ortho-phosphorus and ammonia uncertainties are highest in the growing season, as a result of the coupling to phyto-plankton-zooplankton dynamics. For total kjeldahl-N, nitrate-nitrite and total phosphorus a relatively high accuracy is achieved for the entire year.

5. CONCLUSIONS

The method proposed in this investigation for the analysis of parameter uncertainty and prediction error appears to be quite useful and relatively straightforward to implement. The estimation of the parameter covariance matrix for a twenty parameter-sixteen state variable seasonal phytoplankton model is direct, although it does require multiple integrations of the differential equations (three or four per parameter depending on the accuracy required for the Jacobian elements). Preliminary experience with the parameter estimation algorithm, on the other hand, has not been favorable. Therefore, the parameter covariance matrix computed using the asymptotic maximum likelihood estimate is, strictly speaking, not directly applicable. If it is assumed that the nominal parameter values are reasonably close to the maximum likelihood estimate, then it is probable that the covariance matrix is reasonably close also. In any case its utility can be seen in terms of the insights regarding the estimable parameter groups and the probable error of the individual parameters.

The prediction error equation is a linearization about the nominal parameter values. As a consequence it is also an approximation. For the case of the nominal parameters and the 1972 data set, the prediction error is calculated to be somewhat lower than the actual model error. Whether this is due to the linearization or the lack of maximum likelihood estimates is uncertain at present. Roughly speaking, and based only on limited computational experience, the method appears to produce parameter covariances and prediction errors that are somewhat lower than the actual values, but they appear to be reasonable and, therefore, in spite of the inaccuracy, useful for interpretations and, with qualifications, quantifications of parameter and prediction uncertainty.

With regard to the actual numerical values for the parameter and prediction uncertainties computed for the Lake Ontario model, it should be pointed out that both the parameter covariances and prediction errors are larger than would be the case if the

nominal parameters and exogenous variables for the original calibration (1967-1970) had been modified to more closely match the 1972 data set and exogenous variables which were used to estimate model error variances. As a consequence these larger than normal model errors tend to inflate the parameter covariances and prediction errors. Thus the actual values listed in the tables should be considered only as indicative of the behaviour to be expected from parameter uncertainty covariances and prediction errors and not of their absolute numerical values. A more complete study using a carefully calibrated model would be required to establish the probable range of parameter uncertainties and prediction errors.

Table I.A The biological system, sources (+) and sinks (-)

Process	P (phytoplankton)	Z_1 (herbivorous zooplankton)	Z_2 (carnivorous zooplankton)
Growth	$K_1(T)r \frac{C_N}{K_{mN} + C_N} \cdot \frac{C_p}{K_{mp} + C_p} P$		
Respiration	$- K_2(T)P$		
Herbivorous grazing	$- C_g(T)Z_1P$	$a_1 a_{Zp} \frac{K_{mpl}}{P + K_{mpl}} C_g(T)Z_1P$	
Herbivorous respiration		$- K_3(T)Z_1$	
Carnivorous grazing		$- C_{g2}(T)Z_2Z_1$	$+ a_{ZZ}C_{g2}(T)Z_2Z_1$
Carnivorous respiration			$- K_4(T)Z_2$

Parameter		Nominal value (at 15 C where applicable)
Phytoplankton		
Growth rate	$K_1(T)$	1.51 day^{-1}
Respiration rate	$K_2(T)$	0.07 day^{-1}
N-Michaelis constant	K_{mN}	0.025 mg N/l
PO_4-Michaelis constant	K_{mp}	0.002 mg P/l
Settling velocity	-	0.1 m/day
Zooplankton		
Carbon/Chlorophyll ratio	a_1	0.05 mg C/µg
Herbivorous grazing	$C_g(T)$	0.9 l/mg C day
Herb. assimilation eff.	a_{Zp}	0.6
Herb. respiration	$K_3(T)$	0.015 day^{-1}
Herb. grazing saturation	K_{mpl}	10.0 µg/l
Carnivorous grazing	$C_{g2}(T)$	0.9 l/mg C day
Carn. assimilation eff.	a_{ZZ}	0.6
Carn. respiration	$K_4(T)$	0.015 day^{-1}

Table I.B The nitrogen system, sources (+) and sinks (-)

Process	C_3 (Organic N)	C_4 (NH_3-N)	C_5 (NO_3-N)
Organic nitrogen-ammonia transformation	$-K_{34}(T)c_3$	$+K_{34}(T)c_3$	
Nitrification		$-K_{45}(T)c_4$	$+K_{45}(T)c_4$
Phytoplankton uptake		$-\alpha a_{NP}G_pP$	$-(1-\alpha)a_{NP}G_pP$
Phytoplankton endogenous respiration	$+a_{NP}K_2(T)P$		
Zooplankton endogenous respiration	$+a_{NZ}K_3(T)Z$		
Zooplankton excretion	$+(a_{NP}C_g(T)ZP - a_{NZ}G_ZZ)$		

Table I.C The phosphorus system, sources (+) and sinks(-)

Process	C_6 (Organic P)	C_7 (PO_4-P)
Organic phrosphorus-orthophosphorus transformation	$-K_{67}(T)c_6$	$+K_{67}(T)c_1$
Phytoplankton uptake		$-a_{pP}G_pP$
Phytoplankton endogenous respiration	$+a_{pP}K_2(T)P$	
Zooplankton endogenous respiration	$+a_{pP}K_3(T)Z$	
Zooplankton excretion	$+(a_{pP}C_g(T)ZP - a_{pZ}G_ZZ)$	

Parameter		Nominal value (at 15 C where applicable)
Nitrogen		
Nitrogen/chlorophyll ratio	a_{NP}	0.01 mgN/µg
Organic N → NH_3	$K_{34}(T)$	0.026 day^{-1}
NH_3 → NO_3	$K_{45}(T)$	0.03 day^{-1}
N settling	-	0.001 day^{-1}
Phosphorus		
Phosphorus/chlorophyll ratio	a_{pP}	0.001 mg P/µg
Unavailable P → PO_4	$K_{67}(T)$	0.105 day^{-1}
P settling	-	0.001 day^{-1}

Note: $a_{NZ} = a_{NP}/a_1$

$\alpha = c_4/(c_4+c_5)$

Table II. Model error variance at nominal parameters.

Variable	Data Average (1972) $\bar{\bar{C}}_k$	Model Std. Dev. σ_{η_k}	Coef. of Variation $\sigma_{\eta_k}/\bar{\bar{C}}_k$		Data Average Spatial Std. Dev. $\sqrt{\frac{1}{N_t}\sum_j \sigma^2_{\varepsilon_{jk}}/N_{jk}}$
EPILIMNION	(µg/l)	(µg/l)	(%)	(%)	(µg/l)
Chl a	4.2	3.7	89	6.7	0.28
TKN	180	93	52	3.8	6.9
NH_3-N	10	26	260	5.9	0.59
NO_3-N	170	75	44	2.1	3.5
T-P	20	2.0	10	1.1	0.23
PO_4-P	6.7	5.8	87	3.0	0.20
HYPOLIMNION					
Chl a	2.3	1.6	70	18.0	0.41
TKN	150	26	17	3.5	5.2
NH_3-N	9.8	38	390	5.7	0.56
NO_3-N	240	18	8	1.1	2.6
T-P	21	3.4	16	1.3	0.28
PO_4-P	11	10	91	2.5	0.27

Table III. Parameter uncertainty.

	Coefficients of Variation (%) (all data)				
Phytoplankton					
Growth rate	39	49	16	33	63
Resp. rate	69	63	44	49	81
N. Michaelis	-	240	-	380	479
PO_4 Michaelis	-	306	-	330	411
N/Chl_a	41	34	-	-	53
PO_4/Chl_a	43	40	-	-	63
Settling vel.	68	66	-	-	76
Nitrogen					
ORG.N → NH_3	29	-	-	-	31
NH_3 → NO_3	24	-	-	-	25
N Settling	72	-	-	-	85
Phosphorus					
Unavail P→ PO_4	88	88	-	-	113
P Settling	173	164	-	-	237
Zooplankton					
C/Chl_a	-	-	134	-	235
H. grazing	-	-	118	115	170
H. assim. eff.	-	-	770	250	800
H. resp.	-	-	1600	1100	1650
H. graz. sat.	-	-	1500	-	1560
C. grazing	-	-	77	62	95
C. assim. eff.	-	-	65	43	116
C. resp.	-	-	130	140	160

Table IV. Cross correlation coefficients.

	Growth	Resp	N-Mich	P-Mich	N/Chl	P/Chl	Settl	C.V.%
Growth (K_1)	1	0.82	- 0.18	0.51	- 0.78	- 0.66	0.48	63
Resp (K_2)		1	- 0.03	0.15	- 0.77	0.63	0.42	81
N-Michaelis (K_{MN})			1	- 0.86	0.34	0.17	- 0.17	480
P-Michaelis (K_{MP})				1	- 0.41	- 0.27	0.22	410
N/Chl_a					1	0.75	- 0.52	54
P/Chl_a						1	- 0.74	63
Settling (WCHL)							1	76

Table IVa. Other "strong" correlations (> 0.5).

H. grazing	(CGT)	- N-Mic	(KMN)	0.66	C. assim. eff.	(A8P)	- C/Chl_a	(CCHL)	- 0.81
H. resp.	(K4T)	- H.assim.eff.	(AZP)	0.86	PO_4/Chl_a	(PCHL)	- Unav.P → PO_4	(K67T)	0.61
H. resp.	(K4T)	- P-Mic	(KMP)	- 0.72	PO_4/Chl_a	(PCHL)	- P. settling	(K66)	0.74
C. grazing	(CG8)	- C. resp	(KBT)	- 0.65	Unav.P → PO_4	(K67T)	- P. settling	(K66)	0.77
H. assim.eff.	(AZP)	- H.graz.sat.	(KMPL)	- 0.92	P. settling	(K66)	- Settl.vel.	(WCHL)	- 0.76

Table V. Effect of data availability on parameter cross correlation matrix

	GROWTH	RESP	N - MIC	P - MIC	N/CHL	P/CHL	SETTL	OrgN → NH_3	N SETTL	C.V. %
GROWTH RATE	1	0.76 * (0.73)‡		0.61 (0.55)	-0.80 (-0.74)	-0.52 (-0.45)				49 (65)
RESP RATE		1			-0.81 (-0.91)					62 (110)
N - MICHAELIS			1	-0.83 (-0.77)						240 (250)
P - MICHAELIS				1						310 (345)
N/CHL					1	0.64 (0.64)				34 (35)
P/CHL						1	-0.74 (-0.65)	0.65 (0.54)	0.72 (0.56)	40 (48)
SETTLING							1	-0.60 (-0.52)	-0.86 (-0.88)	66 (74)
ORG.N → NH_3								1	0.66 (0.55)	88 (110)
N SETTLING									1	160 (196)

* Upper rows : all data
‡ Lower rows (between parenthesis) : surface layer data only

Table VI. Effect of missing state variable observations on parameter uncertainty

	coefficients of variation (%) (surface layer data only)		
	all variables	only nitrogen and phosphorus	only phosphorus
Phytoplankton			
Growth rate	48	97	260
Resp. rate	61	280	730
N. Michaelis	-	-	-
PO_4 Michaelis	250	360	2070
N/Chl_a	-	-	-
PO_4/Chl_a	40	60	700
Settling vel.	-	-	-
Nitrogen			
Org. $N \rightarrow NH_3$	-	-	-
$NH_3 \rightarrow NO_3$	-	-	-
N setting	-	-	-
Phosphorus			
Unavail $P \rightarrow PO_4$	110	140	260
P setting	150	300	1060
Zooplankton			
C/Chl_a	220	1200	7500
H. grazing	130	1310	4300
H. assim. eff.	880	3670	6500
H. resp.	1900	6600	11300
H. graz. sat.	1660	7680	15400
C. grazing	97	960	3500
C. assim. eff.	110	190	1800
C. resp	157	980	3800

Table VII.

Comparison of Prediction Error and Model Error

Variable k		Prediction Error Range in $\sqrt{V\{f_k(t_j,\theta)\}}$ ($\mu g/\ell$)	Model Error σ_{η_k} ($\mu g/\ell$)
Epilimnion			
1	Chl a	1.8 - 3.0	3.7
2	H. Zoop	35 - 109	112
3	C. Zoop	3 - 48	53
4	NH_3-N	5 - 21	26
5	NO_3-N	10 - 32	75
6	TKN	16 - 46	93
7	Total P	0.66 - 1.76	2.0
8	PO_4-P	1.7 - 4.7	5.8
Hypolimnion			
9	Chl a	0.2 - 0.88	1.6
12	NH_3-N	5.9 - 12.0	38
13	NO_3-N	6 - 10	18
14	TKN	10 - 13	26
15	Total P	0.67 - 1.0	3.4
16	PO_4-P	1.5 - 3.0	10

Table VIII.

Coefficients of variation (%) for model predictions due to parameter uncertainty.

σ = variance

σ_p = variance neglecting parameter covariance structure

Cruise	Chlorophyll-a		η Zooplankton		Ortho - P		Total P	
	σ	σ_p	σ	σ_p	σ	σ_p	σ	σ_p
1	42	228			18	122	3	8
2	28	126			134	807	4	15
3	32	466	172	2607	176	498	6	23
4	155	3481	50	1628	73	945	8	35
5	51	162	789	2478	140	770	12	121
6	28	175	398	1473	31	190	4	47
7	55	301			10	78	3	22
8					9	53	4	20

REFERENCES

1. Di Toro, D.M., D.J. O'Connor, and R.V. Thomann (1971) "A Dynamic Model of the Phytoplankton Population in the Sacremento-San Joaquin Delta." Adv. in Chemistry Series No. 106. Nonequilibrium Systems in Natural Water Chemistry Am. Chemical Soc., Wash., D.C., pp. 131-180.

2. Di Toro, D.M., R.V. Thomann, D.J. O'Connor and J.L. Mancini (1977) "Estuarine Phytoplankton Biomass Models - Verification Analyses and Preliminary Applications." The Sea: Ideas and Observations on Progress in the Study, ed. E.D. Goldberg, I.N. McCave, J.H. Steele; J. Wiley & Sons, Inc., N.Y., pp. 969-1020.

3. Draper, N.R., and H. Smith (1966) Applied Regression Analysis. J. Wiley & Sons, N.Y., pp. 263-299.

4. Kendall, M.G., and A. Stuart (1963) The Advanced Theory of Statistics. Hafner Publishing Co., N.Y., Vol. 1, Chapt. 17, pp. 1-49.

5. Reid, J.K. (1972) "Fortran Subroutines for the Solution of Sparse Systems of Nonlinear Equations." AERE R 7293. Atomic Energy Research Group, Harwell, Oxfordshire, U.K., pp. 1-51.

6. Searle, S.R. (1971) Linear Models. J. Wiley & Sons, N.Y. Chapt. 9, pp. 376-420.

7. Simons, T.J. (1976) Personal Communication. Canada Centre for Inland Waters, Burlington, Ont.

8. Thomann, R.V., R.P. Winfield and D.M. Di Toro (1974) "Modeling of Phytoplankton in Lake Ontario (IFYGL)." Proc. 17th Conf. Great Lakes Res. Internat. Assoc. Great Lakes Res., Ann Arbor, Mich., pp. 135-149.

9. Thomann, R.V., D.M. Di Toro, R.P. Winfield and D.J. O'Connor (1975) "Mathematical Modeling of Phytoplankton in Lake Ontario, 1. Model development and Verification." Environmental Protection Agency Ecological Research Series. EPA-660/3-75-005, pp. 1-177.

10. Thomann, R.V., R.P. Winfield and J.J. Segna (1979) "Verification Analysis of Lake Ontario and Rochester Embayment Three Dimensional Eutrophication Models." Environmental Protection Agency Ecological Research Series (in press).