Journal of Marine Research, 58, 117–164, 2000

Improving marine ecosystem models: Use of data assimilation and mesocosm experiments

by J. J. Vallino[1]

ABSTRACT

Our inability to accurately model marine food webs severely limits the prognostic capabilities of current generation marine biogeochemistry models. To address this problem we examine the use of data assimilation and mesocosm experiments to facilitate the development of food web models. The components of the data assimilation demonstrated include the construction of measurement models, the adjoint technique to obtain gradient information on the objective function, the use of parameter constraints, incorporation of discrete measurements and assessing parameter observability. We also examine the effectiveness of classic and contemporary optimization routines used in data assimilation.

A standard compartment-type food web model is employed with an emphasis on organic matter production and consumption. Mesocosm experiments designed to examine the interaction of inorganic nitrogen with organic matter provide the data used to constrain the model. Although we are able to obtain reasonable fits between the mesocosm data and food web model, the model lacks the robustness to be applicable across trophic gradients, such as those occurring in coastal environments. The robustness problem is due to inherent structural problems that render the model extremely sensitive to parameter values. Furthermore, parameters governing actual ecosystems are not constants, but rather vary as a function of environmental conditions and species abundance, which increases the sensitivity problem. We conclude by briefly discussing possible improvements in food web models and the need for rigorous comparisons between models and data (a modeling workbench) so that performance of competing models can be assessed. Such a workbench should facilitate systematic improvements in prognostic marine food web models.

1. Introduction

Aquatic food web modeling has primarily focused on conceptual models where the objective is to capture, qualitatively, the behavior of aquatic ecosystems. For example, models have been designed to examine the development of phytoplankton blooms associated with the shoaling of the mixed layer (Evans and Parslow, 1985; Sverdrup, 1953), the importance of the microbial loop (Taylor and Joint, 1990), the chaotic behavior of food webs (Beckers and Nihoul, 1995), and the "paradox of the phytoplankton" (Hutchinson, 1961; Stone, 1990). Although conceptual models have proved to be valuable for elucidating principles governing aquatic ecosystems, we must consider them qualita-

1. Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts, 02543, U.S.A., ***email: jvallino@mbl.edu***

tive because they are often only compared to limited experimental data sets, observations, or to no data at all. These models often lack the breadth to fully capture ecosystem biogeochemistry. Consequently, there has been relatively little improvement in the quantitative, predictive capabilities of aquatic food web models, even though these models have been under development for at least 50 years (Totterdell, 1993). With improvements in transport and ocean circulation models and the growing interests to assess impacts of global change, there is an increasing need to develop quantitative ecosystem models where the focus is not on concepts but rather on prognostic capabilities, efficiency and accuracy (Sarmiento *et al.*, 1993).

The difficulties in quantitative modeling of aquatic foods are well appreciated (Evans and Fasham, 1993a; Platt *et al.*, 1981), and can be categorized by three main challenges: (1) understanding and modeling of 4D complex multidisciplinary dynamics; (2) 4D data acquisition; and (3) rigorous model-data comparisons and data assimilation. Although complete ecosystem models must meet all three challenges, it is possible to decouple the modeling effort into independent components in order to facilitate overall model development. As will be demonstrated, mesocosm experiments (i.e., enclosed, experimental, aquatic ecosystems, see Grice *et al.* (1982)) can facilitate this decoupling.

Marine ecosystems are complex systems that are governed by nonlinear growth behavior of the constituent organisms, by highly dynamic predator-prey interactions between organisms, and by hydrodynamic and other external drivers that can exhibit complex patterns themselves. Hence, developing models that can accurately predict the concentrations of organisms and the rates of the biogeochemical transformations they mediate over large scales in open systems is a nontrivial task (Challenge 1). However, transport processes can be removed and decoupled from the biogeochemical processes by the use of mesocosm experiments because these systems can be designed to be well mixed. By reducing the system dimensions from 4D (i.e., space and time) to 1D (time only), greater effort can be placed on improving the ecosystem model, and this reduction also lowers computational overhead.

The large effort required to obtain comprehensive data sets is unavoidable in biological oceanography (Challenge 2). This challenge is even greater in systems with complex circulation because it is difficult to close elemental balances without either detailed flux measurements across system boundaries or a well-calibrated transport model. Indeed, often both are necessary. Although there is no substitute for field observations, again mesocosm experiments can greatly facilitate problem solution, as these systems can be intensely sampled, and only boundary fluxes across the air-water interface need be measured or accounted for. Mesocosms also allow manipulation of experimental conditions to investigate system dynamics in various regions of state space. Interestingly, while mesocosms have been used to experimentally study ecosystem processes and impact of pollutants (Grice and Reeve, 1982), only a few cases exist where they have been utilized for model development (Baretta-Bekker *et al.*, 1998). In this manuscript, we will rely on mesocosm

experiments as the source for observations. Hence, we will investigate only 1D data assimilation.

Even when observations are available, it is difficult to employ them for model improvement. As a result, aquatic food web models go largely untested and uncalibrated. Challenge 3, which addresses this difficulty, can be broken into two subareas: (a) integration of data and models (data assimilation) to improve model state estimates, model parameter estimates, or both; (b) adaptive improvements of observation strategies (which data to sample, which sensors and platforms to use, and when and where to sample) based on data and on integrated data/model simulation.

A general definition of data assimilation is the integration of models with data to improve the estimation of a system's state. The emphasis of data assimilation can be either data centric, or model centric (McLaughlin, 1995). In the fields of oceanography and meteorology, the former is often the case, while the latter is often called parameter estimation. In oceanography and meteorology, data are often abundant and the models are accurate as they are derived from first principles (conservation of mass, energy and momentum). However, because of sensitivity to initial conditions (i.e., chaos) as well as inaccurate model parameterizations, these models require frequent updates to prevent model divergence from the true state. In this case, data which are more or less randomly distributed in time and space, are interpolated to a model grid, and compared to a model forecast. By appropriate weighting of these two state assessments, an improved estimate of the true state can be generated and used as an initial condition for the next model forecast (Bergamasco *et al*., 1993; Daley, 1991; Evensen, 1994; Tziperman and Thacker, 1989; Wiggins, 1972). The Kalman filter is a classic example of model-data integration in this manner (Burger *et al*., 1998), but other methods exist, including adjoint method, inverse analysis, and optimal interpolation (Courtier *et al*., 1993; Robinson *et al*., 1998). In model-centric data assimilation, data are used to improve estimates of model parameters instead of the model state (Bennett, 1992; Evensen *et al*., 1998; Gunson *et al*., 1999; Robinson *et al*., 1998), although it is possible to achieve both objectives of state and parameter estimation simultaneously (Malanotte-Rizzoli and Tziperman, 1996). This manuscript addresses the parameter estimation component of data assimilation for marine food web models.

Minimizing parameter uncertainty is particularly important in food web models. Because marine food web models are quite sensitive to parameter values, model dynamics can change substantially with slightly different parameter values. Consequently, a poor fit between model output and observations can result from either poor model structure due to inappropriate physics or growth models, or from poor selection of model parameters. This ambiguity between structural uncertainties and parameter uncertainties makes it difficult to distinguish superiority between two or more competing models, and partially explains the plethora of models describing similar, if not the same, phenomena, such as phytoplankton growth (Fasham *et al*., 1990; Geider and Osborne, 1992, pg. 159; Taylor and Joint, 1990),

zooplankton growth (Fasham *et al.*, 1990; Franks *et al.*, 1986; Gunson *et al.*, 1999) or mortality closure terms (Steele and Henderson, 1992). If parameter uncertainty can be removed by assimilation of experimental data, then models with superior functional structure can be identified and selected for improvement systematically. Of course, we need to define a standardized experimental data set that models can be tested against for this approach to be successful (a modeling workbench (Evans and Fasham, 1993b)).

In addition to recursive model development, integration of measurements with models must also be considered (Challenge 3b). State variables used in models often do not correspond to quantities that are directly measurable in the field, which makes it difficult to use all available observations and to compare models to observations. Some measurements may encompass several state variables, while others may only represent a portion of a state variable. Particulate organic carbon (POC) is a frequently and easily obtained measurement. Yet, POC is typically not a state variable in most models because it represents carbon from both living organisms and detritus in various states of degradation. This type of model-observation mismatch is not often a significant problem, as we will show that it is usually possible to construct measurement models that relate state variables to measured variables (Robinson *et al.*, 1998). However, mapping state variables to partial observations of that variable can be problematic. For example, biomass measurements may be available for a particular species, while the model may use an aggregated compartment representing a large number of species. In this case, no obvious mapping exists between the aggregated state variable and the observed quantity without increasing the complexity of the model. These types of problems are best resolved by better integration of experimental design with model development.

In this manuscript, we use data assimilation to find the optimum set of parameter values that minimizes error between model output and observations. For observations we make use of experimental data collected from marine mesocosms. The manuscript illustrates (1) the use of mesocosm experiments to generate data for model development, (2) data assimilation using several different optimization algorithms with mesocosm data and an organic-matter-based food web model, and (3) problems with the structure of current generation food web models. Although there have been several studies focused on parameter estimation in marine food web models, the majority have used only simulated data (Crispi and Mosetti, 1993; Gunson *et al.*, 1999; Ishizaka, 1993; Lawson *et al.*, 1995; Lawson *et al.*, 1996; Marsili-Libelli, 1992) or very limited data (Fasham and Evans, 1995; Marcos and Payre, 1988; Matear, 1995; Prunet *et al.*, 1996) to constrain the model. This manuscript is unique in that it is a rigorous model-data comparison. The Appendix contains the food web model equations and descriptions of the optimization routines.

2. Data assimilation

A standard state space model is used to describe the biogeochemistry and food web dynamics of the marine ecosystem. The state variables represent those quantities of the

system that are dynamic and best approximate the system in the context of the modeling objectives. The state space model in vector form is given by,

$$\frac{d\mathbf{x}(t;\mathbf{k})}{dt} = \mathbf{f}(\mathbf{x}(t;\mathbf{k}), t; \mathbf{p}), \qquad \mathbf{x}(t_0;\mathbf{k}) = \mathbf{x}_0, \qquad \mathbf{x} \in X = \Re^m \times [t_0, t_f], \qquad \mathbf{k} \in K = \Re^m, \qquad \mathbf{f} : X \to X \tag{1}$$

where $\mathbf{x}(t;\mathbf{k})$ is an m-dimension state vector in state space X, $\mathbf{f}$ is a nonlinear vector function that describes the relationship between the state variables and their time derivatives, and $\mathbf{k}$ is an n-dimensional vector comprised of model parameters, $\mathbf{p}$, and initial state conditions, $\mathbf{x}_0$, such that $\mathbf{k}^{\mathrm{T}} = [\mathbf{p}^{\mathrm{T}}, \mathbf{x}_0^{\mathrm{T}}]$. Stochastic state models will not be considered (Miller and Cane, 1996); however, we will assume stochastic observations.

A nonlinear measurement model (Robinson *et al.*, 1998),

$$\mathbf{h}(\mathbf{x}(t;\mathbf{k}), t; \mathbf{p}) : X \to Y, \qquad \mathbf{y}(t) \in Y = \Re^{\ell} \times [t_0, t_f], \tag{2}$$

defines the mapping from the state space, X, to the observation space, Y, where $\mathbf{y}(t)$ is an ℓ-dimensional vector of observations taken at time t. Measurements taken discretely in time, as is often the case, are represented by $\mathbf{y}(t_i)$, where i ranges from 0 to $q - 1$, and q is the total number of discrete observation times. If the state and measurement models (Eqs. 1 and 2) are exact descriptions of the real process, then for the optimal parameter set, $\mathbf{k}^*$,

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t;\mathbf{k}^*), t; \mathbf{p}^*) + \mathbf{v}(t), \qquad t_0 \le t \le t_f \tag{3}$$

where $\mathbf{v}(t)$ is the measurement noise vector with zero mean and covariance $\mathbf{R}(t)$ (i.e., $E\{\mathbf{v}(t)\} = \mathbf{0}$, and $E\{\mathbf{v}(t)\mathbf{v}^{\mathrm{T}}(t)\} = \mathbf{R}(t)$, where $E\{\cdot\}$ is the expectation operator). If an observation for an element of $\mathbf{y}$, $y_j(t_i)$, is unavailable at time t_i, then $y_j(t_i)$ is set to zero and its error variance is set to infinity. The measurement model for the mesocosm experiment is given in the Appendix.

Integration of the state model (Eq. 1) generates a prediction of the state variables over time for a given set of initial conditions, $\mathbf{x}_0$, and model parameters, $\mathbf{p}$. The deterministic measurement model (Eq. 3) translates the predicted state variables to observation space so that the model output can be compared to observations, as given by the residual vector, $(\mathbf{h}(\mathbf{x}(t;\mathbf{k}), t; \mathbf{p}) - \mathbf{y}(t))$. The objective of the present data assimilation is to find the set of initial conditions and model parameters that minimizes some aspect of the residual vector, typically the sum of the squared residuals (i.e., least squares), but other objective functions are possible (Janssen and Heuberger, 1995). Formally, we seek here the set of parameters, $\mathbf{k}$, that minimize the objective (or cost) function,

$$\min_{\mathbf{k}} J(\mathbf{k}) = \int_{t_0}^{t_f} (\mathbf{h}(\mathbf{x}(t;\mathbf{k}), t; \mathbf{p}) - \mathbf{y}(t))^{\mathrm{T}} \mathbf{R}(t)^{-1} (\mathbf{h}(\mathbf{x}(t;\mathbf{k}), t; \boldsymbol{p}) - \mathbf{y}(t))\, dt, \tag{4}$$

subject to constraints imposed by Eqs. 1 and 3. Typically, covariances between measurement errors are assumed negligible, so that $\mathbf{R}(t)$ is approximated as a diagonal matrix,

where the diagonal elements are the measurement error variances (discussed below). The data assimilation problem statement, Eqs. 1-4, is also a nonlinear optimization problem.

3. Optimization

An optimum, **k***, of a scalar function occurs where the gradient of the objective function in parameter space vanishes,

$$\left.\frac{\partial J(\mathbf{k})}{\partial \mathbf{k}}\right|_{\mathbf{k}=\mathbf{k}^*} = \mathbf{0} \text{ (or } \nabla_{\mathbf{k}} J(\mathbf{k})|_{\mathbf{k}=\mathbf{k}^*} = \mathbf{0}),$$

and the Hessian matrix of $J(\mathbf{k}^*)$ at this point,

$$\left.\frac{\partial^2 J(\mathbf{k})}{\partial \mathbf{k}^2}\right|_{\mathbf{k}=\mathbf{k}^*},$$

is either positive definite (a minimum) or negative definite (a maximum). However, because the hyper-surface defined by the nonlinear function $J(\mathbf{k})$ need not be quadratic, it is possible to have more than one optimum parameter set. Considering only minima of $J(\mathbf{k})$, a minimum is said to be a global minimum if no other lesser or equal minima exist in or on the domain of interest; otherwise, it is considered a local minimum. Clearly, it is desirable to find the global minimum of $J(\mathbf{k})$, but this is a nontrivial problem because the global minimum is only a local property of the objective function. Consequently, it is often necessary to exhaustively search all parameter space to locate the global minimum, which is usually quite costly. Consequently, development of efficient global optimization algorithms is an area of intense research (Barhen *et al.*, 1997).

Optimization routines fall into two categories: those that search for local optima and those that search for global optima. The principal numerical techniques for local optimization are the methods of steepest descent, simplex, and conjugate gradient (Press *et al.*, 1986) (see the Appendix). The majority of local optimization methods for smooth objective functions employ variations of the conjugate gradient technique (Fletcher and Reeves, 1964). The basic idea of this technique is to choose a set of conjugate search directions such that a minimization conducted along one search direction does not corrupt a minimization conducted along a previous search direction. Corruption of a previous search is why methods like steepest descent are inefficient (Press *et al.*, 1986). For a quadratic objective function,

$$\overline{J}(\overline{\mathbf{k}}) = \overline{\mathbf{k}}^{\mathrm{T}}\mathbf{A}\overline{\mathbf{k}} + \mathbf{b}^{\mathrm{T}}\overline{\mathbf{k}} + c, \tag{5}$$

the optimum is found in n conjugate line-searches for an n-dimension system. Of course, the objective function defined by Eq. 4 is not quadratic. However, in a neighborhood sufficiently close to an optimum, $\mathbf{k}^{\dagger}$, the objective function can be locally approximated, via a Taylor series expansion, by the quadratic function given by Eq. 5, where the matrix $\mathbf{A}$ and vector $\mathbf{b}$ are the Hessian and gradient of $J(\mathbf{k})$, and the constant c is the objective function, all evaluated around the current search point, $\mathbf{k}^{\dagger}$, and with $\overline{\mathbf{k}}$ replaced by $(\mathbf{k} - \mathbf{k}^{\dagger})$.

The details of most conjugate gradient methods differ in how the Hessian and the gradient of the objective function are constructed and stored. Powel's method attempts to construct the conjugate search directions solely from evaluations of the objective function. Fletcher-Reeves and quasi-Newton methods construct conjugate search directions by evaluating both the objective function and its gradient. Others require information on the Hessian at any given point (Chow *et al.*, 1994).

Since knowledge of $\nabla_{\mathbf{k}} J(\mathbf{k})$ can greatly facilitate location of optima, it is often desirable to calculate $\nabla_{\mathbf{k}} J(\mathbf{k})$ for those optimization routines that can use it. However, since an analytical solution for Eq. 1 typically does not exist, the objective function given by Eq. 4 only implicitly depends on $\mathbf{k}$, so that $\nabla_{\mathbf{k}} J(\mathbf{k})$ must be determined numerically. One approach to calculate $\nabla_{\mathbf{k}} J(\mathbf{k})$ is known as the sensitivity method (see the Appendix), but this approach was found to be too computationally intensive.

A more elegant technique to compute $\nabla_{\mathbf{k}} J(\mathbf{k})$, based on variational calculus and extensively employed in optimal control theory (Kirk, 1970), is the adjoint method (Courtier *et al.*, 1993; Crispi and Mosetti, 1993; Marcos and Payre, 1988). In this method, a set of m adjoint variables, $\mathbf{a}(t)$ (also know as costate or Lagrange multipliers), are introduced, which allow the state space model constraints (Eq. 1) to be directly incorporated into an augmented objective function, $\tilde{J}$

$$\tilde{J}(\mathbf{k}) = J(\mathbf{k}) + \int_{t_0}^{t_f} \mathbf{a}(t)^{\mathrm{T}} \left(\mathbf{f}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p}) - \frac{d\mathbf{x}(t; \mathbf{k})}{dt} \right) dt. \tag{6}$$

The solution of the augmented objective function can then be found by variational calculus, taking derivatives with respect to $\mathbf{p}$, $\mathbf{x}_0$, and $\mathbf{a}$. Using Eq. 6 and conveniently defining the evolution of the adjoint variables by the following adjoint equations and boundary conditions,

$$\frac{d\mathbf{a}(t)}{dt} = -\left(\frac{\partial \mathbf{f}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p})}{\partial \mathbf{x}(t; \mathbf{k})} \right)^{\mathrm{T}} \mathbf{a}(t) - 2 \left(\frac{\partial \mathbf{h}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p})}{\partial \mathbf{x}(t; \mathbf{k})} \right)^{\mathrm{T}} \mathbf{R}(t)^{-1} (\mathbf{h}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p}) - \mathbf{y}(t)) \tag{7}$$

$$\mathbf{a}(t_f) = \mathbf{0}$$

the gradient of $\tilde{J}(\mathbf{k})$ is given by:

$$\frac{\partial \tilde{J}(\mathbf{k})}{\partial \mathbf{p}} = \int_{t_0}^{t_f} \left[\left(\frac{\partial \mathbf{f}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p})}{\partial \mathbf{p}} \right)^{\mathrm{T}} \mathbf{a}(t) + 2 \left(\frac{\partial \mathbf{h}(\mathbf{x}(t; \mathbf{k}); t; \mathbf{p}}{\partial \mathbf{p}} \right)^{\mathrm{T}} \mathbf{R}(t)^{-1} (\mathbf{h}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p}) - \mathbf{y}(t)) \right] dt \tag{8}$$

$$\frac{\partial \tilde{J}(\mathbf{k})}{\partial \mathbf{x}_0} = \mathbf{a}(t_0).$$

The solution, $\mathbf{a}(t)\ \forall\ t \in [t_0, t_f]$, is obtained by first integrating the state equation (Eq. 1) forward in time, followed by backward integration from t_f to t_0 of Eq. 7. The adjoint variables represent the sensitivity of $\tilde{J}$ with respect to $\mathbf{x}(t)$ at time t (Hall and Cacuci, 1983). Once $\mathbf{x}(t; \mathbf{k})$ and $\mathbf{a}(t)$ are determined, the gradient is obtained by integrating the right-hand

side of Eq. 8. The advantage of the adjoint method is that the additional number of differential equations to be solved (Eq. 7) over one iteration is equal to the dimension of the state space, m, which represents significant computation savings over the sensitivity method (see the Appendix). When the adjoint method could not be used to calculate the gradient, a finite difference method was employed (Prunet *et al*., 1996) (see the Appendix).

4. Implementation

In this section we describe application of nonlinear optimization for estimating parameters in a fairly complex aquatic food web model based on experimental data obtained from a mesocosm experiment.

a. Experimental setup and data

Estuaries and coastal zones receive a significant input of organic material exported from terrestrial ecosystems (Peterson *et al*., 1995; Smith and Hollibaugh, 1993; Smith and Mackenzie, 1987). In order to examine how marine food web communities process and use this dissolved organic matter (DOM), a mesocosm experiment was conducted in Woods Hole, MA, USA in which 7 m^3 seawater enclosures were augmented with combinations of dissolved inorganic nitrogen (DIN) and DOM. The enclosures were polyethylene bags equipped with flotation collars and structural hoops which were deployed in Great Harbor. The DOM in these experiments was prepared by leaching leaf litter in seawater. The mesocosm experiment consisted of four treatments: Bag A, control (no additions); Bag B, one-time addition of DOM at the start of the experiment resulting in an increase of 300 μM DOC (dissolved organic carbon); Bag C, daily additions of nitrate, phosphate, and silica equivalent to 5 μM N, 0.5 μM P, and 7 μM Si, respectively; Bag D, treatments B and C combined. All treatments also received $NaH^{13}CO_3$ as a C-tracer.

The experiment was conducted outdoors under ambient lighting (Fig. 1) from 10 Sep. to 30 Sep. 1994. Since the focus of this paper is on data assimilation techniques and aquatic food web models, experimental interpretation will not be presented here and only the following subset of the full suite of measurements will be used for data assimilation (Eq. 3): dissolved organic carbon (DOC, μM C); particulate organic carbon (POC, μM C); particulate organic nitrogen (PON, μM N); dissolved inorganic nitrogen (DIN, μM N); chlorophyll a (μg l^{-1}); net primary productivity at a specified depth (NPP, μM C d^{-1}); bacterial productivity (BP, μM C d^{-1}); light extinction coefficient (K, m^{-1}). Measurements were taken daily or every other day.

b. Food web model

The food web model (Eq. 1) used to capture the dynamics of the mesocosm experiment (Fig. 2, and see the Appendix) consists of the following ten state variables: autotrophs ($A(t)$, μM C), heterotrophs ($Z(t)$, μM C), dissolved inorganic nitrogen ($N(t)$, μM N),

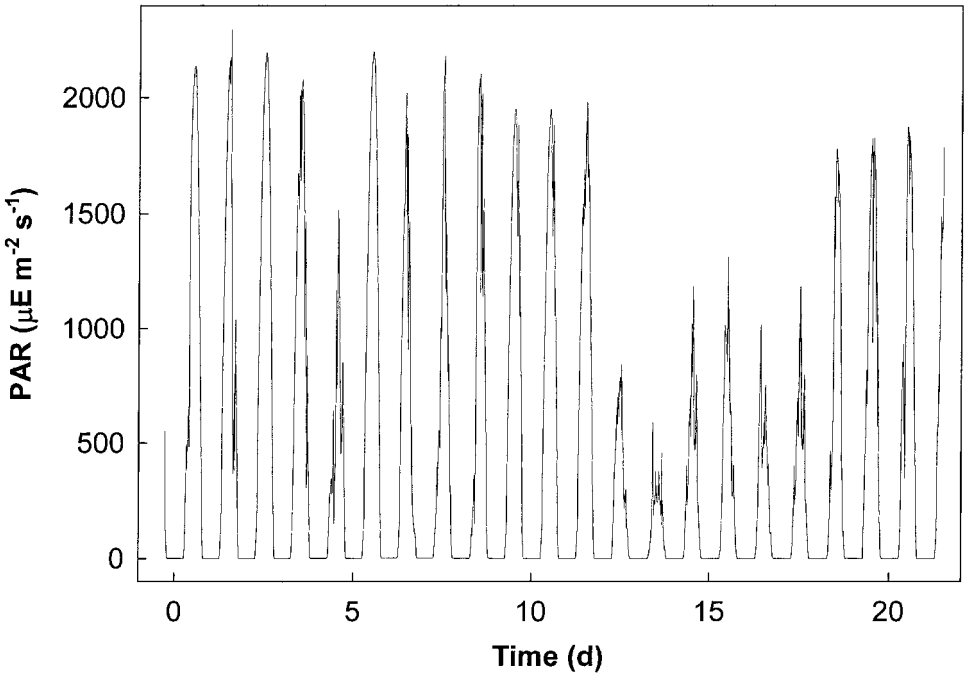

Figure 1. Photosynthetic active radiation (PAR) measured during the course of the mesocosm experiments. Note, due to temporary failure of equipment, data for day 14 were reconstructed from other PAR measurements.

dissolved labile organic C ($O_{CL}(t)$, μM C), dissolved labile organic N ($O_{NL}(t)$, μM N), dissolved refractory organic C ($O_{CR}(t)$, μM C), dissolved refractory organic N ($O_{NR}(t)$, μM N), detrital C ($D_C(t)$, μM C), detrital N ($D_N(t)$, μM N), and bacteria ($B(t)$, μM C). In this model C and N are linked where the biotic compartments (i.e., $A(t)$, $Z(t)$, and $B(t)$) are assumed to have fixed C:N ratios, while the C:N ratios of the nonliving organic compartments are free to vary. For processes, bacteria utilize labile organic matter pools and can either remineralize organic N or immobilize DIN depending on the C:N ratio of the labile DOM. The aggregated heterotrophs pool graze both the bacteria and autotrophs and remineralize N. The autotrophs compete with bacteria for DIN, and excrete both labile and refractory DOM. Decomposition of detrital material into DOM and decomposition of refractory DOM into labile DOM are governed by first order kinetics. The overall model is similar to other aggregated, Monod-based growth models (Evans and Parslow, 1985; Fasham *et al.*, 1990; Kremer and Nixon, 1978; Moloney and Field, 1991; Moran *et al.*, 1988; Pace *et al.*, 1984), except that a significant emphasis has been placed on modeling organic matter production and consumption, which is consistent with the goals of the mesocosm experiment (see the Appendix).

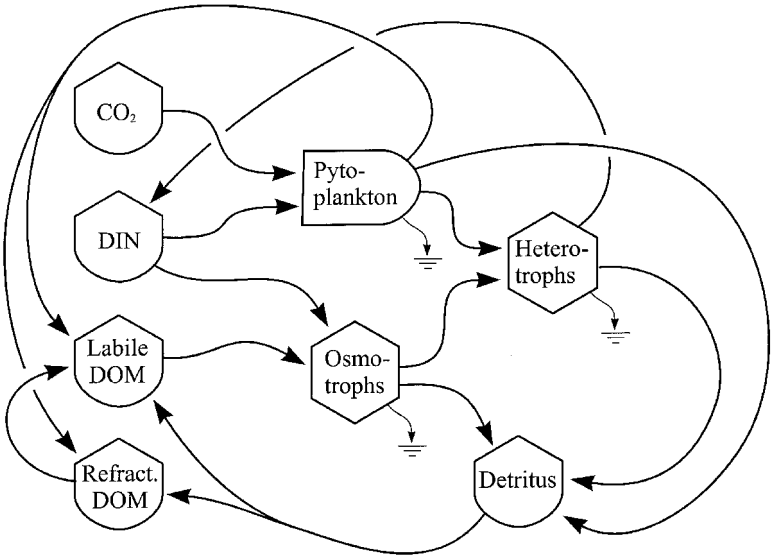

Figure 2. Diagram of food web model used to describe dynamics of the mesocosm experiment. See text and the Appendix for model description.

c. Parameters, scaling and bounds

There are 29 parameters governing the growth kinetics of the biotic compartments and the decomposition of the organic matter pools (Table 1, and see the Appendix). In addition, 10 parameters specify the initial conditions of the state variables, which brings the total adjustable parameter count to 39. Three of the 39 parameters were directly measured, so they were held constant for all optimization runs: light attenuation at the air-sea interface (η_I), mesocosm bag depth (h), and initial DIN concentration ($N(t_0)$). The remaining 36 parameters define the vector $\mathbf{k}$, which are the control variables that the optimization routines manipulate to minimize the objective function (Eq. 4).

Many optimization routines assume an unbounded parameter space, K; however, some parameter values, $\mathbf{k}$, selected from parameter space K will produce an unstable state-space model (Eq. 1) that cannot be numerically integrated. Consequently, it is necessary to place at least simple upper and lower bounds on the parameters, of the form $\mathbf{k}^L \leq \mathbf{k} \leq \mathbf{k}^U$, to insure a mass-conserving state-space model (Table 1). These bounds were crudely chosen to keep parameters greater than or equal to zero and less than 10 times their typical maximum values (Moloney and Field, 1991).

There are two main techniques to impose bounds on $\mathbf{k}$ for optimization routines that do not directly support them. A penalty function, $\Omega(\mathbf{k})$, of the form,

$$\Omega(\mathbf{k}) = \sum_{i=1}^{n} \beta_i \left(\frac{1}{k_i^U - k_i} + \frac{1}{k_i - k_i^L} \right), \tag{9}$$

can be added to the objective function, where β_i is an appropriately chosen scaling factor (Wang and Luus, 1980). The augmented objective, $J'(\mathbf{k})$, then becomes:

$$J'(\mathbf{k}) = \begin{cases} J(\mathbf{k}) + \Omega(\mathbf{k}) & \text{for} \quad \mathbf{k} \in K_B = \{\mathbf{k} | k_i^L < k_i < k_i^U, \quad \forall i \in \{1, \ldots, n\}\} \\ \infty & \text{for} \quad \mathbf{k} \notin K_B \end{cases} \tag{10}$$

The penalty function approach has the disadvantage that the gradient of $J(\mathbf{k})$ also requires modification. This approach caused convergence problems with some of the optimization routines, so its use was discontinued. Instead, we implement a second approach which employs a mapping function from an unbounded space, K', into the bounded space, $K_B \subset K$, defined by $\mathbf{k}^L \leq \mathbf{k} \leq \mathbf{k}^U$. One such transform is the sin-squared function (Box, 1966) given here,

$$k_j = k_j^L + (k_j^U - k_j^L) \sin^2 (k_j'); \qquad j = 1, \ldots, n; \qquad \mathbf{k}' \in K'. \tag{11}$$

Consequently, the optimization routine operates in K'-space, which is unbounded but is always transformed back into the bounded K_B-space via Eq. 11 prior to integration of the state-space model. As most of the optimization routines require an initial guess for $\mathbf{k}'^{(0)}$, this was obtained by applying the inverse transform to $\mathbf{k}^{(0)}$ as follows:

$$k_j'^{(0)} = \arcsin\left(\sqrt{\frac{k_j^{(0)} - k_j^L}{k_j^U - k_j^L}} \right); \qquad j = 1, \ldots, n \tag{12}$$

Note, that the inverse transform need only be used once during the initial call to the optimization routine. Since the optimization routine operates in K'-space, but the gradient is calculated via Eq. 8 in K_B-space, the gradient must be transformed into K'-space as follows:

$$\frac{\partial J(\mathbf{k}')}{\partial \mathbf{k}'} = \frac{\partial \mathbf{k}}{\partial \mathbf{k}'} \frac{\partial J(\mathbf{k})}{\partial \mathbf{k}}. \tag{13}$$

For the sin-squared transform, the $\partial \mathbf{k}/\partial \mathbf{k}'$ matrix is diagonal and its elements are given by

$$\frac{\partial k_j}{\partial k_j'} = 2(k_j^U - k_j^L) \sin (k_j') \cos (k_j'); \qquad j = 1, \ldots, n \tag{14}$$

Optimization routines, especially those based on the conjugate gradient technique, require the parameters to be scaled so that their magnitude is of $O(1)$; otherwise, the routines will often fail. One of the beneficial consequences of the sin-squared transform (Eq. 11) is that the parameters in K'-space are also scaled to $O(1)$ by the transform. For those optimization routines that internally manage simple bounds so that the sin-squared transform was not required, parameters were scaled by a simple linear transform,

$$k_j' = \frac{k_j - k_j^L}{k_j^U - k_j^L}; \qquad j = 1, \ldots, n, \tag{15}$$

and the gradient appropriately scaled (Eq. 13).

Table 1. Parameters governing the growth dynamics of the mesocosm food web model described in the Appendix. Also given are the lower and upper bounds on parameters, the initial guess, $\mathbf{k}^{(0)}$, and the best solution, $\mathbf{k}^*$ (see text).

Parameter	Description	$k_i^L - k_i^U$	$\mathbf{k}^{(0)}$	$\mathbf{k}^*$
φ_B^M	Maximum specific uptake rate of O_{CL} by bacteria	0.01–50	5.0	49.9 d^{-1}
k_{OB}	Half saturation constant for O_{CL} consumption by B	0.01–100	1.0	48.8 μM
ξ_B^M	Maximum growth efficiency of bacteria	0.01–1	0.70	0.804
ρ_Z	C:N ratio of heterotrophs	3–8	6.6	4.72 at.
ξ_Z	Growth efficiency of heterotrophs	0.01–1	0.5	0.151
φ_Z^M	Maximum specific feeding rate of heterotrophs	0.1–10	1.0	3.20 d^{-1}
k_Z	Half saturation constant for A and B consumption by Z	0.1–200	10.0	200 μM
m_Z	Maximum mortality rate of heterotrophs	0–10	0.1	0.033 d^{-1}
φ_A^M	Maximum specific fixation rate of autotrophs	0.1–20	2.0	5.43 d^{-1}
k_{NA}	Half saturation constant for N uptake by A	0.1–50	1.0	0.101 μM
ξ_A	Growth efficiency of autotrophs	0.1–1	0.8	0.998
f_{EA}	Fraction of net production excreted	0–1	0.1	0.564
ρ_A	C:N ratio of autotrophs	4–20	6.6	10.8 at.
ρ_{EA}	C:N ratio of exudate	$3–10^5$	10.0	43200 at.
m_A	Maximum mortality rate of autotrophs	0–10	0.1	0.674 d^{-1}
α	P-I slope	10^{-3}–0.09	0.0054	0.0890 m^2 s d^{-1} μE^{-1}
k_w	Light extinction coefficient of water	0.1–10	0.35	0.935 m^{-1}
k_p	Light extinction coefficient of POC	10^{-5}–1	0.003	0.00428 m^{-1} μM C
η_I	Light attenuation at sea-air interface	0.5–1	0.731	0.731
h	Depth of mesocosm bag	1.8–2.5	2.0	2.0 m
f_{LEA}	Fraction of exudate that is labile	0–1	0.8	0.999
f_{DL}	Fraction of detritus that is labile	0–1	0.8	0.331
d_{DL}	Decomposition rate of detritus	0–50	0.1	49.6 d^{-1}
d_{RL}	Decomposition rate of O_{CR} and O_{NR}	0–0.5	0.001	0.128 d^{-1}
k_{NB}	Half saturation constant of N uptake by B	0.01–50	1.0	49.1 μM
c_{chla}	Carbon to chlorophyll a ratio	0.1–10	4.2	3.76 μmol C (μg chl a)$^{-1}$
k_d	Light extinction coefficient of O_{CR}	10^{-6}–1	0.0025	1.58×10^{-5} m^{-1} μM C
ρ_B	C:N ratio of bacteria	3–7	4.5	3.57 at.
m_B	Maximum mortality rate of bacteria	0–50	0.1	48.4 d^{-1}

Table 1. (Continued)

Parameter	Description	$k_i^L - k_i^U$	$\mathbf{k}^{(0)}$	$\mathbf{k}^*$
$I_o^+(t)$	Light intensity at surface of water (driver var.)	NA	(Fig. 1)	(Fig. 1) μE m^{-2} s^{-1}
$A(t_0)$	Initial autotrophs concentration	0.1–10	5.0	4.29 μM C
$Z(t_0)$	Initial zooplankton concentration	0.1–30	1.0	28.5 μM C
$N(t_0)$	Initial DIN concentration	20–100	46.5	46.5 μM N
$O_{CL}(t_0)$	Initial labile DOC concentration	1–300	100.0	101 μM C
$O_{NL}(t_0)$	Initial labile DON concentration	10^{-3}–20	10.0	16.8 μM N
$O_{CR}(t_0)$	Initial refractory DOC concentration	100–500	350.0	359 μM C
$O_{NR}(t_0)$	Initial refractory DON concentration	0.1–100	0.35	0.158 μM N
$D_C(t_0)$	Initial detrital carbon concentration	10^{-3}–50	5.0	2.35 μM C
$D_N(t_0)$	Initial detrital nitrogen concentration	0.01–10	0.10	0.834 μM N
$B(t_0)$	Initial bacteria concentration	0.1–15	1.0	0.102 μM C

d. Discrete data and measurement model

The definition of $J(\mathbf{k})$ (Eq. 4) and the derivation of the adjoint and gradient equations (Eqs. 7 and 8) assume that the measurements are continuous in time; however, actual measurements are inherently discrete in time. This is usually not a problem, since it is common practice to interpolate a smooth function such as a cubic spline through the data, thereby generating a measurement function continuous in time (Eq. 3). This approach is valid provided the frequency of measurements exceeds the inherent dynamic frequency of the observed quantity. For instance, interpolating daily observations of phytoplankton concentration would be a valid approximation of the true function. If the frequency of the observed quantity is greater than the frequency of measurements, then simple interpolation may not be a good approximation. Certainly, such a case arises for daily observations of net primary production, bacterial production, and perhaps DOM concentration.

Although a discrete-time representation of the state model (Eq. 1), adjoint (Eq. 7) and gradient (Eq. 8) equations could be used (Gunson *et al.*, 1999; Lawson *et al.*, 1995; Thacker and Long, 1988), this would require linearization of the state model in order to calculate the state transition matrix, $\mathbf{\Phi}$, at sample times t_i, where $\mathbf{\Phi}(t_i, t_{i+1}) : \mathbf{x}(t_i) \rightarrow \mathbf{x}(t_{i+1})$. However, this approach is likely to be numerically unstable, since ecosystem models are highly nonlinear and sample times occur infrequently. Artificial sample times could be introduced between actual times, but this would result in high computational overhead. Consequently, a continuous-time model was used for the state and adjoint equations with the following modification of the residual vector.

In order to use low frequency measurements of highly dynamic quantities, we implemented a time-dependent weighting function on the residuals. Inspection of Eqs. 4, 7 and 8 reveals that complete knowledge of $\mathbf{y}(t)$ is not required, because the expressions only

depend on the weighted residual vector, $\mathbf{R}^{-1/2}(t)(\mathbf{h}(\mathbf{x}(t), t) - \mathbf{y}(t))$, where the covariance weighting matrix $\mathbf{R}(t)$, defined by

$$\mathrm{E}[(\mathbf{h}(\mathbf{x}(t)) - \mathbf{y}(t))^{\mathrm{T}}(\mathbf{h}(\mathbf{x}(t)) - \mathbf{y}(t))] = \mathbf{R}(t) = \begin{bmatrix} \sigma_1^2(t) & & 0 \\ & \ddots & \\ 0 & & \sigma_l^2(t) \end{bmatrix}, \tag{16}$$

represents instrument uncertainty and noise associated with measurements. However, we do not know $\mathbf{y}(t) \forall t$. Only when $t = t_k$ is $y(t_k)$ known. Although we can express the continuous residual function as

$$\mathbf{R}^{-1/2}(t)(\mathbf{h}(\mathbf{x}(t), t) - \mathbf{y}(t)) = \mathbf{R}^{-1/2}(t_k)(\mathbf{h}(\mathbf{x}(t), t) - \mathbf{y}(t_k))\delta_{t_k t}, \tag{17}$$

where $\delta_{t_k t}$ is the Kronecker delta, this leads to numerical integration problems due to the discontinuity introduced by $\delta_{t_k t}$. Instead, we associate with $\mathbf{y}(t)$ an uncertainty that increases beyond the instrument precision when t does not correspond to an observation time, t_i. The measurement error model hence varies with time. We therefore represent the uncertainty in $\mathbf{y}(t)$ as a product of the measurement uncertainty, $\sigma^m(t)$, and a time dilation uncertainty about an observation, $\Gamma(t)$, as follows:

$$\sigma_j(t) = \sigma_j^m(t)\Gamma(t); \qquad j = 1, \ldots, \ell. \tag{18}$$

Since $\sigma^m(t)$ depends on $\mathbf{y}(t)$, a continuous-time function is produced from the discrete observations, $\mathbf{y}(t_i)$, using the following piecewise continuous zero-order interpolating function (Fig. 3a):

$$\mathbf{y}(t) = \mathbf{y}(t_i) \qquad \text{for} \qquad \tfrac{1}{2}(t_{i-1} + t_i) < t \leq \tfrac{1}{2}(t_i + t_{i+1}); \qquad i = 0, \ldots, q - 1. \tag{19}$$

From $\mathbf{y}(t)$, the continuous-time, measurement-uncertainty vector is defined as

$$\sigma_j^m(t) = \sigma_j^r |y_j(t)| + \sigma_j^a; \qquad j = 1, \ldots, \ell, \tag{20}$$

where σ_j^r and σ_j^a are the relative and absolute standard deviations of measurement j, respectively (Table 2). The time dilation uncertainty function is modeled as a series of inverse Gaussian functions centered about each observation time, t_i

$$\Gamma(t) = \sum_{i=0}^{q-1} \exp\left(\frac{t - t_i}{\tau}\right)^2 \tag{21}$$

where τ specifies the time-bandwidth over which an observation, $\mathbf{y}(t)$, is allowed to influence the residual (Fig. 3b). The Gaussian function used for $\Gamma(t)$ has the advantage of being smooth and continuously differentiable, in contrast to the Kronecker delta (Eq. 17).

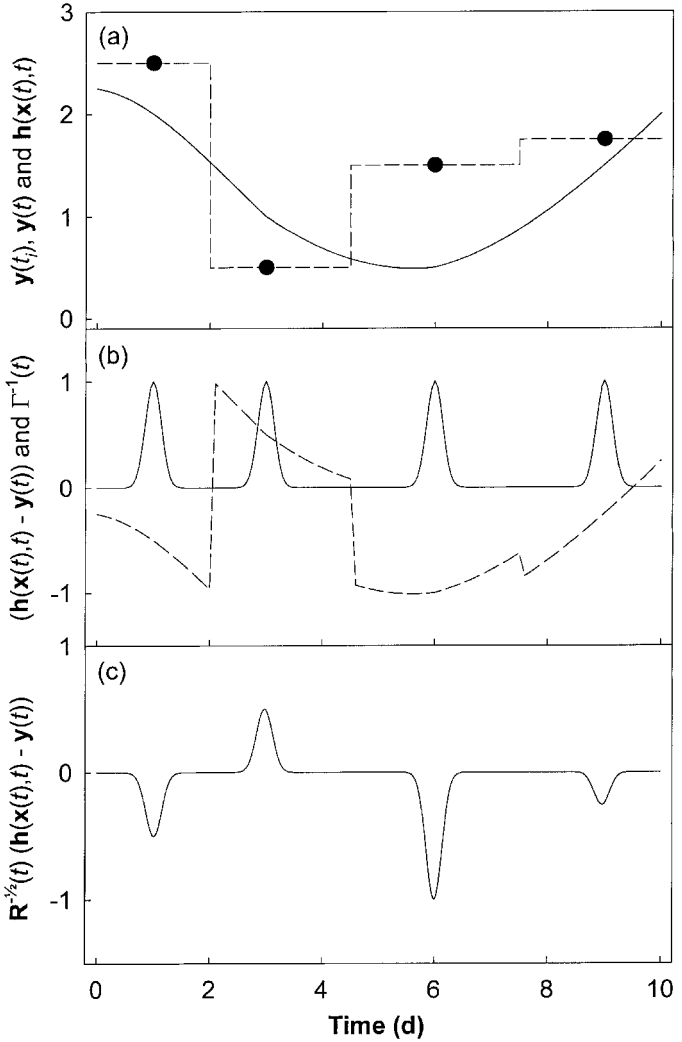

Figure 3. Example of generating a continuous weighted residual. (a) Model output, $\mathbf{h}(\mathbf{x}(t), t)$ (solid line), and observations ($\mathbf{y}(t_i)$, filled circles) connected by piecewise continuous function ($\mathbf{y}(t)$, Eq. 19, dashed line). (b) Unweighted residual function ($\mathbf{h}(\mathbf{x}(t), t) - \mathbf{y}(t)$, dashed line) and inverse of the time-dilation uncertainty function ($\Gamma^{-1}(t)$, Eq. 21, solid line). (c) Weighted residual function, $\mathbf{R}(t)^{-1}(\mathbf{h}(\mathbf{x}(t), t) - \mathbf{y}(t))$. For this example τ in Eq. 21 was set to 0.2 d and a measurement uncertainty, σ^m, of 1.0 was used.

The value of τ is based on the time before and after a given observation for which the datum is expected to be valid. For the mesocosm data, we used a value of 0.05 d for τ for all optimization runs. An example of constructing the weighted residual vector from discrete measurements is illustrated in Figure 3.

Table 2. Relative and absolute standard errors associated with the measured variables, $\mathbf{y}(t_i)$ for the mesocosm experiment. See Eq. 20.

Measurement $\mathbf{y}(t_i)$	Rel. Error σ^r	Abs. Error σ^a
$D_{OC}(t)$ (μM C)	0.03	5.0
$P_{OC}(t)$ (μM C)	0.05	0.5
$P_{ON}(t)$ (μM N)	0.05	0.1
$D_{IN}(t)$ (μM N)	0.02	0.5
$G(t)$ (μg l^{-1})	0.05	0.1
$N_{PP}(t)$ (μM C d^{-1})	0.05	1.0
$B_P(t)$ (μM C d^{-1})	0.10	0.5
$K(t)$ (m^{-1})	0.05	0.1

e. Parameter observability and model underdeterminess

A parameter is observable if a statistically significant change in its value leads to a statistically significant change in the objective function (Eq. 4). Since a parameter that is not observable cannot affect the objective function, nonobservable parameters should be removed from the set of adjustable parameters, as their values cannot be determined from the available measurements. One means to determine the observability of parameters is to examine the sensitivity of $J(\mathbf{k})$ with respect to $\mathbf{k}$ (e.g., $\nabla_{\mathbf{k}} J(\mathbf{k})$) (Fasham *et al.*, 1990; Prunet *et al.*, 1996). For a quadratic objective function, a parameter, k_j, for which $\nabla_{k_j} J(\mathbf{k})$ equals 0 for $\mathbf{k} \neq \mathbf{k}^*$ is considered nonobservable. However, for nonquadratic objective functions, $\nabla_{k_j} J(\mathbf{k})$ can equal zero over some regions of $\{k_j | k_j^L \leq k_j \leq k_j^U\}$ and be nonzero over other regions (see Section 4c.). That is, a parameter may be locally nonobservable, but still be globally observable. For a parameter to be globally nonobservable, $\nabla_{k_j} J(\mathbf{k})$ must equal 0 $\forall k_j \in \{k_j | k_j^L \leq k_j \leq k_j^U\}$. Consequently, it is necessary to evaluate $\nabla_{\mathbf{k}} J(\mathbf{k})$ at several different locations in parameter space to identify nonobservable parameters, k_j, for which $\nabla_{k_j} J(\mathbf{k}) = 0 \; \forall k_j \in \{k_j | k_j^L \leq k_j \leq k_j^U\}$ holds. Since an exhaustive search is impractical for this crude analysis, we compared values of J at a local minimum, $\mathbf{k}^*$, to J evaluated at the upper and lower bounds on $\mathbf{k}$, as given by

$$\frac{J(\mathbf{k}^* + \Delta k_j^U \mathbf{e}_j) - J(\mathbf{k}^*)}{J(\mathbf{k}^*)} \quad \text{and} \quad \frac{J(\mathbf{k}^* - \Delta k_j^L \mathbf{e}_j) - J(\mathbf{k}^*)}{J(\mathbf{k}^*)}; \qquad j = 1, \ldots, n, \quad (22)$$

where $\Delta k_j^U = k_j^U - k_j^*$, $\Delta k_j^L = k_j^* - k_j^L$, and $\mathbf{e}_j$ is an elementary vector with 1 at element j and zeros elsewhere.

The above description is a special case of nonobservability. It is possible that a combination of parameters are dependent, so that $\nabla_{\mathbf{k}_j} J(\mathbf{k}) \neq 0$ for individual parameters, but $\nabla_{\mathbf{Mk}} J(\mathbf{k}) = 0$ for some combination of parameters, $\mathbf{Mk}$. For example, consider an observable parameter, k_1, that is replaced by $(k_2 + k_3)$; neither k_2 nor k_3 can be uniquely determined. In this case, the system is not completely observable (Jazwinski, 1970, pg. 231) and is also referred to as underdetermined. That is, there are more unknown

parameters than constraining independent equations. For linear, discrete systems such as $\mathbf{y}(t) = \mathbf{A}(t)\mathbf{k}$, the dimensions and rank of $\mathbf{A}$ govern whether the system is exactly-, over- or under-determined (cf. Noble and Daniel, 1977). For nonlinear systems, no simple methods exist to determine whether the parameters can be uniquely defined from the observations, $\mathbf{y}(t)$, and the constraints imposed by the minimization problem (Eq. 4). However, we can examine whether the parameters are uniquely defined at a given optimum point, $\mathbf{k}^*$, by expanding the objective function in the neighborhood sufficiently close to the optimum, $\mathbf{k}^\dagger$, which gives

$$J(\mathbf{k}) = J(\mathbf{k}^\dagger) + \left.\frac{\partial J^{\mathrm{T}}}{\partial \mathbf{k}}\right|_{\mathbf{k}^\dagger} (\mathbf{k} - \mathbf{k}^\dagger) + \frac{1}{2}(\mathbf{k} - \mathbf{k}^\dagger)^{\mathrm{T}} \left.\frac{\partial^2 J}{\partial \mathbf{k}^2}\right|_{\mathbf{k}^\dagger} (\mathbf{k} - \mathbf{k}^\dagger) + O((\mathbf{k} - \mathbf{k}^\dagger)^3). \quad (23)$$

Taking the derivative of this expansion with respect to $\mathbf{k}$ and dropping the higher order terms produces

$$\frac{\partial J}{\partial \mathbf{k}} \approx \left.\frac{\partial J}{\partial \mathbf{k}}\right|_{\mathbf{k}^\dagger} + \left.\frac{\partial^2 J}{\partial \mathbf{k}^2}\right|_{\mathbf{k}^\dagger} (\mathbf{k} - \mathbf{k}^\dagger). \quad (24)$$

Since

$$\frac{\partial J}{\partial \mathbf{k}} = 0$$

at an optimum, the right-hand side of Eq. 24 can be solved for $\mathbf{k}^*$ to yield

$$\mathbf{k}^* = -\left(\left.\frac{\partial^2 J}{\partial \mathbf{k}^2}\right|_{\mathbf{k}^\dagger}\right)^{-1} \left.\frac{\partial J}{\partial \mathbf{k}}\right|_{\mathbf{k}^\dagger} + \mathbf{k}^\dagger, \quad (25)$$

provided the Hessian matrix,

$$\left.\frac{\partial^2 J}{\partial \mathbf{k}^2}\right|_{\mathbf{k}^\dagger},$$

is nonsingular. Note, that if the system is linear, then Eq. 25 reduces to

$$\mathbf{k}^* = \left(\int_{t_0}^{t_f} \mathbf{A}^{\mathrm{T}}(t)\mathbf{R}^{-1}\mathbf{A}(t)\, dt\right)^{-1} \int_{t_0}^{t_f} \mathbf{A}^{\mathrm{T}}(t)\mathbf{R}^{-1}\mathbf{y}(t)\, dt \quad (26)$$

which is the standard weighted least-squares solution to Eq. 4 for $\mathbf{h}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p}) = \mathbf{A}(t)\mathbf{k}$. For discrete-time systems, the integrals of Eq. 26 are dropped since the time dependency is embedded in the vector and matrix elements. Consequently, by evaluating the rank of the Hessian matrix at a minimum, $\mathbf{k}^*$, we can assess whether the system is underdetermined at that point. Furthermore, if the Hessian is nonsingular, its inverse gives the covariance

matrix for the model parameters (Matear, 1995), assuming the higher order nonlinear terms in the Taylor expansion (Eq. 23) are small.

To numerically estimate the Hessian of $J(\mathbf{k})$ at a point $\mathbf{k}^* \in K_B$, the gradient of J was calculated via Eqs. 7 and 8 at $\mathbf{k}^* \pm \Delta\mathbf{k}$, then central differences were used to obtain the second order derivatives, as given by

$$\left.\frac{J^2(\mathbf{k})}{\partial k_i \partial k_j}\right|_{\mathbf{k}^*} = \frac{1}{2\Delta k_i}\left(\frac{\partial J(\mathbf{k} + \Delta k_i \mathbf{e}_i)}{\partial k_j} - \frac{\partial J(\mathbf{k} - \Delta k_i \mathbf{e}_i)}{\partial k_j}\right)\Bigg|_{\mathbf{k}^*} \tag{27}$$

where Δk_j is given by Eq. A4 in the Appendix. To improve numerical stability in subsequent analyses, the Hessian matrix was normalized as follows:

$$\tilde{\mathbf{H}} = (k_i^U - k_i^L)\frac{J^2(\mathbf{k})}{\partial k_i \partial k_j}(k_j^U - k_j^L) \tag{28}$$

To examine the condition of $\tilde{\mathbf{H}}$, a singular value decomposition (SVD) was employed, so that $\tilde{\mathbf{H}}$ could be expanded as

$$\tilde{\mathbf{H}} = \mathbf{U}\Lambda\mathbf{V}^{\mathrm{T}} \tag{29}$$

where the diagonal matrix, Λ, contains the singular values of $\tilde{\mathbf{H}}$. A matrix is considered computationally singular (rank deficient) if the ratio of the maximum to minimum singular values (also know as the matrix condition number) exceeds the inverse of the computer's machine precision, which in this case is approximately 10^{-16} for a PC using double precision arithmetic. However, the condition number of $\tilde{\mathbf{H}}$ compared to machine precision is overly optimistic, as the precision of the measurements is typically far less than machine precision. Consequently, singular values of $\tilde{\mathbf{H}}$ are considered significant if

$$\frac{\lambda_j}{\lambda_1} > 10^{-\xi}, \qquad j = 1, \ldots, n \tag{30}$$

where λ_i are the singular values and ξ is the number of significant digits associated with the measurements (Tziperman and Thacker, 1989).

If the Hessian matrix is singular as determined by Eq. 30, then several techniques exist to identify those parameters that are uniquely defined by the measurements from those that are not. Sensitivity analysis (i.e., Eq. 22 and (Fasham *et al.*, 1990)) does identify those parameters that are not well defined, but does not provide information on correlation between parameters. Consequently, sensitivity analysis cannot detect linearly dependent parameters at $\mathbf{k}^*$. To obtain this information, the correlation matrix can be calculated from the Hessian (Matear, 1995). However, this approach cannot be used if $\tilde{\mathbf{H}}$ is computationally singular, and we found the correlation matrix only marginally useful for large dimensional problems. Another approach is to calculate the resolution matrix of $\tilde{\mathbf{H}}$ (Tziperman and Thacker, 1989; Wiggins, 1972). If we rewrite Eq. 24 as $\tilde{\mathbf{H}}\mathbf{k} = \mathbf{z}$ and make use of the SVD

of $\tilde{\mathbf{H}}$ (Eq. 29), then it is easy to show from the properties of the SVD that

$$\mathbf{V}\mathbf{V}^{\mathrm{T}}\mathbf{k} = \mathbf{V}\Lambda^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{z}, \tag{31}$$

where $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ is the resolution matrix. Because $\mathbf{V}$ is an orthonormal matrix, when $\tilde{\mathbf{H}}$ is full rank (nonsingular) $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ is the identity matrix, so that the parameters, $\mathbf{k}$, are all uniquely defined (or resolved). However, when $\tilde{\mathbf{H}}$ is singular, Eq. 31 shows that the magnitudes of the diagonal terms of $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ define the extent to which each parameter is resolved, and the magnitude of the off-diagonal terms identify those parameters that cannot be well resolved. Specifically, each row vector of $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ (or column vector, since $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ is symmetric) identifies the extent of linear dependency between parameters. It should be noted that the above analysis does not apply globally and is only valid in the neighborhood of the minimum about which the objective function is linearized, Eq. 23.

f. Optimization routines

Twelve different algorithms (Table 3) were employed to solve the nonlinear optimization problem (Eqs. 1, 3 and 4), four of which attempt to locate the global optimum. These algorithms are described in the Appendix along with details on the numerical computations.

5. Results

We first present the parameter observability analysis in order to determine which parameters or initial conditions cannot be resolved and should be removed from the adjustable parameter set, $\mathbf{k}$. In Section 5b, the results of the data assimilation are presented for each of the 12 optimization routines tested (Table 3). Performance of the optimization routines is only for the assimilation of the DOM + DIN mesocosm data (Bag D).

a. Parameter observability

Sensitivity analysis (Eq. 22) applied to the food web model shows that all parameters and initial conditions (Table 1) can significantly influence the objective function (Eq. 4). The only parameters which did not produce significant change ($>100\%$) in $J(\mathbf{k})$ (Eq. 22) were $D_C(t_0)$ and $D_N(t_0)$, which produced only 4% and 6% changes, respectively. However, caution must be used when applying this crude test, as dependent parameters will not be detected. Parameter dependency can only be determined at a given optimum.

The maximum and minimum singular values of the Hessian of $J(\mathbf{k})$ calculated from Eq. 29 at the optimum located by the simulated annealing algorithm (see below) were found to be approximately 10^8 and 10^{-2}, respectively, which gives a condition number of 10^{10}. Although the Hessian condition number is much smaller than the inverse of the machine precision, 10^{16}, Eq. 30 indicates that ten singular values are of questionable significance given a measurement precision of five digits. Examination of all the singular values of $\tilde{\mathbf{H}}$ (Fig. 4) shows that the last five or six singular values deviate from the general

Table 3. Optimization routines used for data assimilation in the mesocosm model. See the Appendix for descriptions and references.

Routine Name	Algorithm	Optima Search	Gradient Required	Constraints Allowed*	Notes
SA	Simulated annealing	Global	No	SB	
GLOBAL	Quasi-Newton with stochastic searching.	Global	No	SB	1
DN2FB	Adaptive Newton with trust region.	Local	Yes	SB	2, 3
PRAXIS	Powel's Conjugate gradient w/restarting.	Local	No	No	
DNLS1	Levenberg-Marquardt	Local	Yes	No	2, 3
SUBPLEX	Modified simplex	Local	No	SB	
BBVSCG	Quasi-Newton and Conjugate gradient	Local	Yes	No	
VE08	Quasi-Newton	Local	Yes	SB	
TN	Truncated Newton	Local	Yes	SB	4
SIGMA	Stochastic differential equations	Global	No	SB	5
GA	Genetic algorithm	Global	No	SB	
TENSOR	Tensor method	Local	Yes	No	6

*SB: Simple upper and lower bounds.

1. Quasi-Newton local search is not bounded, so sin-transform (Eq. 11) was employed.
2. Uses internal finite difference code to calculate gradient.
3. Uses vector objective function of residuals (Eq. A5), so adjoint method (Eqs. 7 and 8) not used.
4. Finite difference calculation of Hessian caused bounds to be violated, so used sin-transform (Eq. 11) to implement bounds.
5. Obtained better performance using sin-transform (Eq. 11) than codes SB constraints.
6. Uses internal finite differences to calculate Hessian matrix.

trend, which is usually an indication that these singular values may not be significant. To examine how the last six singular values ($i = 31, \ldots, 36$) affect the observability of parameters, we calculated the resolution matrix, $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ (Eq. 31), with the last six column vectors of $\mathbf{V}$ removed (corresponding to the six smallest singular values). The diagonal elements of the resulting resolution matrix (Table 4) show that parameters ρ_{EA}, d_{DL}, $D_C(t_0)$, and $D_N(t_0)$ are not well resolved, and theoretically could be replaced by linear combinations of the other 32 parameters thereby reducing the dimension of $\mathbf{k}$ by four. The diagonal elements of the resolution matrix (Table 4) can also be interpreted to mean that the objective function is insensitive to the values of ρ_{EA}, d_{DL}, $D_C(t_0)$, and $D_N(t_0)$ in the neighborhood of the SA minimum. Of course, this analysis is only locally valid. Indeed, the observability analysis from Eq. 22, which defines a more global perturbation, only identifies $D_C(t_0)$ and $D_N(t_0)$ as being poorly resolved globally. Since $D_C(t_0)$ and $D_N(t_0)$ are not particularly important parameters (they are only initial conditions), no attempt was made to reduce the number of adjustable model parameters for subsequent analyses.

Table 4. Diagonal elements of the resolution matrix, $\mathbf{V}\mathbf{V}^{\mathrm{T}}$, (Eq. 31) with the six smallest singular values removed, so that $\mathbf{V} \in \Re^{36\times30}$.

Param.	Resolution	Param.	Resolution	Param.	Resolution
ρ_{EA}	0.0000014	k_{OB}	0.903	m_A	0.998
d_{DL}	0.007	φ_Z^M	0.929	φ_A^M	0.999
$D_C(t_0)$	0.009	m_B	0.944	α	0.999
$D_N(t_0)$	0.021	$A(t_0)$	0.952	D_{RL}	0.999
k_{NB}	0.592	ξ_B^M	0.977	f_{EA}	0.999
$Z(t_0)$	0.778	ξ_A	0.981	K_{NA}	0.999
$O_{NL}(t_0)$	0.798	ρ_A	0.984	k_w	1.000
ρ_B	0.800	c_{chla}	0.990	$B(t_0)$	1.000
k_Z	0.808	$O_{CL}(t_0)$	0.992	ξ_Z	1.000
$O_{NR}(t_0)$	0.817	f_{DL}	0.993	m_Z	1.000
ρ_Z	0.852	$O_{CR}(t_0)$	0.994	k_p	1.000
φ_B^M	0.890	f_{LEA}	0.995	k_d	1.000

b. Optimization routines

The simulated annealing routine, SA, located the minimum with the smallest objective function value of 170 (Table 5). However, location of this minimum came at considerable expense, requiring 350,000 function evaluations and 253 hr of CPU time (133 MHz Intel Pentium). The three local search routines, DN2FB, PRAXIS, and DNLS1 did almost as well as SA, but required only a fraction of the number of function evaluations and CPU time. The two global routines, SIGMA and GA, did rather poorly, both consuming large amounts of CPU time and locating minima with relatively large final costs. TENSOR also did poorly, which is a result of having to numerically calculate the Hessian. The other

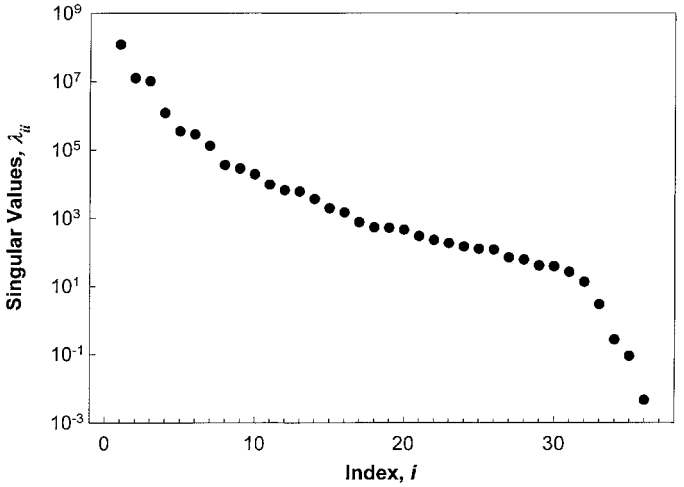

Figure 4. Singular values of Hessian matrix evaluated at the minimum located by the simulated annealing algorithm.

Table 5. Computation requirements and value of objective function associated with the minimum found by each of the optimizations routines. The initial objective function value was 39430.

Routine Name	Function Calls	Gradient Calls	CPU Time* (hr)	Final Cost $J(t_f)$	Notes
SA	350000	—	253	170	1
GLOBAL	181273	—	347	204	2
DN2FB	3537	NG	7.77	237	3, 4
PRAXIS	8455	—	6.65	248	
DNLS1	566	NG	1.15	258	3
SUBPLEX	6946	—	5.1	292	
BBVSCG	169	169	0.74	337	4
VE08	241	241	1.35	345	4
TN	539	539	2.13	471	4
SIGMA	179422	—	485	546	
GA	200020	—	321	577	1
TENSOR	57902	82	278	693	4

*CPU: 133 MHz Intel Pentium.
1. Terminated by iteration limit.
2. Located 19 other minima.
3. Used numerical gradient (NG) since routines use vector objective function (see Eq. A5).
4. Convergence error associated with minimum located.

routines fell in between the two extremes (Table 5). We should emphasize here that many of the optimization routines tested require several parameters, specific to the particular routine, to be specified by the user. Unless obviously incorrect, we used the recommended default parameter values for the optimization routines. Without extensive use of the routine, it is difficult to have a good intuitive understanding for the ideal choice of algorithmic parameters. Consequently, it is likely that performance of several of the routines could be improved by judicious selection of the algorithmic parameters.

Comparing the parameters associated with the minimum found by each optimization routine (Fig. 5) reveals that each minimum (Table 5) is a different local optimum. Indeed, the twelve optima found appear to span the entire bounded parameter space, K_B, and several of the parameters in an optimal set lie on the boundary of K_B. The presence of such a large number of vastly different optima is just one of the problems associated with current generation food web models, as will be discussed below.

A comparison between the measurements obtained from the mesocosm experiment and the model output based on the optimum solution found by the SA routine (Table 1), shows good fits for DIN (Fig. 6d) and Chl-*a* (Fig. 6e), while the fits for the other variables could stand improvement, especially POC and PON (Figs. 6b and 6c). Model-data fits of phytoplankton and bacterial productivity (Figs. 6f and 6g, respectively) are fairly good, but are difficult to interpret due to their highly dynamic nature. The model output based on the initial parameter guess is illustrated as the dashed line (Fig. 6).

Although PAR (Fig. 1) and the daily addition of nitrate introduce oscillations to the

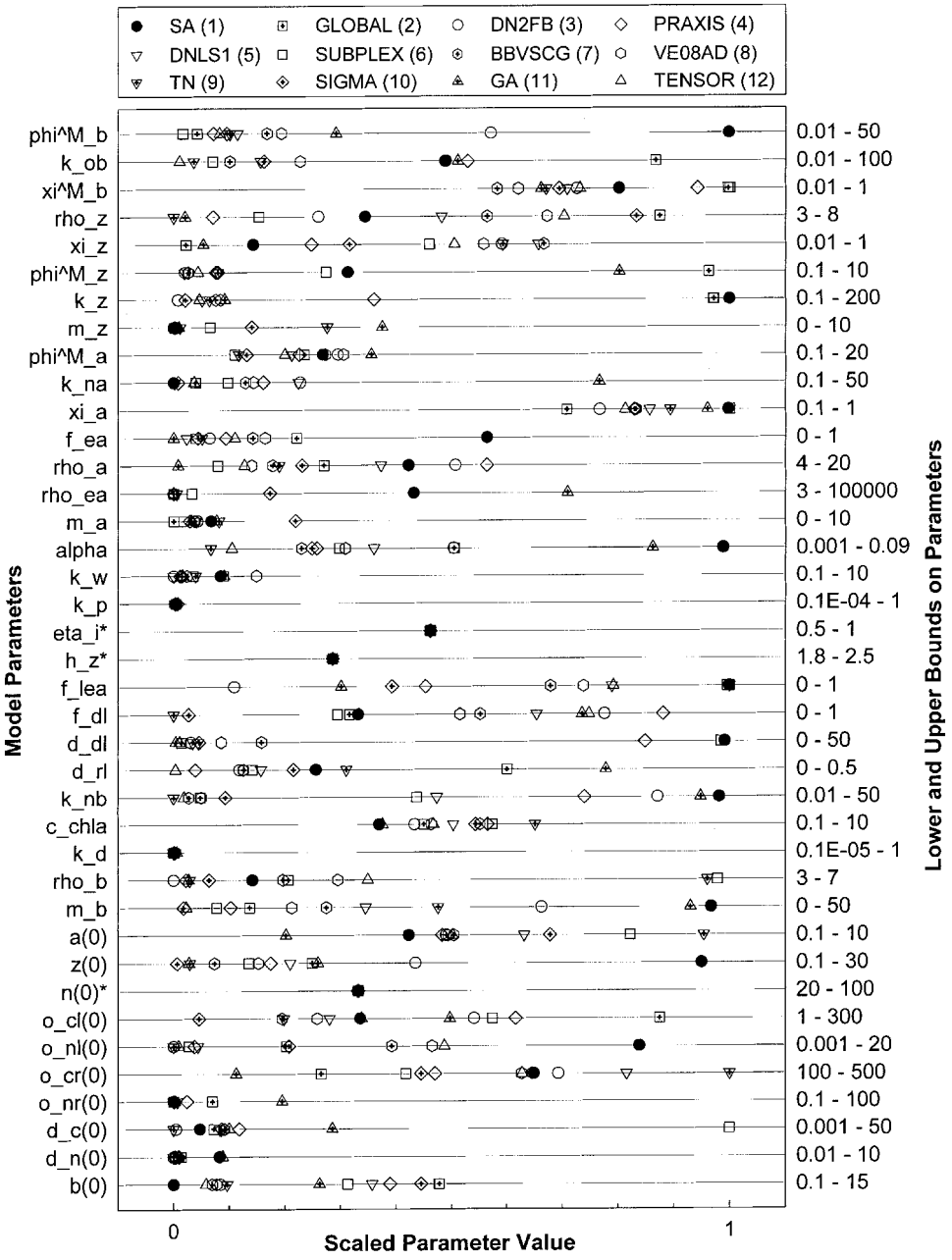

Figure 5. Parameter values, scaled by Eq. 15, associated with each of the minima located by the twelve optimization routines. Parameters marked with an asterisk were held constant for data assimilation. Also see Tables 3 and 5.

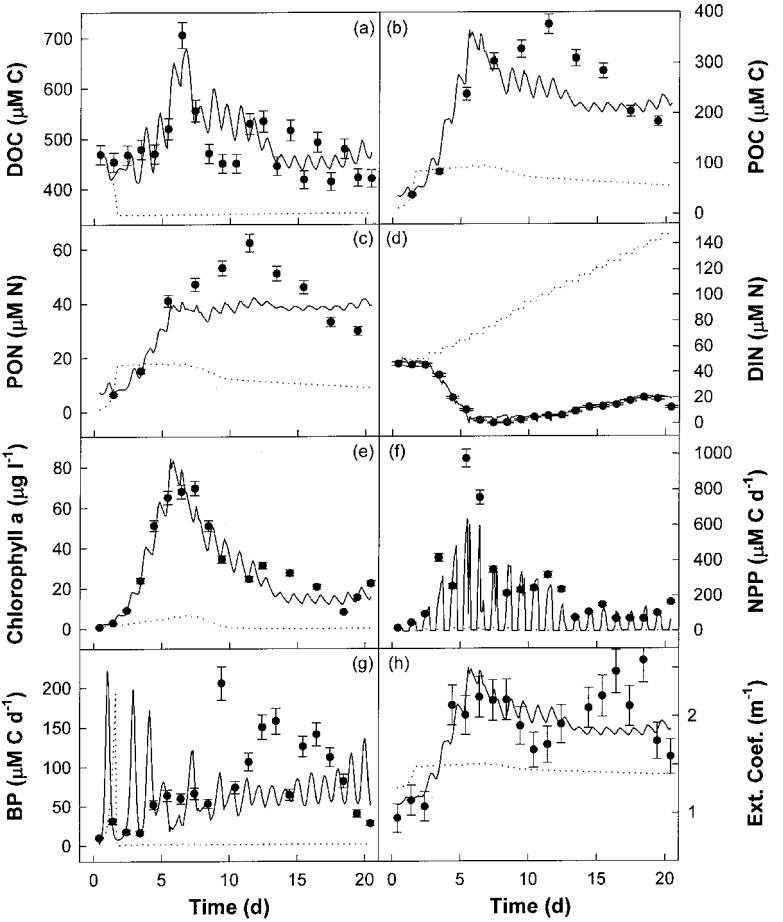

Figure 6. Comparison between food web model simulations (lines) and the eight mesocosm observed variables (fill circles) for the DOM + DIN treatment (Bag D). Model simulations are based on initial parameter guesses (dashed line, Table 1) and the optimum parameter set (Table 1) obtained by the simulated annealing (SA) routine (solid line). Error bars are based on Eq. 20 and values listed in Table 2.

system, these driver oscillations are amplified in the state and output variables due to the parameter values associated with the optimum located by the SA routine (Fig. 6). In particular, the extreme growth kinetics for bacteria (φ_B^M, k_{OB}, m_B, Table 1), the rapid decomposition rate of detritus (d_{DL}, Table 1) and the high DOM exudation rate by phytoplankton (f_{EA}, Table 1) produce the significant oscillatory behavior in DOC and bacterial production (Fig. 6a,g). Other optimal solutions with higher final costs (Table 5) did not exhibit such oscillations due to less extreme parameter values; however, these parameter sets did not produce as good a fit to the DOC data (Fig. 7).

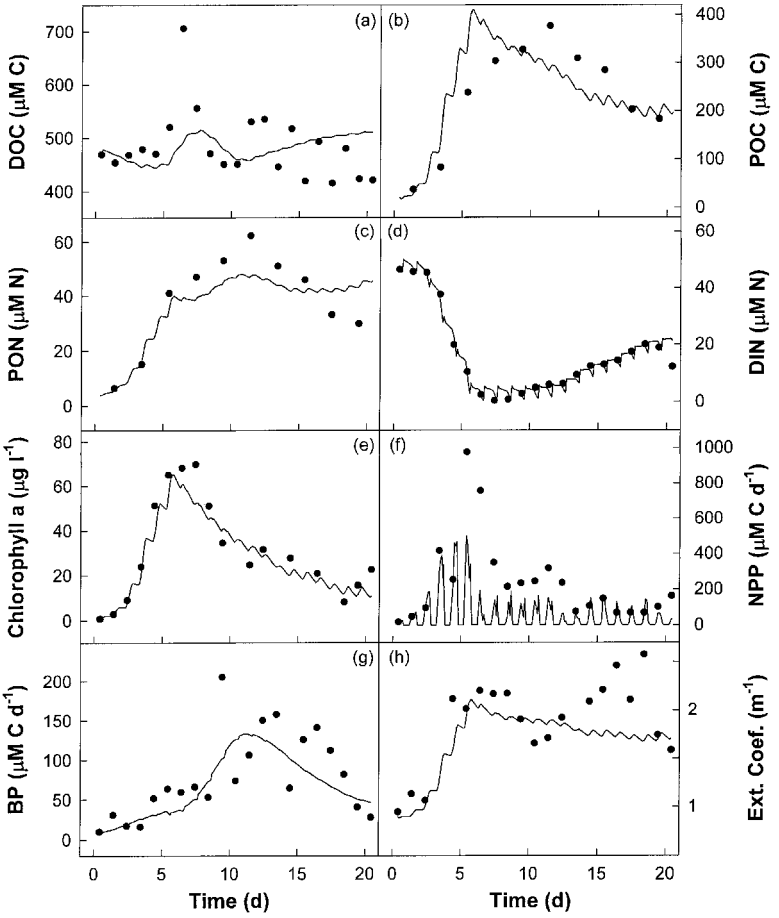

Figure 7. Same as Figure 6, except model fit based on optimum parameter set obtained by the PRAXIS routine (Table 5).

6. Discussion

In general, the fit between the food web model and the observations (Fig. 6) is inadequate if we intend to use the model as a prognostic tool. The poor fit is not unusual for these models, as several other investigators have had difficulty in fitting ecosystem models to observations (Fasham and Evans, 1995; Lawson *et al.*, 1996; Matear, 1995; Prunet *et al.*, 1996). Here we examine possible areas that may prevent good fits between model and observations and provide some suggestions as to how food web models might be improved.

a. Parameter observability

The large condition number of the Hessian matrix about the SA minimum indicates that the values of some parameters may be difficult to determine, at least at this point in K_B.

Examination of the singular values (Fig. 4) and resolution matrix of $\tilde{\mathbf{H}}$ (Table 4) indicate that the parameters ρ_{EA}, d_{DL}, $D_C(t_0)$, *and* $D_N(t_0)$ are poorly resolved. However, due to strong nonlinearities in the food web model, it is likely that poor resolution of the parameters ρ_{EA} and d_{DL} is only associated with the SA minimum. For instance, for large values of the C to N ratio of phytoplankton exudate, ρ_{EA}, we would not expect the objective function to be sensitive to changes in ρ_{EA} because the difference in N content of DOM exudate with a C:N of 43200 versus 30000 is basically zero. Conversely, when ρ_{EA} is small (i.e., similar in value to ρ_A) a small perturbation in ρ_{EA} will have relatively major effects on N dynamics, which would cause significant changes in the value of the objective function. Consequently, identifying dependent (or poorly resolved) parameters in highly nonlinear models is a nontrivial task, because parameter resolution is a function of the location of the optima.

Although the presence of dependent parameters in a model is not desirable, such dependency does not prevent the model from fitting the observations; that is, parameter dependency does not explain the poor fit of the food web model to some of the observations (Fig. 6). The presence of dependent parameters simply means that there exists more than one solution (parameter set) that will fit the observations equally well. If the main object of the data assimilation is to produce a model that accurately tracks state, $\mathbf{x}(t)$, and output, $\mathbf{y}(t)$, variables, then parameter dependency is not critical. However, if some of the parameters in the model have intrinsic value to the modeler (for instance, they will be used in another model or can be directly measured), then parameter dependency needs to be removed. Removal of dependent parameters also reduces computational requirements because it decreases the dimension of searchable parameter space, K_B.

b. Objective function

In our analysis we used a least-squares objective function ($J(\mathbf{k})$, Eq. 4) to describe the quality-of-fit between the model output and the observations. Interestingly, the best least-squares fit between the food web model and the mesocosm data (Fig. 6) was not the most appealing fit due to the oscillatory nature of some of the state and output variables (Fig. 6a,g) and the extreme values of many of the parameters associated with OM processing (Table 1). Visually, other solutions appeared to fit the data marginally better than the SA solution (e.g., PRAXIS, Fig. 7), but had a slightly larger objective function (Table 5). Since the parameter set obtained by data assimilation ultimately depends on the functionality of the objective function, use of different criteria for $J(\mathbf{k})$, such as maximal absolute error, may be worth investigating (Janssen and Heuberger, 1995). However, such modifications are relatively minor in this case, and it is not expected they would improve the general fit between the food web model and mesocosm data (Fig. 6) since the model likely contains significant structural errors (discussed below). Model structural errors are also evident by the large residual error compared to the assumed measurement errors (Fig. 6).

c. Local optima and optimization

It is clear from Table 5 and Figure 5 that the hyper-surface described by the objective function $J(\mathbf{k})$ is replete with local optima. Indeed, routine GLOBAL located 20 unique optima with $J(\mathbf{k})$ ranging from 204 to 9443. To get some sense of the $J(\mathbf{k})$ surface, we visually examined a large number of 2D and 3D slices through the optimum located by the SA routine. Although many of these slices revealed well-behaved surfaces (i.e., describable by low order polynomials, Fig. 8a), several of the surfaces examined were highly irregular, especially surfaces defined by parameters associated with OM processing (Fig. 8b). In addition to the several smooth local minimum valleys on the k_{OB} versus $O_{CL}(t_0)$ surface (Fig. 8b), there are also highly spiked areas on the surface that appear to be numerical instability problems. Inspection of these spikes reveals no numerical problems. Instead, the spikes occur when a peak in modeled bacterial production coincides with an observation (Fig. 6g). Small shifts in either k_{OB} or $O_{CL}(t_0)$ time-displace the $B_P(t)$ peak, so that a spike in $J(\mathbf{k})$ appears and disappears as the $B_P(t)$ peak wanders on and off an observation as a function k_{OB} or $O_{CL}(t_0)$ (consider Fig. 3 with spikes in model output). It seems likely that this phenomenon can occur for other highly dynamic variables. Although the spikes perhaps could be removed by not including such measurements as $B_P(t)$ and $N_{PP}(t)$, it is usually better to include all possible measurements to fully constrain the model. Even if the spikes were not present, $J(\mathbf{k})$ would still contain many local optima both within and on the boundary of K_B (Fig. 8a, b). Consequently, selection of an appropriate optimization routine that can perform well under the above circumstances is an important component of data assimilation.

The simulated annealing algorithm in routine SA performed the best in this study (Table 5), so we recommend its use for similar type problems (also see routine ASA (Ingber, 1993), Matear (1995), and Siarry *et al.* (1997)). The fact that this algorithm has two to three orders of magnitude greater computational requirements than classical algorithms is becoming less of a problem since computation costs continue to drop while CPU speeds increase. However, food web models embedded in ocean circulation models (Gunson *et al.*, 1999; Sarmiento *et al.*, 1993) or other transport models (Hopkinson and Vallino, 1995; Oguz *et al.*, 1996) are still computationally intensive, so use of efficient optimization routines is paramount for these cases. For models with high computational requirements, we recommend using either DN2FB or PRAXIS, with a preference going to PRAXIS since this routine does not require gradient information. We also note that convergence to the global optimum using these local optimization routines can be improved by employing a subset of the observation data in an iterative manner (Wang and Luus, 1980). In this approach, an optimum is located using a subset of the data, then more data are introduced and a new optimum is located. This iteration continues until all observation data have been assimilated.

In our analysis, we have examined local and global optimization algorithms that require the gradient of the objective function as well as ones that do not (Table 3). Our results show

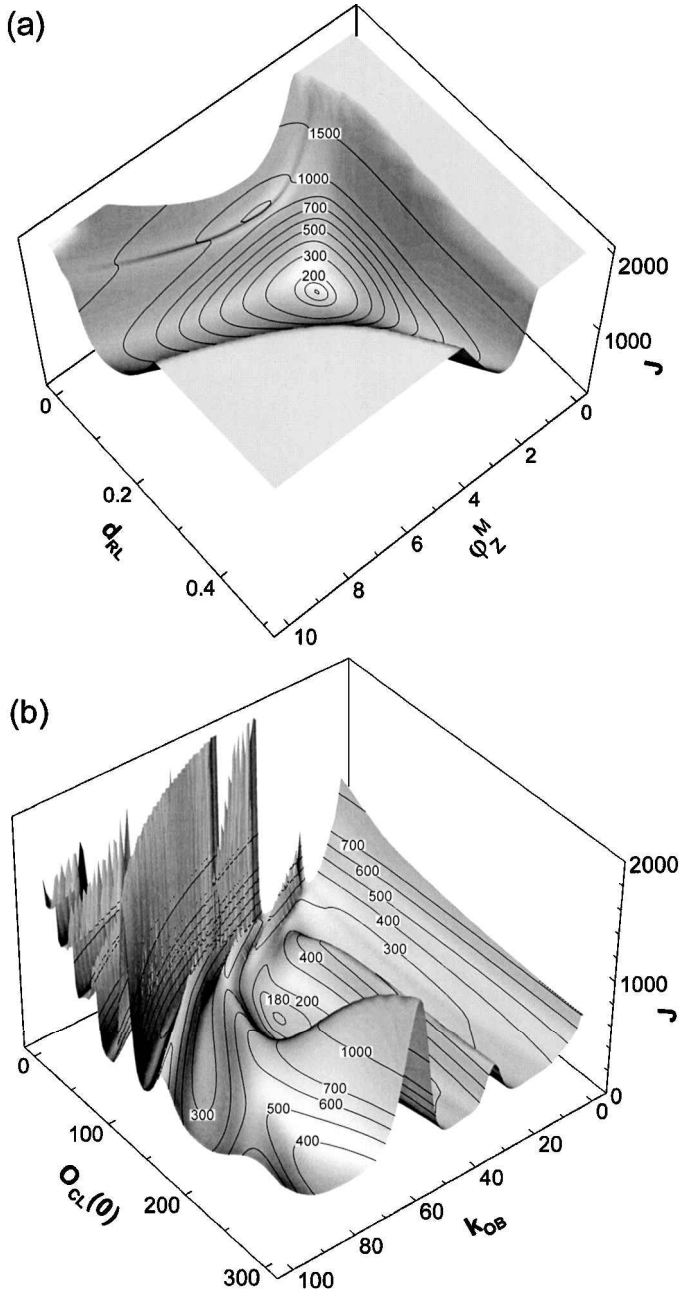

Figure 8. Two-dimensional slices through the objective function hyper-surface (Eq. 4) about the optimum located by the SA routine (Table 1). (a) Objective surface as a function of maximum zooplankton growth rate, φ_Z^M (d^{-1}), and the first order decomposition rate of refractory OM to labile OM, d_{RL} (d^{-1}). (b) Objective surface as a function of the half saturation constant for labile DOC, k_{OB} (μM), and initial concentration of labile DOC, $O_{CL}(t_0)$ (μM).

that the routines that use gradient information converged to an optimum faster, often significantly faster, than those routines not using gradient information (except TENSOR, Table 5). We also found that the gradient calculated via the adjoint method (Eqs. 7 and 8) had greater accuracy and precision and was 10 to 20 times faster than forward finite-differences (Eqs. A3 and A4 in the Appendix). Yet many of the gradient-requiring routines, especially those where the adjoint equations were used (BBVSCG, VE08, TN, and TENSOR), ranked poorly in terms of final cost and often terminated with algorithm errors (Table 5). This is perhaps not surprising given the highly irregular nature of the objective function (Fig. 8b). Since the programming and computational requirements associated with the adjoint equations and finite differences are high, we recommend the use of routines that do not require calculation of the gradient, except for those cases where computational requirements are extreme, as noted above. Given the good performance of the Levenberg-Marquardt-like routines (DN2FB and DNLS1, Table 5), we also recommend development of adjoint equations for the calculation of gradient information associated with vector objective functions (Eq. A5), as required by these routines.

Does the presence of many local optima (Fig. 8) prevent a good fit between model and observations (Fig. 6)? Since we cannot prove that a given optimum is global, there always exists the possibility that a different solution, possibly outside of K_B, would allow the model to fit the observations perfectly, especially for large dimensional problems where an exhaustive search over all parameter space is computationally impossible. To examine the possibility, albeit unlikely, that the optimization is the source of poor model fit, we used the food web model with the parameters obtained by the SA routine (Table 1) to generate simulated observations at one-day intervals. Using the same initial guess for **k** as before (Table 1), we then examined how well the optimization routines SA, GLOBAL and PRAXIS could fit the simulated data. For this case, known as a twin experiment, an exact solution exits and is known *a priori.* We found similar results for all three routines. Good fits were obtained between the model and simulated data for all output variables except for DOC, $D_{OC}(t)$, and bacterial production, $B_P(t)$ (Fig. 9). The poor fit to $D_{OC}(t)$ and $B_P(t)$, however, is understandable, since the SA solution produces a very dynamic model (Fig. 6) which is difficult to resolve with only daily sampling. The twin experiment was also able to recover most of the parameter values, except some associated DOM and bacteria dynamics. The twin experiment also indicates that the sin-squared transform (Eq. 11) did not inhibit recovery of the parameters, since routines SA and GLOBAL do not use the transform but PRAXIS does, yet all three routines were able to recover the parameters (Fig. 9). Consequently, it is unlikely that poor optimization is the root cause of the less-than-perfect model fit.

d. Model structure and aggregation uncertainty

The root cause of the poor fit between the model and observations (Fig. 6) is the underlying model structure. Model structure refers to the state variables used to represent the fundamental system dynamics, how the state variables are interconnected, and the type

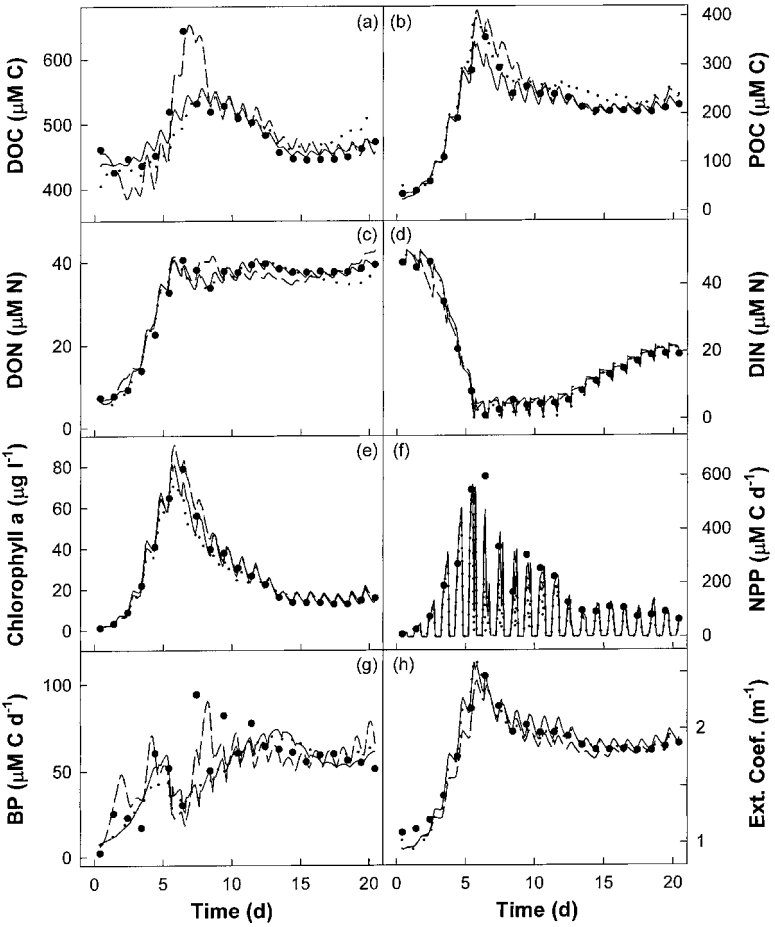

Figure 9. Twin experiment to examine the ability of data assimilation to recover parameters from simulated data (filled circles) based on model parameters obtained from the best solution, **k*** (Table 1). Optimization routines tested: SA (solid line), GLOBAL (dashed line) and PRAXIS (dotted line).

of growth kinetics used to describe the interconnecting flows. There are two main sources of model structure uncertainty. One source is simply the lack of adequate knowledge about the system, which in our case is mostly associated with organic matter production and consumption. The second uncertainty is due to model aggregation.

Due to the extreme complexity of aquatic ecosystems, almost all aquatic food web models aggregate a large number of functionally similar species into a single compartment (i.e., zooplankton) or perhaps several subcompartments when size is also considered (Moloney and Field, 1991). These aggregated compartments are then modeled as a single species, where the kinetic parameters are often derived from laboratory experiments involving pure cultures. Although the growth kinetics used for aggregated state variables

are gross approximations, their use continues due to the lack of comprehensive knowledge on ecosystem food webs and the difficulty with aggregation in nonlinear models (Iwasa *et al.*, 1989).

Aggregated models, nevertheless, maintain a high degree of nonlinearity and associated dynamics, including chaos (Beckers and Nihoul, 1995; Kot *et al.*, 1992), and typically exhibit strong predator-prey dynamics due to the auto-catalytic nature of the growth kinetics (i.e., growth rate is a function of biomass). As a result, aggregated models are quite sensitive to parameter values and initial conditions. For example, Steele and Henderson (1992) have noted that slight changes in the mortality function in simple food web models can lead to significant differences in model dynamics. Such sensitivity to parameter values explains the large number of local optima we observed in the $J(\mathbf{k})$ surface (Fig. 8) and the extreme range of values associated with a given parameter obtained by the optimization routines examined in this study (Fig. 5). Furthermore, model aggregation often makes it difficult to obtain model parameters from direct measurements. For instance, Wallach and Genard (1998) have shown that a better fit between model and observations can usually be obtained if the model parameters are allowed to assume values that differ from their values reported in the literature, which appears consistent with our results (Table 1). Unfortunately, this makes it difficult to use parameter values reported in the literature that are based on culture experiments or allometric relationships (Moloney and Field, 1991) when aggregated models are used. This also means parameters obtained by data assimilation for one model cannot be readily used for other models.

Current generation food web models, including ours, also suffer from the use of fixed-value parameters. All parameters used in our food web model are treated as constants, yet many of the parameters can change value either due to physiological changes in the organisms or shifts in species composition. For example, it is well known that the C:N ratio of phytoplankton can vary widely depending on N availability or growth rate (Goldman *et al.*, 1979), yet this parameter is held constant in our model. Likewise, since state variables represent an aggregation of functionally similar organisms, alteration of the effective uptake kinetics of nutrients, such as NO_3 or NH_4, can change as different species dominate the functional group. Although one can attempt to make model parameters a function of environmental conditions, the complexity of the resulting models increase in both the number of state variables and parameters. This increased complexity does not guarantee a better fit. For instance, to better capture the observed DOC dynamics (Fig. 6a), we tried modeling the C and N content of phytoplankton explicitly to allow for internal C or N storage (i.e., variable phytoplankton C:N ratio) (e.g., Geider *et al.*, 1998; Tusseau *et al.*, 1997). Even though this model had two more state variables and several more parameters, it did not fit the DOM data better than the model given in the Appendix. We also note here that a reduced-complexity model using only first order kinetics (instead of Monod, or Holling type-III) was tried, but this pseudo-linear model fit the data poorly (Fig. 10). Consequently, nonlinear kinetics must be retained in order to accurately capture food web dynamics.

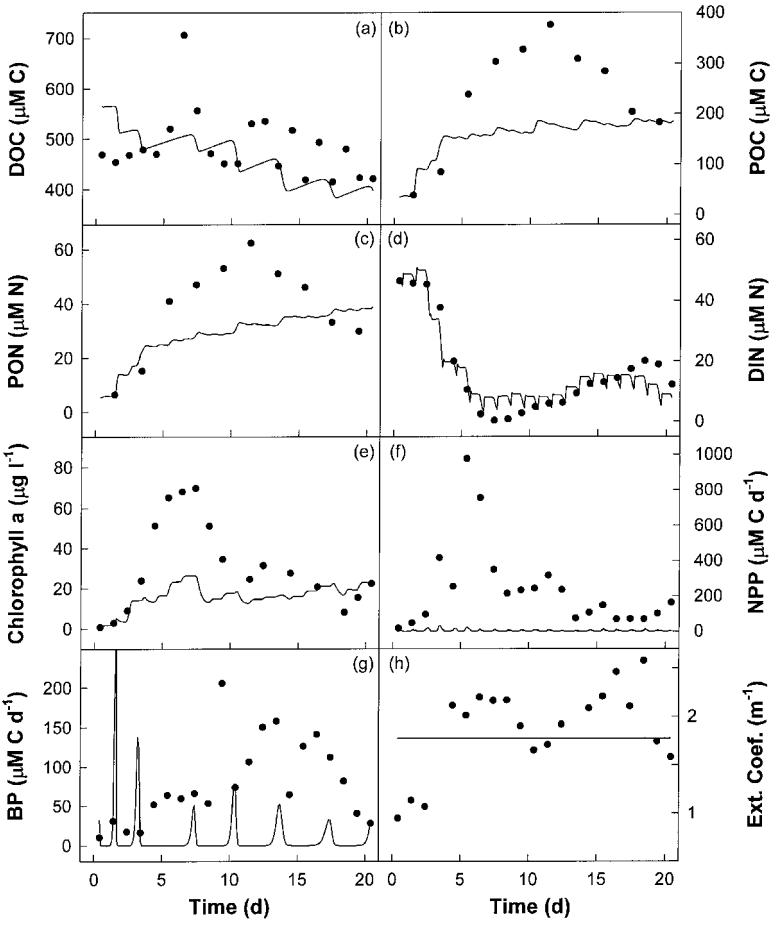

Figure 10. Same as Figure 6, except a linear food web model was used to fit the observations.

The problem with fixed-valued parameters (or inaccurate model structure) becomes quite apparent when we compared the food web model simulation using the optimum parameter set (Table 1) to data from the same experiment, but from a different mesocosm treatment. Although the model does an adequate job of fitting the DOM only treatment (Bag B, Fig. 11), the model only marginally reproduces the control treatment (Bag A, Fig. 12) and completely breaks down for the DIN-only treatment (Bag C, Fig. 13). The model is able to reproduce all four treatments if data assimilation is used to find new parameter sets for each treatment, but the new parameter sets differ significantly form the parameter set obtained from the DIN + DOM treatment (i.e., Table 1).

Although the fixed-value parameter constraint may be less of a problem for pelagic food web models where environmental conditions are more stable, it is a significant problem for systems where environmental conditions can change significantly over relatively short

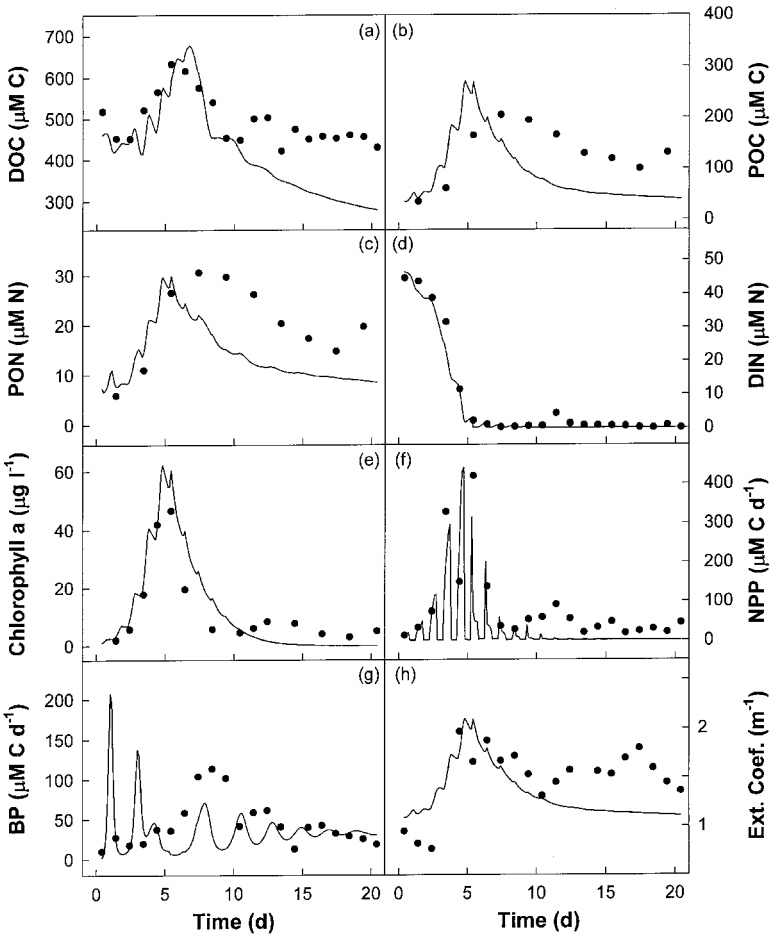

Figure 11. Same as Figure 6, except observations are from the DOM-only treatment (Bag B). Model simulations are based on the optimum parameter set obtained by the SA routine (Table 1), except the daily addition of DIN has been turned off during simulation.

temporal or spatial scales, such as in estuaries. Indeed, our mesocosm experiment was intended to contrast ecosystem dynamics between the relatively oligotrophic marine end-member (treatment A) and the eutrophic freshwater end-member (treatment D). Unfortunately, there are no easy solutions to remove problems associated with model structure uncertainty, but we believe aquatic ecosystems model can be improved.

e. Suggestions for model improvements

One means to remove problems associated with model aggregation is to explicitly represent all organisms and their interactions. This reductionist approach is likely to fail, of course, since real ecosystems consist of hundreds, if not thousands, of species exhibiting

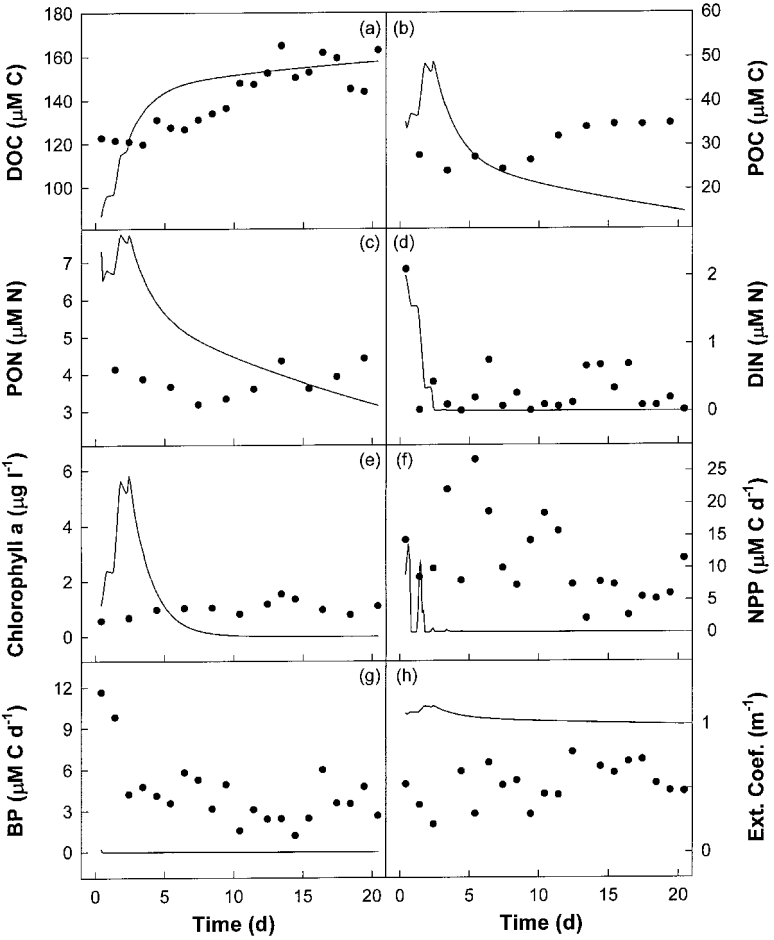

Figure 12. Same as Figures 6 and 11, except observations are from the control treatment (Bag A), and initial conditions for DOM ($O_{CL}(t_0)$, $O_{NL}(t_0)$, $O_{CR}(t_0)$, and $O_{NR}(t_0)$) and DIN ($N(t_0)$) have been changed to reflect the lower initial concentrations in Bag A, and the daily addition of DIN has been turned off.

autotrophic, heterotrophic, or even mixotrophic feeding behaviors, that can change with ontogeny (Turner and Roff, 1993). It is unlikely that we will ever be able to accurately model all these organisms and their feeding behaviors. Indeed, a recent model of phytoplankton growth required three state variables and ten parameters to capture observed dynamics for a single species (Geider *et al.*, 1998). Instead, it will probably be more productive to continue modeling these systems as consisting of just a few trophic or functional groups (Totterdell *et al.*, 1993); however, the underlying growth kinetics should not be modeled as a single organism, as is currently done in many models. For the aggregation approach to succeed, a greater effort must be placed on developing growth models that represent the feeding characteristics of an aggregated consortium of species or

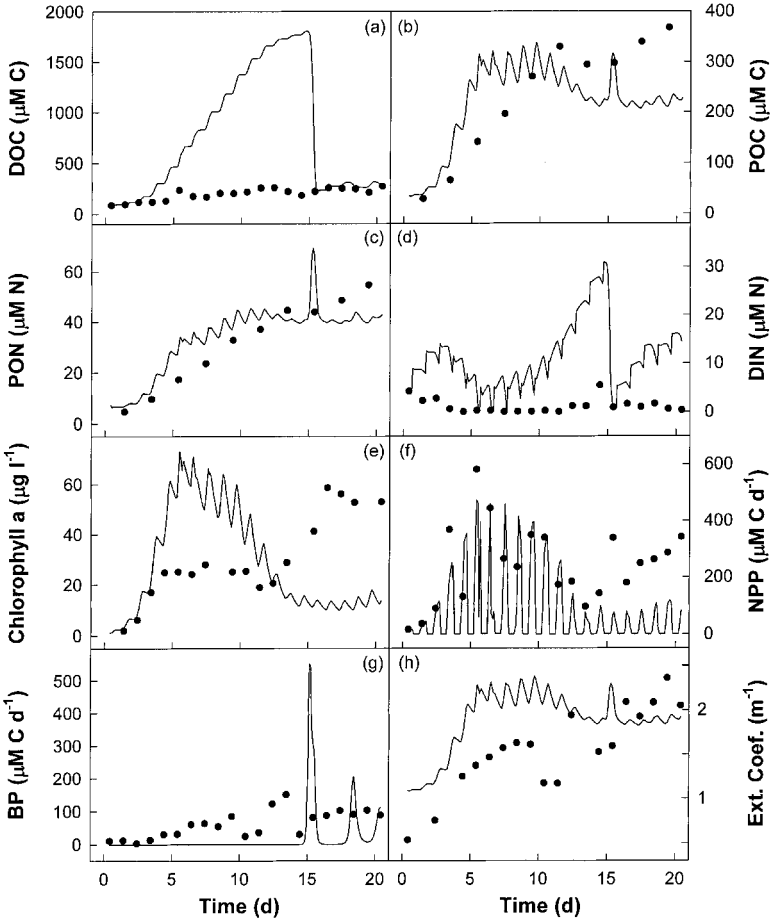

Figure 13. Same as Figures 6 and 11, except observations are from the DIN Only treatment (Bag C), and initial conditions for DOM ($O_{CL}(t_0)$, $O_{NL}(t_0)$, $O_{CR}(t_0)$, and $O_{NR}(t_0)$) and DIN ($N(t_0)$) have been changed to reflect the lower initial concentrations in the experiment.

individuals. For example, intratrophic feeding can be modeled as a general type of cannibalism (Pitchford and Brindley, 1998). Such aggregated models will require development from fundamental concepts and holistic understanding of ecosystem function, as opposed to systematic aggregation of complex models (Iwasa *et al.*, 1989). This is a significant departure from standard approaches, where the emphasis has been on isolation of organisms and reductionist modeling.

Use of more abstract representations of aquatic ecosystems that attempt to capture emergent properties of these systems may prove to be more robust with respect to parameter uncertainty and data requirements (Platt *et al.*, 1981; Vallino *et al.*, 1996). However, these models have not seen wide applicability due to the complex concepts involved whose understanding is still in its infancy. Nevertheless, classical engineering

approaches to modeling ecosystem dynamics that rely solely on mass and energy conservation have their limitations, so novel approaches will likely be required for improving ecosystem dynamic models.

We have found the use of mesocosm experiments coupled to data assimilation to be a powerful tool for the development and testing of aquatic ecosystem models. Indeed, the food web model given in the Appendix is the result of data assimilation with several iterations of models with various state dimensions and levels of complexity, including a linear model (Fig. 10). Currently, food web models are not rigorously tested against experimental data, so it is difficult to know if a new model represents a real improvement in predictive capabilities. Furthermore, if the model does not fit observations, it is unknown whether the poor fit of the model is due to parameter uncertainty, structural uncertainty, or both. By using data assimilation coupled with extensive experimental data, we can determine the best a given model can fit an experimental observation set (assuming the global optimum can be found). If, after data assimilation, the model still fits the data poorly, we can assume with high probability that the model suffers from structural uncertainties. In other words, data assimilation can be used to remove parameter uncertainty and thereby reveal structural uncertainty. The modeling community can then focus on improving model structure. Mesocosm experiments also facilitate development of robust models, since environments ranging from oligotrophic to eutrophic conditions can be readily produced, and perturbation studies can be conducted to discriminate between models based on transient responses. Mesocosm experiments of different size, scope and complexity can be conducted (from small tanks to much larger ocean enclosures) to address various modeling needs.

Detractors of mesocosms claim the experiments are not useful due to side effects introduced by the containing walls, which do not exist in real systems. Although mesocosm walls do influence ecosystem dynamics, their influence can be explicitly described by developing two or more submodels. Thus, one model is developed for the water column and a separate model accounts for processes occurring on the walls and/or benthos. The models are developed so that they can be readily separated. This modularity allows the mesocosm wall model to be dropped when the water column model is embedded into a transport model. Of course, with this approach, the mesocosm wall must be an integral part of the experimental observations, so that the effects of the walls can be accurately modeled.

There is also a need to develop experimental measurements that address specific modeling needs. These new measurements may, or may not, be easily interpretable on their own. For example, measuring the fraction of POM that is living would allow direct constraints to be imposed on biotic compartments, but may be of little interest to experimentalists. Until we aggressively combine modeling with experimental observations, we are unlikely to see significant improvements in aquatic ecosystem models.

Acknowledgments. The mesocosm data are a crucial component of this manuscript and would not have been possible without the expertise and long hours contributed by Chuck Hopkinson, Linda Deegan, Anne Giblin, John Hobbie, Hap Garritt, Jane Tucker, Michele Bahr and Ishi Buffam. The

detailed comments provided by three anonymous reviewers are greatly appreciated. This work was partially supported by grants from the National Science Foundation (OCE-9214461 and OCE-9726921) and the Lakian Foundation.

APPENDIX

This Appendix contains (1) alternative approaches for calculating the gradient of the objective function, $J(\mathbf{k})$, (2) descriptions of the optimization routines employed in the study, (3) the equations governing the state and measurement models for the mesocosm experiment, and (4) the numerical routines employed.

Alternatives for calculating $\nabla_{\mathbf{k}} J(\mathbf{k})$

Sensitivity method

In this method (Biegler *et al.*, 1986), $\nabla_{\mathbf{k}} J(\mathbf{k})$ is calculated by taking the derivative of Eq. 4 with respect to $\mathbf{k}$, as follows,

$$\begin{aligned}\frac{\partial J(\mathbf{k})}{\partial \mathbf{p}} &= 2\int_{t_0}^{t_f}\left(\frac{\partial \mathbf{h}(\mathbf{x}(t;\mathbf{k}),t;\mathbf{p})}{\partial \mathbf{p}} + \frac{\partial \mathbf{h}(\mathbf{x}(t;\mathbf{k});t;\mathbf{p})}{\partial \mathbf{x}(t;\mathbf{k})}\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{p}}\right)^{\mathrm{T}} \\ &\quad \cdot \boldsymbol{R}(t)^{-1}(\mathbf{h}(\mathbf{x}(t;\mathbf{k}),t;\mathbf{p}) - \mathbf{y}(t))\,dt \qquad \text{(A1)} \\ \frac{\partial J(\mathbf{k})}{\partial \mathbf{x}_0} &= 2\int_{t_0}^{t_f}\left(\frac{\partial \mathbf{h}(\mathbf{x}(t;\mathbf{k}),t;\mathbf{p})}{\partial \mathbf{x}(t;\mathbf{k})}\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{x}_0}\right)^{\mathrm{T}} \boldsymbol{R}(t)^{-1}(\mathbf{h}(\mathbf{x}(t;\mathbf{k}),t;\mathbf{p}) - \mathbf{y}(t))\,dt\end{aligned}$$

where use of the chain rule has been made. Eq. A1 contains two unknown matrices,

$$\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{p}} \quad \text{and} \quad \frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{x}_0}.$$

These matrices can be determined by taking the derivative of Eq. 1 with respect to $\mathbf{p}$ and $\mathbf{x}_0$ and again employ the chain rule as follows:

$$\begin{aligned}\frac{d}{dt}\left(\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{p}}\right) &= \frac{\partial \mathbf{f}(\mathbf{x}(t;\mathbf{k}),t;\mathbf{p})}{\partial \mathbf{x}(t;\mathbf{k})}\left(\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{p}}\right) + \frac{\partial \mathbf{f}(\mathbf{x}(t;\mathbf{k}),t;\mathbf{p})}{\partial \mathbf{p}}, \qquad \left(\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{p}}\right)_{t=t_0} = \mathbf{0} \\ \frac{d}{dt}\left(\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{x}_0}\right) &= \frac{\partial \mathbf{f}(\mathbf{x}(t;\mathbf{k}),t;\mathbf{p})}{\partial \mathbf{x}(t;\mathbf{k})}\left(\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{x}_0}\right), \qquad \left(\frac{\partial \mathbf{x}(t;\mathbf{k})}{\partial \mathbf{x}_0}\right)_{t=t_0} = \mathbf{I}.\end{aligned} \qquad \text{(A2)}$$

It is clear from Eq. A2, that in order to calculate $\nabla_{\mathbf{k}} J(\mathbf{k})$, an $m \times n$ matrix of differential equations must be solved each time the optimization routine requires the gradient. Due to the high computation overhead, this approach was abandoned in favor of the adjoint method (Eqs. 7 and 8).

Finite differences method

When the adjoint method could not be used to calculate the gradient of $J(\mathbf{k})$, a finite difference method was used. The forward difference estimate of the gradient is

given by

$$\frac{\partial J(\mathbf{k})}{\partial k_j} \approx \frac{J(\mathbf{k} + \Delta k_j \mathbf{e}_j) - J(\mathbf{k})}{\Delta k_j}, \qquad j = 1, \ldots, n \tag{A3}$$

with

$$\Delta k_j = \max\{\delta, |k_j|\delta\} \qquad \text{and} \qquad \delta = \max\left\{\sqrt{\varepsilon}, \frac{\kappa}{10}\right\} \tag{A4}$$

where $\mathbf{e}_i$ is an elementary vector with 1 at element i and zeros elsewhere, ε is the machine precision, and κ is the user requested accuracy (Kraft, 1994). Clearly, as Δk_j approaches zero in Eq. A3, gradient accuracy increases, but precision goes toward zero. Consequently, the finite difference approach sacrifices accuracy for precision or vice versa (Herstine, 1998). This approach often results in a large loss of precision, perhaps maintaining only half the precision compared to the adjoint or sensitivity methods.

Optimization routines

Four global and eight local optimization routines (Table 3) were investigated to find the optimal parameter set(s), $\mathbf{k}^*$, that minimizes the nonlinear objective function $J(\mathbf{k})$, Eq. 4. This section briefly describes these routines. For reference, the most rudimentary method, steepest descent, is briefly described.

The method of steepest descent proceeds in an iterative fashion (as most methods do) from the current point, $\mathbf{k}^{(i)}$, in parameter space to an optimum in one dimension along a search direction described by the negative of the function gradient at the current point, $-\nabla_{\mathbf{k}} J(\mathbf{k})|_{\mathbf{k}=\mathbf{k}^{(i)}}$. As with most local optimization techniques, the iteration terminates when some specified criteria are met, such as minimal change in $J(\mathbf{k})$, $\nabla_{\mathbf{k}} J(\mathbf{k})$ approaches zero, or insignificant difference between $\mathbf{k}^{(i)}$ and $\mathbf{k}^{(i+1)}$. The steepest descent method works adequately when the current point is far from the local optimum. However, the method becomes inefficient in the neighborhood of the optimum because the search direction specified by the gradient does not take advantage of the quadratic nature of the objective function in the vicinity of the optimum. Consequently, steepest descent methods are typically not used, or are employed only initially in search algorithms.

Routine SA (Corana *et al*., 1987; Goffe *et al*., 1994) uses simulated annealing to locate a global optimum. The algorithm employs a stochastic search, whereby a simulated temperature governs the likelihood of accepting a current guess. At high-simulated temperature, there is a high probability that a suboptimal guess will be incorporated into the search. This simulated thermal noise helps to prevent the routine from being trapped in a local minimum. The algorithm is based on phase transitions in condensed matter, such as freezing.

Routine GLOBAL (Boender *et al*., 1982; Csendes, 1988) combines a standard conjugate gradient-type search with a stochastic clustering multistart driver to locate the global minimum. The stochastic multistart driver provides initial locations to begin the deterministic gradient-based search.

Routines DN2FB (Dennis *et al.*, 1981a, b) and DNLS1 (Vandevender and Haskell, 1982) employ a variation of the Levenberg-Marquardt least-squares algorithm (Conway *et al.*, 1970). These algorithms take advantage of the form of the minimization function (i.e., sum of squared residuals, or χ^2 distribution), which allows direct calculation of the Hessian matrix (Press *et al.*, 1986 pg. 522). The disadvantage of these routines in the current context is that the objective function is represented as a vector instead of scalar and is discrete in time, as given by,

$$\hat{J}_{(j-1)q+i+1}(\mathbf{k}) = \frac{(h_j(\mathbf{x}(t_i;\mathbf{k}), t_i; \mathbf{p}) - y_j(t_i))^2}{\sigma_j^2}; \qquad i = 0, \ldots, q-1; \qquad j = 1, \ldots, \ell. \quad \text{(A5)}$$

Consequently, the adjoint equations could not be used to provide the gradient of the vector objective function, so finite differences were used (Eqs. A3 and A4).

Routines PRAXIS (Brent, 1973), BBVSCG (Buckley and Lenir, 1985; Buckley, 1994), VE08 (Griewank and Toint, 1982), and TN (Nash, 1984) employ variations of the conjugate gradient algorithm to locate local minimum; however, PRAXIS does not require gradient information.

The routine TENSOR (Chow *et al.*, 1994) also uses a conjugate gradient approach; however, instead of approximating $J(\mathbf{k})$ by a quadratic function about the optimum, TENSOR employs a fourth-order (quartic) approximation. Consequently, in addition to $J(\mathbf{k})$ and its gradient, the routine also explicitly requires the Hessian of $J(\mathbf{k})$. Solution of the adjoint equations provided the gradient of $J(\mathbf{k})$, but finite differences (internal to TENSOR) were used to calculate the Hessian.

Routine SUBPLEX (Rowan, 1990) employs a variation of the simplex algorithm to subspaces of the overall parameter space. Simplex methods (Marsili-Libelli, 1992; Nelder and Mead, 1965) (not to be confused with the simplex method of linear programming) employ a simplex (i.e., the hyper-volume defined by $n + 1$ points whose edges span all of parameter space) to locate an optima. The $n + 1$ vertex points of the simplex are moved in a simple geometric manner so as to first move the entire simplex "down-hill" (for minimization) and then contract around the local optimum. The advantages of the simplex method are that (1) gradient information is not required, (2) discontinuous or discrete functions can be handled, and (3) it is easily implemented. Like steepest descent, however, the simplex method does not converge to the optimum in the fewest number of steps for quadratic objective functions and only locate local optima.

Routine SIGMA (Aluffi-Pentini *et al.*, 1988a,b) uses the solution of a stochastic differential equation derived from the steepest descent equation perturbed by white noise. In theory, the noise perturbation of the differential equation allows a trajectory to be bumped across barriers so that the algorithm can potentially locate the global optimum.

The routine GA (Carroll, 1996) uses a genetic algorithm to locate the global optimum. Genetic algorithms locate optima by simulating natural selection as follows. Initialization

begins by generating a random population of individuals that each represents a unique parameter set, **k**. The fitness (i.e., $J(\mathbf{k})$) of each individual is determined, and the fittest (i.e., smallest $J(\mathbf{k})$) are used to generate the next population of individuals in a manor analogous to gene crossover and mutation. The algorithm terminates after a specified number of generations have occurred.

All optimization routines started with the same initial guess for **k** (Table 1), except for routines GLOBAL and GA, which randomly select initial conditions. This initial guess produced a value of 39430 for the objective function (Eq. 4).

Mesocosm state and measurement models

Below is a description of the model developed for the mesocosm experiment. Because of our focus on dissolved organic matter (DOM), processes involving DOM production and consumption are represented in greater detail than is typically found in pelagic food web models (e.g., Fasham *et al.*, 1990). See Table 1 in main text for description of parameters.

State equations

State vector

$$\mathbf{x}^{\mathrm{T}}(t) = [A(t), Z(t), N(t), O_{\mathrm{CL}}(t), O_{\mathrm{NL}}(t), O_{CR}(t), O_{NR}(t), D_C(t), D_N(t), B(t)] \tag{A6}$$

Autotroph balance

$$\dot{A}(t) = (\varphi_A(t) - R_A(t) - E_A^C(t) - \varphi_{mA}(t))A(t) - \varphi_{AZ}(t)Z(t) \tag{A7}$$

Heterotroph balance

$$\dot{Z}(t) = (\varphi_{AZ}(t) + \varphi_{BZ}(t) - R_{ZA}(t) - R_{ZB}(t) - \varphi_{mZ}(t))Z(t) \tag{A8}$$

Bacteria balance

$$\dot{B}(t) = (\varphi_B(t) - R_B(t) - \varphi_{mB}(t))B(t) - \varphi_{BZ}(t)Z(t) \tag{A9}$$

Dissolved inorganic nitrogen balance

$$\dot{N}(t) = -U_{\mathrm{A}}^N(t)A(t) - U_B^N(t)B(t) + (E_{AZ}^N(t) + E_{BZ}^N(t))Z(t) \tag{A10}$$

Dissolved labile organic carbon balance

$$\dot{O}_{CL}(t) = f_{LEA}E_A^C(t)A(t) - \varphi_B(t)B(t) + d_{RL}O_{CR}(t) + f_{DL}d_{DL}D_C(t) \tag{A11}$$

Dissolved labile organic nitrogen balance

$$\dot{O}_{NL}(t) = \frac{1}{\rho_{EA}} f_{LEA}E_A^C(t)A(t) - \varphi_B(t)B(t)\frac{O_{NL}(t)}{O_{CL}(t)} + d_{RL}O_{NR}(t) + f_{DL}d_{DL}D_N(t) \tag{A12}$$

Dissolved refractory organic carbon balance

$$\dot{O}_{CR}(t) = (1 - f_{LEA})E_A^C(t)A(t) + (1 - f_{DL})d_{DL}D_C(t) - d_{RL}O_{CR}(t) \tag{A13}$$

Dissolved refractory organic nitrogen balance

$$\dot{O}_{NR}(t) = \frac{1}{\rho_{EA}}(1 - f_{LEA})E_A^C(t)A(t) + (1 - f_{DL})d_{DL}D_N(t) - d_{RL}O_{NR}(t) \tag{A14}$$

Detrital carbon balance

$$\dot{D}_C(t) = \varphi_{mA}(t)A(t) + \varphi_{mB}(t)B(t) + \varphi_{mZ}(t)Z(t) - d_{DL}D_C(t) \tag{A15}$$

Detrital nitrogen balance

$$\dot{D}_N(t) = \frac{1}{\rho_A}\varphi_{mA}(t)A(t) + \frac{1}{\rho_B}\varphi_{mB}(t)B(t) + \frac{1}{\rho_Z}\varphi_{mZ}(t)Z(t) - d_{DL}D_N(t) \tag{A16}$$

Autotroph growth equations

Autotrophs are limited by both DIN and light availability (Fasham *et al*., 1990), excrete f_{EA} fraction of net primary productivity as DOM, respiration is growth associated, and mortality is a function of DIN availability.

$$\varphi_A(t) = \Theta(I_o^+(t), t)\frac{N(t)}{N(t) + k_{NA}} \tag{A17}$$

$$\begin{aligned}\Theta(I_o^+(t), t) &= \frac{1}{h}\int_0^h \frac{\varphi_A^M \alpha I(t, z)}{\sqrt{(\varphi_A^M)^2 + (\alpha I(t, z))^2}}\,dz \\ &= \frac{\varphi_A^M}{h(k_w + k_p P_M(t) + k_d O_{CR}(t))}\ln\left[\frac{\alpha I(t, 0) + \sqrt{(\alpha I(t, 0))^2 + (\varphi_A^M)^2}}{\alpha I(t, h) + \sqrt{(\alpha I(t, h))^2 + (\varphi_A^M)^2}}\right]\end{aligned} \tag{A18}$$

$$P_M(t) = A(t) + B(t) + Z(t) + D_C(t)$$

$$I(t, z) = \eta_I I_o^+(t)\exp\left[-(k_w + k_p P_M(t) + k_d O_{CR}(t))z\right]$$

$$R_A = (1 - \xi_A)\varphi_A(t) \tag{A19}$$

$$E_A^C(t) = f_{EA}\xi_A\varphi_A(t) \tag{A20}$$

$$U_A^N(t) = \left(\frac{1 - f_{EA}}{\rho_A} - \frac{f_{EA}}{\rho_{EA}}\right)\xi_A\varphi_A(t) \tag{A21}$$

$$\varphi_{mA}(t) = m_A\left(1 - \frac{N(t)}{N(t) + k_{NA}}\right) \tag{A22}$$

Bacterial growth equations

Bacteria (osmotrophs) consume DOM and immobilize or remineralize DIN as a function of the N content of DOM. Bacteria will respire all DOC as total N availability goes to zero. Bacterial mortality is a function of DOC (i.e., energy) availability and respiration is growth associated.

$$\varphi_B(t) = \frac{\varphi_B^M O_{CL}(t)}{O_{CL}(t) + k_{OB}} \tag{A23}$$

$$\xi_B(t) = \xi_B^M \left[\frac{N(t)}{N(t) + k_{NB}} (1 - \zeta) + \zeta \right] \qquad \text{where } \zeta = \begin{cases} \dfrac{\rho_B O_{NL}(t)}{\xi_B^M O_{\mathrm{CL}}(t)} & \text{for } \dfrac{\rho_B O_{NL}(t)}{\xi_B^M O_{CL}(t)} < 1 \\ 1 & \text{for } \dfrac{\rho_B O_{NL}(t)}{\xi_B^M O_{CL}(t)} \geq 1 \end{cases} \tag{A24}$$

$$R_B(t) = (1 - \xi_B(t))\phi_B(t) \tag{A25}$$

$$U_B^N(t) = \left(\frac{\xi_B(t)}{\rho_B} - \frac{O_{NL}(t)}{O_{CL}(t)} \right) \varphi_B(t) \tag{A26}$$

$$\varphi_{mB}(t) = m_B \left(1 - \frac{\varphi_B(t)}{\varphi_B^M} \right) \tag{A27}$$

Heterotroph growth equations

Heterotrophs consume both autotrophs and bacteria using modified Holling type III growth kinetics (Holling, 1965). Heterotrophic growth can be either C or N limited, but when C limited, the excess N is excreted as ammonium. Mortality is a function of food availability.

$$\varphi_{AZ}(t) = \frac{\varphi_Z^M A(t)^2}{A(t)^2 + B(t)^2 + k_z^2} \tag{A28}$$

$$R_{ZA}(t) = \begin{cases} (1 - \xi_Z)\varphi_{AZ}(t) & \text{for } \dfrac{\xi_Z \rho_A}{\rho_Z} \leq 1 \\ \left(1 - \dfrac{\rho_Z}{\rho_A} \right) \varphi_{AZ}(t) & \text{for } \dfrac{\xi_Z \rho_A}{\rho Z} > 1 \end{cases} \tag{A29}$$

$$E_{AZ}^N(t) = \begin{cases} \left(\dfrac{1}{\rho_A} - \dfrac{\xi_Z}{\rho_Z} \right) \varphi_{AZ}(t) & \text{for } \dfrac{\xi_Z \rho_A}{\rho_Z} \leq 1 \\ 0 & \text{for } \dfrac{\xi_Z \rho_A}{\rho_Z} > 1 \end{cases} \tag{A30}$$

$$\varphi_{BZ}(t) = \frac{\varphi_Z^M B(t)^2}{A(t)^2 + B(t)^2 + k_Z^2} \tag{A31}$$

$$R_{ZB}(t) = \begin{cases} (1 - \xi_Z)\varphi_{BZ}(t) & \text{for} \quad \dfrac{\xi_Z \rho_B}{\rho_Z} \le 1 \\ \left(1 - \dfrac{\rho_Z}{\rho_B}\right)\varphi_{BZ}(t) & \text{for} \quad \dfrac{\xi_Z \rho_B}{\rho_Z} > 1 \end{cases} \tag{A32}$$

$$E_{BZ}^N(t) = \begin{cases} \left(\dfrac{1}{\rho_B} - \dfrac{\xi_Z}{\rho_Z}\right)\varphi_{BZ}(t) & \text{for} \quad \dfrac{\xi_Z \rho_B}{\rho_Z} \le 1 \\ 0 & \text{for} \quad \dfrac{\xi_Z \rho_B}{\rho_Z} > 1 \end{cases} \tag{A33}$$

$$\varphi_{mZ}(t) = m_z \left(1 - \frac{\varphi_{AZ}(t) + \varphi_{BZ}(t)}{\varphi_Z^M}\right) \tag{A34}$$

Measurement model

Here we detail the mapping from state space to observation space, or formally:

$$\mathbf{h}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p}) \to \mathbf{y}(t) \tag{A35}$$

Measurement vector

$$\mathbf{y}^{\mathrm{T}}(t) = [D_{OC}(t), P_{OC}(t), P_{ON}(t), D_{IN}(t), G(t), N_{PP}(t), B_P(t), K(t)] \tag{A36}$$

Dissolved organic carbon (μM C)

$$O_{CL}(t) + O_{CR}(t) \to D_{OC}(t) \tag{A37}$$

Particulate organic carbon (μM C)

$$A(t) + B(t) + Z(t) + D_C(t) \to P_{OC}(t) \tag{A38}$$

Particulate organic nitrogen (μM N)

$$\frac{A(t)}{\rho_A} + \frac{B(t)}{\rho_B} + \frac{Z(t)}{\rho_Z} + D_N(t) \to P_{ON}(t) \tag{A39}$$

Dissolved inorganic nitrogen (μM N)

$$N(t) \to D_{IN}(t) \tag{A40}$$

Chlorophyll *a* (μg l^{-1})

$$\frac{A(t)}{c_{\mathrm{chl}a}} \to G(t) \tag{A41}$$

Net primary productivity at depth z_l (μM C d^{-1})

$$(1 - f_{EA})\xi_A \frac{N(t)}{N(t) + k_{NA}} \frac{\varphi_A^M \alpha I(t, z_l)}{\sqrt{(\varphi_A^M)^2 + (\alpha I(t, z_l))^2}} A(t) \rightarrow N_{PP}(t, z_l) \tag{A42}$$

where z_l is the depth at which the incubation was performed.

Bacterial productivity (μM C d^{-1})

$$B_P(t) = \xi_B(t)\varphi_B(t)B(t) \tag{A43}$$

Light extinction coefficient (m^{-1})

$$k_w + k_P P_M(t) + k_d O_{CR}(t) \rightarrow K(t) \tag{A39}$$

Numerical integration and quadrature

The routines DDRIV3 and DQAGP from the SLATEC library (Vandevender and Haskell, 1982) were used for numerical integration of the differential equations and numerical quadrature of the objective function (Eq. 4), respectively. DDRIV3 employs an error controlled adjustable time-step and dynamically selects either Adams' method (for nonstiff equations) or Gear's method (for stiff equations). Because of the spiky nature of the residual vector (Fig. 3c) driving Eqs. 4, 7 and 8, both integration and quadrature routines were forced to evaluate the functions at times corresponding to observations, t_i, to insure that the numerical solution did not step over the residual correction term.

Code to generate derivatives of $\mathbf{f}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p})$ and $\mathbf{h}(\mathbf{x}(t; \mathbf{k}); \mathbf{p})$ with respect to $\mathbf{x}(t; \mathbf{k})$ and $\mathbf{p}$ for the adjoint method (Eqs. 7 and 8) was symbolically derived using ADIFOR 2.0 (Bischof *et al.*, 1992) from the source code of $\mathbf{f}(\mathbf{x}(t; \mathbf{k}), t; \mathbf{p})$ and $\mathbf{h}(\mathbf{x}(t; \mathbf{k}); \mathbf{p})$, respectively. The LINPACK singular value decomposition (SVD) routine DSVDC (Dongarra *et al.*, 1979) was used to examine the maximum and minimum singular values of the normalized Hessian matrix (Eq. 28).

REFERENCES

Aluffi-Pentini, F., V. Parisi and F. Zirillim. 1988a. Algorithm 667 SIGMA—A stochastic-integration global minimization algorithm. ACM Trans. Math. Software, *14,* 366-380 (http://www.netlib.org/toms/667).

—— 1988b. A global optimization algorithm using stochasitc differential equations. ACM Trans. Math. Software, *14,* 345-365 (http://www.netlib.org/toms/667).

Baretta-Bekker, J. G., J. W. Baretta, A. S. Hansen and B. Riemann. 1998. An improved model of carbon and nutrient dynamics in the microbial food web in marine enclosures. Aquat. Microb. Ecol., *14,* 91-108.

Barhen, J., V. Protopopescu and D. Reister. 1997. TRUST: a deterministic algorithm for global optimization. Science, *276,* 1094-1097.

Beckers, J. M. and J. C. J. Nihoul. 1995. A simple two species ecological model exhibiting chaotic behavior. Mathl. Comput. Modeling, *21,* 3-11.

Bennett, A. F. 1992. Inverse Methods in Physical Oceanography, Cambridge University Press, Cambridge.

Bergamasco, A., P. Malanotte-Rizzoli, W. C. Thacker and R. B. Long. 1993. The seasonal steady circulation of the Eastern Mediterranean determined with the adjoint method. Deep-Sea Res., *40,* 1269-1298.

Biegler, L. T., J. J. Damiano and G. E. Blau. 1986. Nonlinear parameter estimation: a case study comparison. AIChE J., *32,* 29-43.

Bischof, C., A. Carle, G. Corliss, A. Griewank and P. Hovland. 1992. ADIFOR: Generating derivative codes from Fortran programs. Scientific Prog., *1,* 11-29 (http://www-c.mcs.anl.gov/adifor/).

Boender, C. G. E., A. H. G. Rinnooy-Kan, G. T. Timmer and L. Stougie. 1982. A stochastic method for global optimization. Mathematical Programming, *22,* 125-140.

Box, M. J. 1966. A comparison of several current optimization methods, and the use of transformations in constrained problems. Comput. J., *9,* 67-77.

Brent, R. P. 1973. Algorithms for Minimization without Derivatives, Prentice-Hall, NJ, 195 pp. (http://www.netlib.org/opt/praxis).

Buckley, A. G. 1994. Algorithm 734: A Fortran 90 code for unconstrained nonlinear minimization. ACM Trans. Math. Software, *20,* 354-372 (http://www.netlib.org/toms/index.html).

Buckley, A. and A. Lenir. 1985. Algorithm 630 BBVSCG—A variable-storage algorithm for function minimization. ACM Trans. Math. Software, *11,* 103-119 (http://www.netlib.org/toms/index.html).

Burger, G., P. J. van Leeuwen and G. Evensen. 1998. On the analysis scheme in the ensemble Kalman filter. Mon. Weather Rev., *126,* 1719-1724 (http://www.nrsc.no/Modelling/).

Carroll, D. L. 1996. Chemical laser modeling with genetic algorithms. AIAA J., *34,* 338-346 (http://www.staff.uiuc.edu/~carroll/ga.html).

Chow, T., E. Eskow and R. Schnabel. 1994. Algorithm 739: A software package for unconstrained optimization using tensor methods. ACM Trans. Math. Software, *20,* 518-530 (http://www.netlib.org/toms/739).

Conway, G. R., N. R. Glass and J. Wilcox. 1970. Fitting nonlinear models to biological data by Marquardt's algorithm. Ecology, *51,* 503-507.

Corana, A., M. Marchesi, C. Martini and S. Ridella. 1987. Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. ACM Trans. Math. Software, *13,* 262-280 (http://www.netlib.org/opt/simann.f).

Courtier, P., J. Derber, R. Errico, J.-F. Louis and T. Vukicevic. 1993. Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. Tellus, *45A,* 342-357.

Crispi, G. and R. Mosetti. 1993. Adjoint estimation of aquatic ecosystem parameters. COENOSES, *8,* 11-14.

Csendes, T. 1988. Nonlinear parameter estimation by global optimization—Efficiency and reliability. Acta Cybernetica, *8,* 361-370 (ftp://ftp.jate.uszeged.hu/pub/math/optimization/index.html).

Daley, R. 1991. Atmospheric Data Analysis, Cambridge University Press, Cambridge.

Dennis, J. E., D. M. Gay and R. E. Welsch. 1981a. Algorithm 573 NL2SOL—An adaptive nonlinear least-squares algorithm. ACM Trans. Math. Software, *7,* 369-383 (http://gams.nist.gov/toms/Overview.html).

—— 1981b. An adaptive nonlinear least-squares algorithm. ACM Trans. Math. Software, *7,* 348-368 (http://www.bell-labs.com/project/PORT/).

Dongarra, J. J., C. B. Moler, J. R. Bunch and G. W. Stewart. 1979. LINPACK Users' Guide, SIAM, Philadelphia.

Evans, G. T. and M. J. R. Fasham. 1993a. Towards a model of Ocean Biogeochemical Processes, Springer-Verlag, Berlin, 350 pp.

—— 1993b. Themes in modelling ocean biogeochemical processes, *in* Towards a Model of Ocean Biogeochemical Processes, G. T. Evans and M. J. R. Fasham, eds., Springer-Verlag, Berlin, 1-19.

Evans, G. T. and J. S. Parslow. 1985. A model of annual plankton cycles. Biol. Oceanogr., *3,* 327-347.

Evensen, G. 1994. Inverse methods and data assimilation nonlinear ocean models. Physica D, *77,* 108-129 (http://www.nrsc.no/Modelling/).

Evensen, G., D. Dee and J. Schröter. 1998. Parameter estimation in dynamical models, *in* Ocean Forecasting: Conceptual Basis and Applications, N. Pinardi and J. D. Woods, eds., Springer-Verlag, Berlin.

Fasham, M. J. R., H. W. Ducklow and S. M. McKelvie. 1990. A nitrogen-based model of plankton dynamics in the ocean mixed layer. J. Mar. Res., *48,* 591-639.

Fasham, M. J. R. and G. T. Evans. 1995. The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station. Philos. Trans. R. Soc. Lond. , B., *348,* 203-209.

Fletcher, R. and C. M. Reeves. 1964. Function minimization by conjugate gradients. Comput. J., *7,* 149-154.

Franks, P. J. S., J. S. Wroblewski and G. R. Flierl. 1986. Behavior of a simple plankton model with food-level acclimation by herbivores. Mar. Biol., *91,* 121-129.

Geider, R. J., H. L. MacIntyre and T. M. Kana. 1998. A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. Limnol. Oceanogr., *43,* 679-694.

Geider, R. J. and B. A. Osborne. 1992. Algal Photosynthesis, Chapman and Hall, NY, 256 pp.

Goffe, W. L., G. D. Ferrier and J. Rogers. 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics, *60,* 65-99 (http://www.netlib.org/opt/simann.f).

Goldman, J. C., J. J. McCarthy and D. G. Peavey. 1979. Growth rate influence on the chemical composition of phytoplankton in oceanic waters. Nature, *279,* 210-215.

Grice, G. D. and M. R. Reeve. 1982. Marine Mesocosms: Biological and Chemical Research in Experimental Ecosystems, Springer-Verlag, NY.

Griewank, A. and Ph. L. Toint. 1982. Partitioned variable metric updates for large structured optimization problems. Numerische Mathematik, *39,* 119-137 (http://www.netlib.org/opt/ve08).

Gunson, J., A. Oschlies and V. Garçon. 1999. Sensitivity of ecosystem parameters to simulated satellite ocean color data using a couple physical-biological model of the North Atlantic. J. Mar. Res., *57,* 613-639.

Hall, M. C. G. and D. G. Cacuci. 1983. Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models. J. Atmos. Sci., *40,* 2537-2546.

Herstine, M. 1998. Algorithms for high-precision finite differences. Dr. Dobb's Journal, *285,* 52-58.

Holling, C. S. 1965. The functional response of predators to prey density and its role in mimicry and population regulation. Mem. Entom. Soc. Can., *45,* 1-60.

Hopkinson, Jr. C. S. and J. J. Vallino. 1995. The relationships among man's activities in watersheds and estuaries: A model of runoff effects on patterns of estuarine community metabolism. Estuaries, *18,* 598-621.

Hutchinson, G. E. 1961. The paradox of the plankton. Amer. Nat., *95,* 137-145.

Ingber, L. 1993. Simulated annealing: Practice versus theory. Mathematical and Computer Modelling, *18,* 29-57 (http://www.ingber.com/).

Ishizaka, J. 1993. Data assimilation for biogeochemical models, *in* Towards a Model of Ocean Biogeochemical Processes, G. T. Evans and M. J. R. Fasham, eds., Springer-Verlag, Berlin, 295-316.

Iwasa, Y., S. A. Levin and V. Andreasen. 1989. Aggregation in model ecosystems II. approximate aggregation. IMA Journal of Mathematics Applied in Medicine & Biology, *6,* 1-23.

Janssen, P. H. M. and P. S. C. Heuberger. 1995. Calibration of process-orientedmodels. Ecol. Model., *83,* 55-66.

Jazwinski, A. H. 1970. Stochastic Processes and Filtering Theory, Academic Press, NY, 376 pp.

Kirk, D. E. 1970. Optimal Control Theory: An Introduction, Prentice-Hall, Inc., Englewood Cliffs, NJ, 452 pp.

Kot, M., G. S. Sayler and T. W. Schultz. 1992. Complex dynamics in a model microbial system. Bull. Math. Biol., *54,* 619-648.

Kraft, D. 1994. Algorithm 733: TOMP-Fortran modules for optimal control calculations. ACM Trans. Math. Software, *20,* 262-281.

Kremer, J. N. and S. W. Nixon. 1978. A coastal marine ecosystem: Simulation and analysis, *in* Ecological Studies: Analysis and Synthesis, W. D. Billings, F. Golley, O.L. Lange and J. S. Olson, eds., Springer-Verlag, NY, 217 pp.

Lawson, L. M., E. E. Hofmann and Y. H. Spitz. 1996. Times series sampling and data assimilation in a simple marine ecosystem model. Deep-Sea Res., *43,* 625-651.

Lawson, L. M., Y. H. Spitz, E. E. Hofmann and R. L. Long. 1995. A data assimilation technique applied to a predator-prey model. Bull. Math. Biol., *57,* 593-617.

Malanotte-Rizzoli, P. and E. Tziperman. 1996. The oceanographic data assimilation problem: overview, motivation and purposes, *in* Modern Approaches to Data Assimilation in Ocean Modeling, P. Malanotte-Rizzoli, ed., Elsevier, Amsterdam, 3-17.

Marcos, B. and G. Payre. 1988. Parameters estimation of an aquatic biological system by the adjoint method. Mathematics and Computers in Simulation, *30,* 405-418.

Marsili-Libelli, S. 1992. Parameter estimation of ecological models. Ecol. Model., *62,* 233-258.

Matear, R. J. 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P. J. Mar. Res., *53,* 571-607.

McLaughlin, D. 1995. Recent developments in hydrologic data assimilation. Rev. Geophys., *33* (Suppl.), 977-984 (http://www.agu.org/revgeophys/mclaug01/mclaugh01.html).

Miller, R. N. and M. A. Cane. 1996. Tropical data assimilation: theoretical aspects, *in* Modern Approaches to Data Assimilation in Ocean Modeling, P. Malanotte-Rizzoli, ed., Elsevier, NY, 207-233.

Moloney, C. L. and J. G. Field. 1991. The size-based dynamics of plankton food webs. I. A simulation model of carbon and nitrogen flows. J. Plankton Res., *13,* 1003-1038.

Moran, M. A., T. Legovic, R. Benner and R. E. Hodson. 1988. Carbon flow from lignocellulose: a simulation analysis of a detritus-based ecosystem. Ecology, *69,* 1525-1536.

Nash, S. G. 1984. Newton-type minimization via the Lanczos method. SIAM Journal of Numerical Analysis, *21,* 770-788 (http://www.netlib.org/opt/tn).

Nelder, J. A. and R. Mead. 1965. A simplex method for function minimization. Comput. J., *7,* 308-313.

Noble, B. and J. W. Daniel. 1977. Applied Linear Algebra, Prentice-Hall, Inc., Englewood Cliffs, NJ, 477 pp.

Oguz, T., H. W. Ducklow, P. Malanotte-Rizzoli, S. Tugrul, N. P. Nezlin and U. Unluata. 1996. Simulation of annual plankton productivity cycle in the Black Sea by a one-dimensional physical-biologicalmodel. J. Geophys. Res., *101,* 16585-16599.

Pace, M. L., J. E. Glasser and L. R. Pomeroy. 1984. A simulation analysis of continental shelf food webs. Mar. Biol., *82,* 47-63.

Peterson, B., B. Fry, M. Hullar, S. Saupe and R. Wright. 1995. The distribution and stable carbon isotopic composition of dissolved organic carbon in estuaries. Estuaries, *17,* 111-121.

Pitchford, J. and J. Brindley. 1998. Intratrophic predation in simple predator-prey models. Bull. Math. Biol., *60,* 937-953.

Platt, T., K. H. Mann and R. E. Ulanowicz. 1981. Mathematical Models in Biological Oceanography, The Unesco Press, Paris, 156 pp.

Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling. 1986. Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, Cambridge, 818 pp.

Prunet, P., J.-F. Minster, D. Ruiz-Pino and I. Dadou. 1996. Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean 1. Method and preliminary results. Global Biogeochem. Cycles, *10,* 111-138.

Robinson, A. R., P. F. J. Lermusiaux and N. Q. Sloan III. 1998. Data assimilation, *in* The Global Coastal Ocean: Processes and Methods, K. H. Brink and A. R. Robinson, eds., Wiley, NY, 541-594.

Rowan, T. 1990. Functional stability analysis of numerical algorithms, University of Texas at Austin, Ph.D. Thesis, (http://www.netlib.org/opt/subplex,http://www.epm.ornl.gov/~rowan/thesis/).

Sarmiento, J. L., R. D. Slater, M. J. R. Fasham, H. W. Ducklow, J. R. Toggweiler and G. T. Evans. 1993. A seasonal three-dimensional ecosystem model of nitrogen cycling in the North Atlantic euphotic zone. Global Biogeochem. Cycles, *7,* 417-450.

Siarry, P., G. Berthiau, F. Durdin and J. Haussy. 1997. Enhanced simulated annealing for globally minimizing functions of many-continuos variables. ACM Trans. Math. Software, *23,* 209-228.

Smith, S. V. and J. T. Hollibaugh. 1993. Coastal metabolism and the oceanic organic carbon balance. Rev. Geophys, *31,* 75-89.

Smith, S. V. and F. T. Mackenzie. 1987. The ocean as a net heterotrophic system: implications from the carbon biogeochemical cycle. Global Biogeochem. Cycles, *1,* 187-198.

Steele, J. H. and E. W. Henderson. 1992. The role of predation in plankton models. J. Plankton Res., *14,* 157-172.

Stone, L. 1990. Phytoplankton-bacteria-protozoa interactions: a qualitative model portraying indirect effects. Mar. Ecol. Prog. Ser., *64,* 137-145.

Sverdrup, H. U. 1953. On conditions for the vernal blooming of phytoplankton. J. Conseil Exp. Mer., *18,* 287-295.

Taylor, A. H. and I. Joint. 1990. A steady-state analysis of the 'microbial loop' in stratified systems. Mar. Ecol. Prog. Ser., *59,* 1-17.

Thacker, W. C. and R. B. Long. 1988. Fitting dynamics to data. J. Geophys. Res., *93,* 1227-1240.

Totterdell, I. J. 1993. An annotated bibliography of marine biological models, *in* Towards a Model of Ocean Biogeochemical Processes, G. T. Evans and M. J. R. Fasham, eds., Springer-Verlag, Berlin, 317-339.

Totterdell, I. J., R. A. Armstrong, H. Drange, J. S. Parslow, T. M. Powell and A. H. Taylor. 1993. Trophic resolution, *in* Towards a Model of Ocean Biogeochemical Processes, G. T. Evans and M. J. R. Fasham, eds., Springer-Verlag, Berlin, 71-92.

Turner, J. T. and J. C. Roff. 1993. Trophic levels and trophospecies in marine plankton: lessons from the microbial food web. Mar. Microbial Food Webs, *7,* 225-248.

Tusseau, M.-H., C. Lancelot, J.-M. Martin and B. Tassin. 1997. 1-D coupled physical-biological model of the northwestern Mediterranean Sea. Deep-Sea Res., *44,* 851-880.

Tziperman, E. and W. C. Thacker. 1989. An optimal-control/adjoint-equations approach to studying the oceanic general circulation. J. Phys. Oceanogr., *19,* 1471-1485.

Vallino, J. J., C. S. Hopkinson and J. E. Hobbie. 1996. Modeling bacterial utilization of dissolved organic matter: Optimization replaces Monod growth kinetics. Limnol. Oceanogr., *41,* 1591-1609.

Vandevender, W. H. and K. H. Haskell. 1982. The SLATEC mathematical subroutine library. SIGNUM Newsletter, *17,* 16-21 (http://www.netlib.org/slatec/index.html).

Wallach, D. and M. Genard. 1998. Effect of uncertainty in input and parameter values on model prediction error. Ecol. Model., *105,* 337-345.

Wang, B.-C. and R. Luus. 1980. Increasing the size of region of convergence for parameter estimation through the use of shorter data-length. Int. J. Control, *31,* 947-972.

Wiggins, R. A. 1972. The general linear inverse problem: implication of surface waves and free oscillations for earth structure. Rev. Geophys. Space Phys., *10,* 251-285.

Received: *17 November, 1998*; revised: *6 December, 1999.*