ELSEVIER

Deep-Sea Research II 53 (2006) 576–600

DEEP-SEA RESEARCH
PART II

www.elsevier.com/locate/dsr2

# Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data

Marjorie A.M. Friedrichs[a,b,*], Raleigh R. Hood[c], Jerry D. Wiggert[a]

[a]*Center for Coastal Physical Oceanography, Old Dominion University, Crittenton Hall, 768 W 52nd Street, Norfolk, VA 23529, USA*
[b]*Virginia Institute of Marine Science, College of William and Mary, P.O. Box 1346, Gloucester Point, VA 23062, USA*
[c]*Horn Point Laboratory, Center for Environmental Science, University of Maryland, P.O. Box 775, 2020 Horn Point Road, Cambridge, MD, USA*

Received 19 October 2004; accepted 21 January 2006
Available online 19 May 2006

**Abstract**

In the wake of recent large-scale observational programs such as Joint Global Ocean Flux Study (JGOFS) and Global Ocean Ecosystems Dynamics (GLOBEC), a number of models have been developed to simulate biogeochemical cycling in various oceanographic regions; however, few quantitative comparisons of these models have been made. In order to assess critically which ecosystem structures and model formulations are best able to simulate observed biogeochemical cycling, three fundamentally different ecosystem models of varying complexity are applied within a consistent one-dimensional (1-D) framework at 15.5°N, 61.5°E in the Arabian Sea. Each model is forced by three different sets of physical forcing fields: two of these are derived from the solutions of different 3-D physical models, while the third is derived from moored observations. In situ concentrations of plankton and nitrogenous nutrients, and rates of production and export measured during the US JGOFS Arabian Sea Expedition are used for assimilation and evaluation. After objectively optimizing each model, their performance is quantitatively compared to assess which model structure best represents the fundamental underlying biogeochemical processes. Results are highly sensitive to the number of parameters optimized for each ecosystem model. A set of cross-validation experiments designed to assess predictive capability demonstrates that when too many parameters are allowed to vary, the more complex models are unable to reproduce any unassimilated data, suggesting that under these conditions the models have little predictive skill. Optimizing only an objectively and systematically selected subset of uncorrelated model parameters minimizes this problem. The method developed to accomplish this parameter selection is presented. After this optimization method is applied, all three models behave similarly, implying that the additional complexity of a multiple size-class model may not be advantageous. Furthermore, a change in physical model (mixed-layer depth and vertical velocity fields) typically produces a far greater change in biogeochemical model response than does a change in ecosystem model complexity, highlighting the fact that biogeochemical variability is largely determined by the physical environment both in situ and in the model domain.


*Keywords:* Marine ecosystem modeling; Biogeochemical model complexity; Data assimilation; Arabian Sea

*Corresponding author. Tel.: +1 757 683 5562; fax: +1 757 683 5550.
*E-mail address:* marjy@ccpo.odu.edu (M.A.M. Friedrichs).

0967-0645/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dsr2.2006.01.026

## 1. Introduction

An important legacy of the US Joint Global Ocean Flux Study (JGOFS) Synthesis and Modeling Project will be the formulation of a broad suite of models designed to simulate biogeochemical cycling at the various process and time-series study sites. These models vary from the simplest four to five state variable models (e.g., Denman and Pena, 2002; Schartau and Oschlies, 2003; Kantha, 2004) to extremely complex 20 to 30+ state variable, multi-nutrient models (e.g., Bissett et al., 1999; Moore et al., 2002; Gregg et al., 2003). Although each of these modeling studies has advanced our understanding of biogeochemical cycling within the open ocean, few quantitative comparisons of these models have been made. Generally these models use distinct forcing fields, and their performance is evaluated using different criteria (Arhonditsis and Brett, 2004). As a result, it is difficult to examine critically which ecosystem structures and formulations are most successful in a specific region, and how much complexity is required to simulate accurately major observed biogeochemical cycles. Although current dogma dictates that the simplest single size-class ecosystem models cannot adequately explain the plethora of biogeochemical observations produced by programs such as JGOFS, it has not been demonstrated that models of greater complexity will inherently produce the best estimates of bulk biogeochemical quantities and fluxes, or that they will exhibit greater predictive ability. This is because the number of parameters that must be specified from observations increases by as much as the square of the number of state variables, and quickly surpasses our ability to constrain them properly from observations (Denman, 2003).

Objectively assessing the performance of marine ecosystem models characterized by varying levels of complexity is not a straightforward task (Arhonditsis and Brett, 2004). First, the models must be forced with identical physical fields, and second, the models must be evaluated using the same biogeochemical data. Even assuming these criteria are accomplished, because model performance is largely a function of time spent tuning unconstrained parameters, it is often unclear whether a given model reproduces a given data set because of the specific structural characteristics of the model or because more time has been spent tuning the model to a particular data set. Furthermore, models including additional state variables, i.e. more functional groups, size classes, micro-nutrients, etc., inherently contain more unconstrained parameters that require a more comprehensive observational database to uniquely characterize this greater complexity. With more degrees of freedom, it is possible that a highly complex model will reproduce a data set better than its simpler counterpart; however, a model that is better able to reproduce a given data set does not necessarily have a greater predictive ability.

One method of objectively comparing the performance of models of varying complexity is to invoke formal data assimilation and parameter optimization techniques. A number of successful ecosystem modeling studies have recently applied the variational adjoint method to marine ecosystem modeling analyses (McGillicuddy et al., 1998; Spitz et al., 2001; Vallino, 2000; Friedrichs, 2001; Schartau et al., 2001; Fennel et al., 2001; Friedrichs, 2002). In its simplest form, the variational adjoint method is a particularly efficient method that approximates a nonlinear weighted least-squares analysis in which the misfit between the model solution and the observations is minimized and an optimal model parameter set is recovered. This method is particularly well suited for marine ecosystem models, which all inherently contain large numbers of poorly known parameters governing processes such as plankton growth, mortality, nutrient uptake, and remineralization. By applying this data assimilation technique to models of varying complexity, it is possible to minimize the inherently confounding problems associated with subjective tuning.

The 1-D model comparison effort described here is conducted at a site within the Arabian Sea. The Arabian Sea is an excellent location to test a variety of ecosystem models because the region experiences seasonal extremes in forcing and biological response that alternate from calm, stratified, near-oligotrophic conditions during the Intermonsoon periods (March–May and October–November) to strongly forced, eutrophic conditions during the Southwest (June–September) and Northeast (November–February) monsoons (Smith et al., 1998). The two monsoons themselves generate distinctly different biological responses. During the Southwest Monsoon (SWM), coastal upwelling produces high biological production with chlorophyll-rich filaments extending several hundred kilometers off the Omani coast (Gundersen et al., 1998; Manghnani et al., 1998; Lee et al., 2000). In contrast, during the Northeast Monsoon (NEM) convectively driven

nutrient entrainment enhances offshore production (Lee et al., 2000; Wiggert et al., 2000). Thus the Arabian Sea provides a demanding test for ecosystem models, since they must be able to capture the full gamut of pelagic ecosystem behavior and respond appropriately to dramatic seasonal variations in light levels and nutrient supply.

The JGOFS Arabian Sea Process Study data set is arguably the most comprehensive of all the JGOFS regional data sets. The analysis presented here utilizes biogeochemical observations from station S7 (16°N, 62°E). These data were collected on six different cruises throughout 1995 that cover all four monsoon/intermonsoon periods (Smith et al., 1998). This analysis also makes use of particle flux time series measured at a nearby sediment trap (MS-4, 16°N, 61.5°E, Honjo et al., 1999). In addition, as part of a companion effort sponsored by ONR, a mooring (henceforth referred to as the 'WHOI mooring') was placed at 15.5°N, 61.5°E, 80 km southwest of S7 (Weller et al., 1998). Its instrumentation provides the simultaneous surface irradiance and in situ temperature time series required as physical forcing fields for the ecosystem models. Finally, solutions from multiple basin-scale physical models are available for this region, which provide estimates of mixed-layer depth and vertical velocity.

This study compares the nine modeling realizations that arise through the utilization of three different sets of physical forcing fields and three ecosystem models of varying complexity.

## 2. Methods

### 2.1. Ecosystem models

In this analysis, simulations from three different ecosystem models are compared. Each of these models is well documented in the literature, and therefore only their general characteristics and appropriate references are provided.

The simplest model, EM4, is a four-component (phytoplankton, zooplankton, dissolved inorganic nitrogen (DIN) and detritus) ecosystem model that represents a classic diatom–mesozooplankton system. Unlike the other two models, EM4 has been developed and calibrated specifically for the Arabian Sea (McCreary et al., 1996, 2001; Hood et al., 2003).

The second model, EM5, is a five-component (phytoplankton, heterotrophs, DIN, dissolved organic nitrogen (DON) and detritus) ecosystem model developed by Hood et al. (2001) for use at the Bermuda Atlantic Time Series station (BATS) and subsequently applied over the tropical and subtropical Atlantic (Hood et al., 2004; Coles et al., 2004). Although this model was originally developed with a sixth diazotrophic state variable, here the model is implemented without this component. The heterotroph compartment is considered to represent the sum of all heterotrophic processes that are facilitated by bacteria, microzooplankton and mesozooplankton. This model emphasizes the microbial loop by having all organic matter cycle through the heterotroph compartment at relatively high rates.

The most complex ecosystem model applied here, EM8, is an eight-component model containing two size classes of phytoplankton, zooplankton and detritus, as well as ammonium and nitrate. This model was originally developed for use in the equatorial Pacific (Christian et al., 2002), but has more recently been adapted for use in the Indian Ocean (Wiggert et al., 2006). In this study the original version of the model is used with one exception: in order to simplify the model comparison, the iron model component is turned off. With the multiple plankton size classes included in this model, EM8 can be basically thought of as a combination of EM4 and EM5.

### 2.2. Physical forcing fields

Time series of photosynthetically active radiation (PAR), temperature, vertical diffusivity, vertical velocity and mixed-layer depth (MLD) are required to run the 1-D ecosystem models. Identical PAR, temperature and vertical diffusivity time series were used for all model runs. PAR and temperature time series were obtained from the Weller mooring data (Weller et al., 1998; McCreary et al., 2001). Vertical diffusivity was computed using the velocity and temperature data from the WHOI mooring (Pacanowski and Philander, 1981).

The ecosystem models also were forced by three different sets of mixed-layer depth and vertical velocity time series. The Forcing 1 (F1) MLD and vertical velocity were derived from WHOI mooring data. Daily averaged MLD was estimated as the depth at which temperature is 1.0 °C lower than the near-surface temperature (Dickey et al., 1998; McCreary et al., 2001). Below the bottom of the mixed layer, vertical velocities were estimated from

biweekly averaged isotherm displacements; within the mixed layer vertical velocity was set to zero.

The Forcing 2 (F2) MLD and vertical velocity time series were derived from a reduced-gravity circulation model (McCreary et al., 2001). This model consists of four active layers (the surface mixed layer, the diurnal thermocline, the seasonal thermocline and the main thermocline) overlying a quiescent, deep ocean where pressure gradients are assumed to vanish. In this model the mixed-layer thickness is determined using a Kraus–Turner mixed-layer model (Kraus and Turner, 1967) that considers the effects of both buoyancy mixing and wind-induced shear instability. From October 1994 through October 1995, daily mean data from the WHOI mooring are blended into the forcing fields. In addition, diurnal variability is introduced by allowing incoming solar radiation to have a realistic diurnal cycle.

The Forcing 3 (F3) MLD and vertical velocity time series were obtained from a reduced-gravity, primitive equation ocean model (Murtugudde et al., 1996; Murtugudde and Busalacchi, 1998) with a variable depth mixed layer overlying 19 sigma layers. In this model mixed-layer thickness is determined using a "hybrid" mixed-layer model (Chen et al., 1994) that considers both wind stirring and shear instability. The thickness of each of the remaining 19 layers is a constant fraction of the total vertical distance between the base of the mixed layer and bottom of the model domain. Thus whereas the F3 fields were derived from a model with relatively high vertical resolution, the F2 fields were derived from a model with relatively high temporal resolution that resolves the diurnal cycle.

The F1–F3 MLD time series all show similar trends throughout 1994–1995 (Fig. 1): a pronounced deepening of the MLD during the winter NEM and the summer SWM, and shallow MLDs during the Intermonsoonal periods. The F1 and F3 MLDs, however, diverge in November 1994 and July/August 1995, with F3 producing significantly deeper MLDs in both instances. The F2 MLDs are closer to the mooring-derived estimates, which is due to a combination of tuning (Hood et al., 2003) and the application of mooring-derived surface forcing data. These two time-periods were dominated by prominent mesoscale eddy activity (Dickey et al., 1998; Fischer et al., 2002), which caused the mixed layer in this region to shoal significantly. Neither of the physical models used in this study have the ability to capture physical perturbations

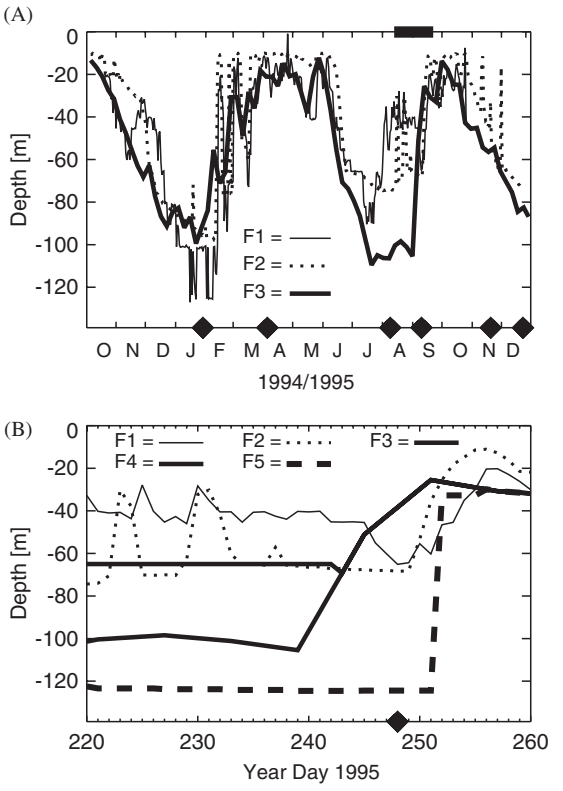


Fig. 1. (A) Time series of MLD obtained from mooring data (F1), the McCreary et al. (2001) model (F2) and the Murtugudde and Busalacchi (1998) model (F3). Black diamonds represent the times of the six JGOFS Arabian Sea Process Study cruises. The black bar on the upper *x*-axis represents the time period of the time series shown in (B). (B) As in (A), but zoomed in to illustrate the detail surrounding the ttn-050 cruise (black diamond on *x*-axis). Two additional time series (F4 and F5) are shown here for comparison (see text of Section 4.2).

associated with these mesoscale events, which may account for these discrepancies.

### 2.3. Model implementation

The ecosystem models are all solved using a second-order Runge–Kutta scheme. Vertical advection is computed using a simple centered, finite-difference, gradient-form advection scheme and vertical diffusion is applied using a Crank–Nicholson vertically variable diffusion operation, with a closed upper boundary and an open bottom boundary. The effect of the horizontal advection of biological quantities was examined through a scaling analysis using output from two 3-D coupled biological–physical models (Hood et al., 2003; Wiggert et al., 2006). In this region, horizontal advection was found

to be small in comparison to vertical advection and other biological source/sink terms, and was thus neglected.

At each time step, all state variables are mixed down to the bottom of the mixed layer. A modified second-order flux-corrected transport scheme is applied to calculate detrital sinking. A uniform sinking velocity of $32\,m\,day^{-1}$ is applied to all detrital components with the exception of the small detritus component of EM8, which sinks at a much slower $0.3\,m\,day^{-1}$. The subsurface light field is computed using the method of Anderson (1993): the water column is divided into three layers, and vertical attenuation coefficients for each layer are determined from a fifth-order polynomial fit. The attenuation coefficients are proportional to the square root of the average pigment in each layer (Friedrichs and Hofmann, 2001.) The models are run from 1 October 1994 through 1 January 1996 with a time step of 1 h and a depth resolution of 10 m.

### 2.4. Biogeochemical data

Five distinct data types are assimilated into the ecosystem models: phytoplankton chlorophyll, zooplankton biomass, DIN, primary production, and export flux. Data from six Arabian Sea Process Study cruises were downloaded from the US JGOFS website: http://usjgofs.whoi.edu/jg/dir/jgofs/arabian. These data cover each season of 1995: mid-NEM (ttn-043: 1/8–1/31); Spring Intermonsoon (ttn-045: 3/14–4/7); mid-SWM (ttn-049: 7/18–8/12); late SWM (ttn-050: 8/18–9/12); Fall Intermonsoon (ttn-053: 10/29–11/25); and early NEM (ttn-054: 11/30–12/26). For each cruise, only data from Station S7 (16.0°N, 62.0°W) are utilized; note, however, that for each data type, observations are not necessarily available from each cruise. In addition to cruise data, particulate nitrogen export flux measurements are utilized from the 800 m sediment trap located at 16.0°N, 61.5°W (Honjo et al., 1999).

Chlorophyll-*a* (measured fluorometrically) and primary production data (24-h in situ incubations) are available at S7 for five of the six cruises (Barber et al., 2001). Total zooplankton carbon ($>200\,\mu m$) data were collected using a double $1\,m^2$ multiple opening–closing net system (MOCNESS) for four of the six cruises (Wishner et al., 1998). For comparison with the nitrogen-based model simulations, a C:N conversion factor of 6.625 was assumed. Total DIN data ($NO_3+NO_2+NH_4$) were computed from data for all six cruises (Morrison et al., 1998). When multiple nutrient profiles were available for the same day, these were averaged prior to assimilation. All data were also interpolated to the model grid, starting at 15 m for zooplankton and 5 m for chlorophyll and nutrients.

In initial experiments in which DIN data were assimilated over the entire model domain (0–150 m), unreasonable optimized ecosystem parameter estimates were obtained. This outcome was a direct result of the assimilation attempting to compensate for the physical models' inability to capture the sharp nutricline because of their overestimated mixing and diffusion. Thus only mixed-layer DIN concentrations were assimilated so the biogeochemical assimilation scheme would not attempt to compensate for this physics-specific shortcoming. The remaining cruise data (primary production, chlorophyll and zooplankton concentrations) were assimilated over the full depth to which data were available.

### 2.5. Variational adjoint method

The variational adjoint method of data assimilation is used to determine the optimal ecosystem parameter values, such that the differences between the model solution and the observations are minimized. This method consists of: (1) a numerical model; (2) a measure of the misfit between the predicted and observed variables (i.e. the cost function); (3) an adjoint, or backwards model, which is used to compute the gradient of the cost function with respect to the subset of model parameters that will be adjusted, also called the model control variables; and (4) an optimization procedure that uses this information to determine the direction and amount by which the control variables must be modified in order to minimize the cost function.

After initial estimates of the control variables are made, the numerical model is run in order to obtain a value of the cost function. The technique developed by Lawson et al. (1995) is used to construct the adjoint code directly from the model code by means of Lagrange multipliers. The adjoint of the model is then run backward in time to find the gradient of the cost function with respect to the model control variables. Values of these gradients are passed to a variable-storage quasi-Newton

optimization procedure (Gilbert and Lemarechal, 1989), which computes the optimal direction towards the minimum of the cost function, and the optimal step size in that direction. New values of the control variables closer to the minimum of the cost function are found, and the model is rerun. This procedure is carried out in an iterative manner until a specified convergence criterion, generally based on the norm of the gradient of the cost function, has been satisfied. Uncertainties in the recovered values of the model control variables are computed from a finite difference approximation of the Hessian matrix, i.e. the matrix of the second derivatives of the cost function with respect to the control variables. When computed at the minimum of the cost function, the inverse of the Hessian matrix can be used to estimate not only the errors for the optimal parameter estimates, but also parameter correlations and the sensitivities of the cost function to each parameter (Tziperman and Thacker, 1989; Matear, 1995).

### 2.5.1. Cost function

The general cost function, $J$, is a measure of the misfit between the predicted variables ($a_{ij}$) and the observed variables ($\hat{a}_{ij}$), and can be expressed as a weighted sum of squares:

$$J = \frac{1}{N}\sum_{i=1}^{M}\sum_{j=1}^{T} W_{ij}^2 (a_{ij} - \hat{a}_{ij})^2.$$

The sums are carried out over the number ($T$) of time-steps as well as the number of variables ($M = 5$) for which observations are available: the rate of primary production, phytoplankton chlorophyll, zooplankton biomass, DIN, and sediment trap export flux. The normalization factor ($N = 180$) represents the total number of observations assimilated.

In this formulation, the weights ($W_{ij}$) are defined to be inversely proportional to the measurement uncertainty. These uncertainties must include not only the accuracy of the observations, but also the confidence we have that our modeled quantities truly represent the data. In order to lessen the subjectivity and difficulties associated with individually assigning these uncertainties to each type of measurement, a reasonable alternative is to define the uncertainty in all cases to be proportional to the range of observational variability. Specifically, in each case, the measurement uncertainty is chosen to be 25% of the standard deviations ($\sigma_i$) of the observations:

$$W_{ij} = \frac{4C_{ij}}{\sigma_i},$$

where $\sigma = 18\,\mathrm{mg\,C\,m^{-3}\,d^{-1}}$ for primary production, $\sigma = 0.22\,\mathrm{mg\,chl\,m^{-3}}$ for phytoplankton chlorophyll, $\sigma = 0.4\,\mathrm{mmol\ C\,m^{-3}}$ for zooplankton biomass, $\sigma = 2.4\,\mathrm{mmol\ N\,m^{-3}}$ for DIN, and $\sigma = 1.4\,\mathrm{mg\ N\,m^{-2}\,d^{-1}}$ for sediment trap export flux. The constant scaling factor (4) is included in order to ensure that the magnitudes of the weights are reasonably representative of the uncertainties associated with the observations, and has no effect on the intercomparison of different $J$ values. $C_{ij}$ acts as a data availability switch: if data are available $C_{ij} = 1$, if not $C_{ij} = 0$. Thus the values $W_{ij}^{-1}$ roughly represent a significance threshold for the magnitude of $a_{ij} - \hat{a}_{ij}$. Defined in this way, two values of $J$ are not significantly different if they vary by less than one.

As expected, the magnitudes of the cost functions are very sensitive to the choice of weights. However, the relative magnitudes of the post-assimilation cost functions obtained for the various models studied here were generally independent of our choice of weights. For example, experiments using a 50% lower weight for zooplankton resulted in very little change in the relative performance of our models.

Experiments also were conducted to examine the effect of imposing upper and lower bounds on the range of allowed parameter values as penalty terms in the cost function equation. These experiments revealed that such penalties resulted in the optimized parameter values either being unaffected by the restriction (if the weight given to that penalty was low) or obtaining the maximum or minimum value of that range (if the weight given to that penalty was high). As a result of this behavior, and in order to ensure a maximum degree of freedom in the model control variables, the penalty terms in the cost function were removed and were not included in any of the results described in the following section. Unrealistic parameter values were not obtained in any of our final model runs, which we attribute to the careful selection of the model control variables.

### 2.5.2. Predictive cost function

Although the cost function is a useful tool for examining model-data misfit for a particular data set, data assimilative models additionally need to be validated against independent, unassimilated data.

Here cross-validation experiments are carried out in which three seasons of data are assimilated, and the resulting parameter estimates are used to determine the misfit (cost function, $J$) for the remaining season of data. This experiment is conducted four times, where each time a different season of data is withheld. The resulting four individual cost functions (all computed using a normalization factor of $N = 180$) are summed to obtain one overall cost function, which is defined as the 'predictive cost function', or $J_P$. Thus whereas $J$ represents how well a model reproduces a given data set, $J_P$ reveals how well a model reproduces unassimilated data. In this analysis, this is interpreted as a measure of a given model's predictive ability.

#### 2.5.3. Assimilation experiments

Two assimilation experiments are conducted with each of the nine combinations afforded by the three ecosystem models (EM4, EM5, EM8) and the three physical forcing fields (F1–F3). In Experiment 1, all ecosystem model parameters are optimized for each model. This consists of 10 parameters for EM4, 16 parameters for EM5, and 19 parameters for EM8. The resulting values of $J$ and $J_P$ are computed.

In Experiment 2, only an objectively and systematically chosen subset of parameters is optimized for each of the nine modeling realizations. Beginning with the results from Experiment 1 where all parameters are optimized, the sensitivities of the cost functions to each model parameter and the correlations between each pair of parameters are computed from the inverse of the Hessian matrix (Matear, 1995.) The parameter to which the cost function is least sensitive, or in other words the optimized parameter with the greatest normalized uncertainty, is then fixed to its original value, and another assimilation simulation is conducted with one fewer control variables. After again computing $J$, $J_P$ and the corresponding Hessian matrix, the parameter to which the cost function is least sensitive is again held fixed, and another assimilation simulation is conducted with one fewer control variables. This process is continued until there are no remaining control variables. The resulting values of $J$ and $J_P$ are examined, and the 'optimal' simulation used for Experiment 2 is determined to be the simulation characterized by the minimum value of $J_P$. This technique provides an objective method for determining an optimized parameter subset that maximizes predictive ability by sequentially removing the optimized parameters that are associated with the greatest uncertainties. In addition, by default this method sequentially removes parameters that are highly correlated to other optimized parameters. Thus in each case the optimized parameter subset consists of a relatively uncorrelated (typically having correlation coefficients less than 0.8) set of parameters to which the cost function is most sensitive. Note that although the minimum value of $J$ will correspond to the maximum number of control variables for a given model, the minimum value of $J_P$ will correspond to a much smaller number of control variables, and will depend not only on the specific ecosystem model chosen, but also on the forcing field used.

## 3. Results

### 3.1. No assimilation

Initial simulated distributions of phytoplankton chlorophyll (Fig. 2) and zooplankton biomass (Fig. 3) obtained from the nine model combinations (EM4, EM5, EM8 combined with F1–F3) indicate that prior to assimilation, the results are more sensitive to the choice of ecosystem model than the choice of physical forcing. For example, independent of which forcing fields are applied, chlorophyll distributions obtained using EM4 generally show low concentrations throughout the study period, with only a few short-lived (less than 3 weeks) blooms associated with the shoaling of the mixed-layer at the end of the NEM and the SWM. In contrast, chlorophyll distributions obtained with EM5 show relatively high concentrations throughout 1995, with blooms typically lasting 2 months or longer. EM8 produces chlorophyll concentrations similar to those of EM5; however, the EM8 blooms are less persistent, typically lasting only 1–3 weeks. Deep-chlorophyll maxima are produced by both EM5 and EM8 during the Spring Intermonsoon, again independent of which physical forcing fields are used, while for EM4 deep chlorophyll maxima never appear.

Given the low chlorophyll concentrations produced by EM4, it is not surprising to see that EM4 also produces consistently low concentrations of zooplankton biomass (Fig. 3). These low concentrations (generally less than $0.8\,\mathrm{mmol\,C\,m^{-3}}$) are in contrast to those produced by EM5, which typically exceed $0.8\,\mathrm{mmol\,C\,m^{-3}}$. Zooplankton biomass

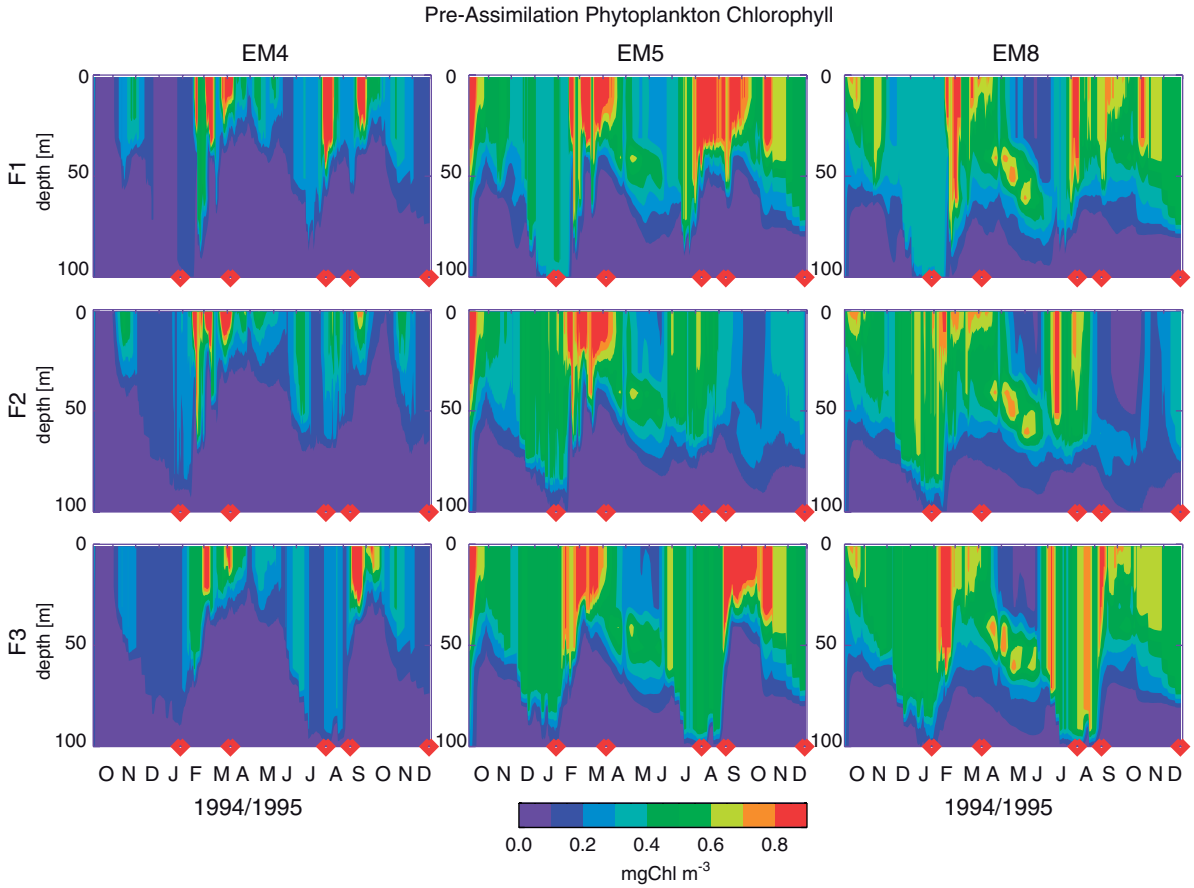


Fig. 2. Pre-assimilation distributions of phytoplankton chlorophyll for the nine modeling realizations obtained using the three ecosystem models (EM4, EM5, EM8) and the three forcing fields (F1–F3). Red diamonds on the *x*-axis denote times of chlorophyll data assimilation.

produced by EM8 is highly variable, with very low surface concentrations during the Spring Intermonsoon, and relatively low concentrations during the early NEM and SWM time periods. Both EM5 and EM8 produce deep subsurface maxima of zooplankton biomass during the SIM. These results are all independent of which physical forcing fields are applied.

The pre-assimilation cost function ($J_0$) can be used to compare quantitatively overall model-data misfit for the multiple combinations of models and forcing fields (Table 1). When F1 is used, the lowest cost function is obtained with EM5 ($J_0 = 19.9$), EM4 produces the lowest cost function when F2 is used ($J_0 = 22.6$), and EM8 produces the lowest cost function when F3 is used ($J_0 = 11.4$). Furthermore, for any given ecosystem model, the lowest cost functions are consistently produced with F3.

### 3.2. Assimilation Experiment 1: optimization of all control variables

In the first assimilation experiment all possible ecosystem parameters are included as control variables in the adjoint analysis. Again, there are 10 control variables for EM4, 16 for EM5, and 19 for EM8.

#### 3.2.1. Experiment 1: post-assimilation distributions

The resulting simulated distributions of phytoplankton chlorophyll and zooplankton biomass (Figs. 4 and 5) are strikingly different from the analogous pre-assimilation distributions (Figs. 2 and 3). Whereas the pre-assimilation distributions primarily depend on which ecosystem model is implemented, the optimized, post-assimilation distributions are strongly dependent on which forcing

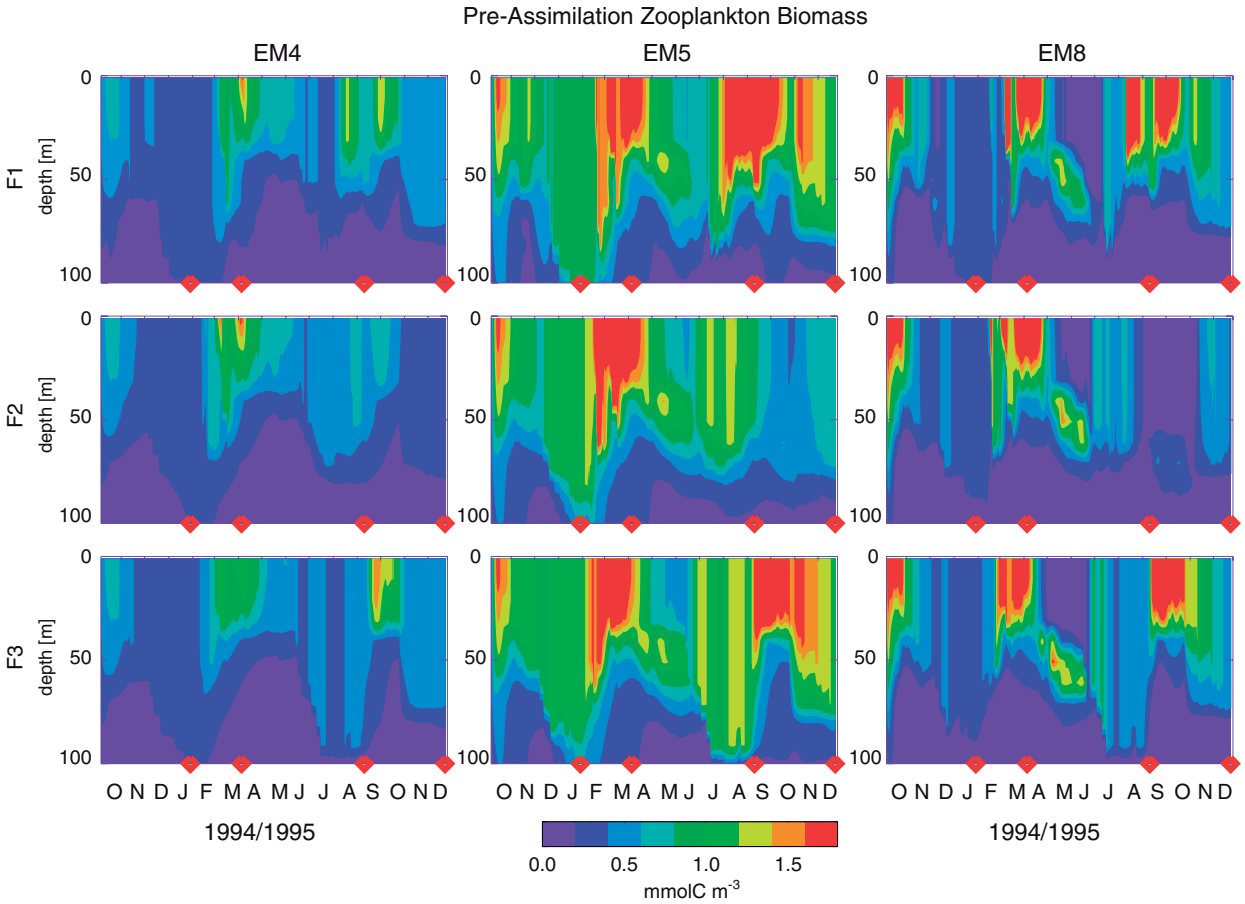


Fig. 3. As in Fig. 2, but for zooplankton biomass.

is used. For example, all three models forced by F3 produce phytoplankton concentrations of 0.4–0.5 mg chl $m^{-3}$ during both the NEM and the SWM. During the Spring Intermonsoon all three models also produce deep chlorophyll maxima with similar magnitudes that are located at similar depths.

Changing the forcing fields has a dramatic effect on the simulated distributions. The 50-m-deep chlorophyll maximum that was evident during the Spring Intermonsoon for EM5 and EM8 prior to assimilation is no longer present for F2, but does appear for each of the ecosystem models when F1 is applied. For all three models the application of F2 yields spurious oscillations of phytoplankton and zooplankton concentrations that are most clearly evident for EM4. These oscillations are at least partly due to the discrete nature (4–6 snapshots) of the assimilated data; presumably the assimilation of higher resolution time series (e.g., ocean color or bio-optical) would result in an oscillatory solution being severely penalized.

Post-assimilation zooplankton distributions are also highly dependent on which forcing fields are applied. The strong peaks in zooplankton biomass, apparent for the EM5 F1 model combination do not exist when alternate forcing fields are used. For EM8, F3 produces much higher concentrations of zooplankton biomass than does F1. As is the case for the chlorophyll distributions, the application of F2 to EM4 results in spurious oscillations of zooplankton biomass. These post-assimilation results indicate that the optimized models are more strongly dependent on the choice of physical forcing fields than were the original model simulations. In addition, the post-assimilation distributions are typically a stronger function of the physical forcing fields than of ecosystem model complexity.

Table 1
Model data misfit with no assimilation ($J_0$), number of control variables, cost function ($J$) and predictive cost function ($J_P$), for three models of varying complexity forced by three different sets of physical fields. Results from both Experiments 1 and 2 are shown

| Forcing | Model | $J_0$ | Experiment #1 | | | Experiment #2 | | |
|---|---|---|---|---|---|---|---|---|
| | | | # Control variables | $J$ | $J_P$ | # Control variables | $J$ | $J_P$ |
| F1 | EM4 | 24.0 | 10 | 10.8 | 21.7 | 5 | 17.1 | 17.5 |
| F1 | EM5 | 19.9 | 16 | 11.3 | 20.0 | 2 | 14.7 | 19.6 |
| F1 | EM8 | 22.7 | 19 | 11.8 | 20.0 | 5 | 12.9 | 19.0 |
| F2 | EM4 | 22.6 | 10 | 9.6 | 20.7 | 6 | 10.9 | 19.9 |
| F2 | EM5 | 25.6 | 16 | 11.8 | 29.5 | 4 | 15.3 | 21.3 |
| F2 | EM8 | 26.2 | 19 | 13.4 | 32.5 | 2 | 18.0 | 18.5 |
| F3 | EM4 | 20.8 | 10 | 6.7 | 10.9 | 3 | 8.9 | 9.8 |
| F3 | EM5 | 15.2 | 16 | 5.1 | 12.4 | 4 | 9.8 | 12.4 |
| F3 | EM8 | 11.4 | 19 | 7.2 | 18.5 | 4 | 8.7 | 10.2 |

For Experiment 1, all control variables within each model are estimated via the variational adjoint method. For Experiment 2, an objectively chosen subset of uncorrelated model parameters is estimated. Uncertainties are $\pm 0.5$.

*3.2.2. Experiment 1: post-assimilation model-data fit*

The assimilation process improves the model-data fit for all assimilated fields to varying degrees: chlorophyll (Fig. 6), zooplankton biomass (Fig. 7), nitrate concentration (Fig. 8), primary production (Fig. 9), and export flux (Fig. 10). Post-assimilation model-data fit for chlorophyll (Fig. 6) is excellent, especially for F3. A dramatic improvement in model-data fit for zooplankton is also apparent (Fig. 7). Because the physical model framework does not allow vertical gradients of zooplankton concentration within the mixed layer, the model-data fits for zooplankton are not expected to be substantially better than those obtained.

Nitrate concentrations are only slightly altered as a result of the assimilation. Pre-assimilation model-data fits within the mixed layer are generally excellent, and thus there was little room for improvement via assimilation. In early 1995 (ttn-043 and ttn-045) the models all produce a nutricline that is weaker than that observed (Fig. 8). Throughout the remainder of the year, the models all produce a sharper nutricline in better agreement with the observations, but they overestimate its depth. As a result, model-data agreement between 50 and 100 m is poor. As discussed in Section 2.4, this is primarily a problem with the physical forcing fields, which no change in any ecosystem parameters can remedy.

Differences between the chlorophyll, zooplankton and nitrate profiles obtained using the three different ecosystem models are greatly reduced as a result of the assimilation. For example, pre-assimilation F3 chlorophyll profiles for cruise ttn-050 range from 0.4 to 1.1 mg Chl $m^{-3}$, whereas the analogous post-assimilation profiles are nearly identical and are independent of which ecosystem model is used (Fig. 6). This is also the case for F1; however, while the F3 results reproduce the data quite closely, the F1 results for this cruise all underestimate the assimilated data by nearly a factor of 2. The assimilation also dramatically decreases the disparity between the zooplankton results of the various ecosystem models (Fig. 7). For example, pre-assimilation F3 zooplankton surface concentrations for the time frame of cruise ttn-050 range from 0.50 to 1.76 mmol C $m^{-3}$. After assimilation, the modeled values differ by only 0.3 mmol C $m^{-3}$. In general, the post-assimilation zooplankton profiles, as well as the post-assimilation nitrate profiles, are relatively independent of which forcing or ecosystem model is implemented.

Pre-assimilation primary productivity profiles are generally lower than the observations, especially within the upper 25 m (Fig. 9). The post-assimilation EM4 profiles are nearly as high as observed, but the EM8 profiles always underestimate the productivity observations, despite the assimilation of these data. As is the case for the other fields, pre-assimilation export flux time series (Fig. 10) are relatively independent of which forcing is used, while the post-assimilation time series are very sensitive to the forcing choice. F1 yields relatively similar time series for all the ecosystem models, but this is not the case for the other forcing fields. The oscillations visible in the phytoplankton and zoo-

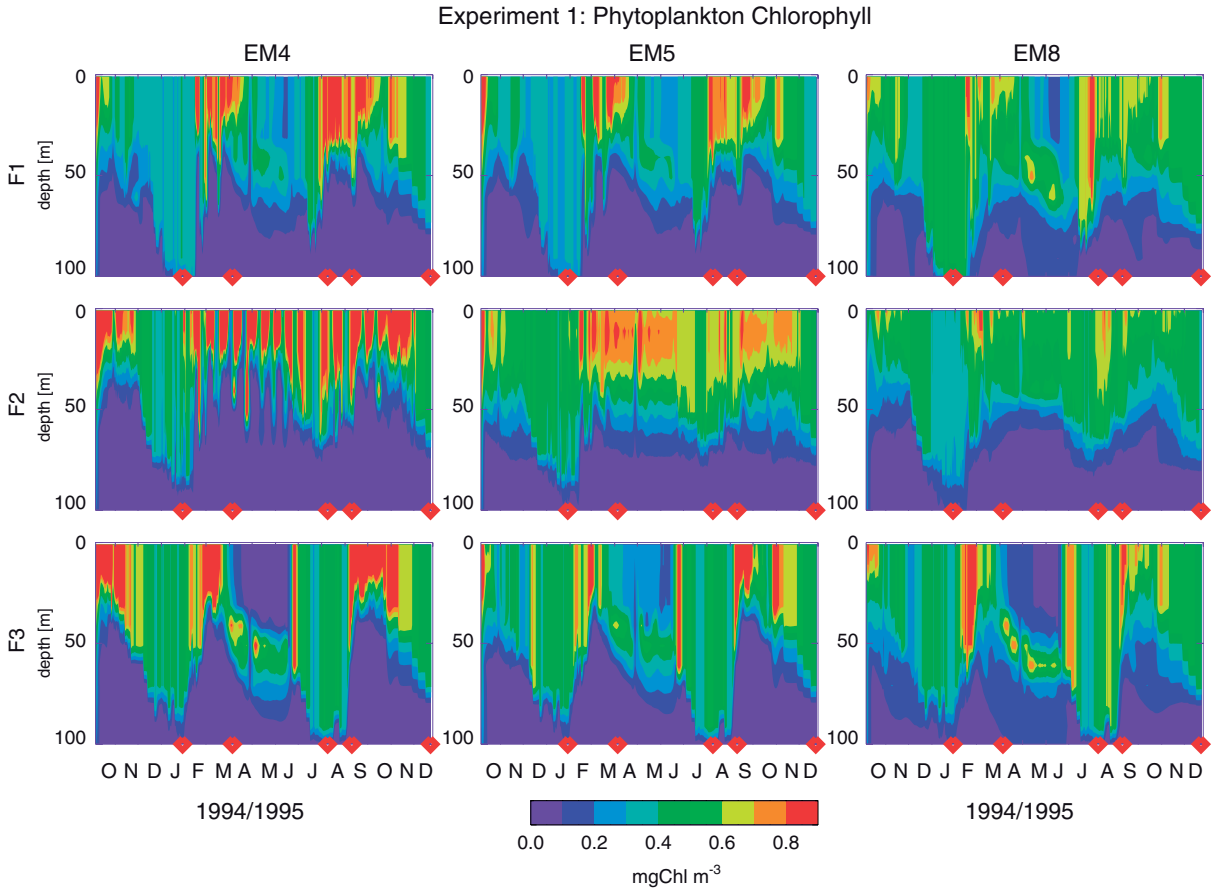


Fig. 4. As in Fig. 2, but for the post-assimilation (Experiment 1) distributions of phytoplankton chlorophyll.

plankton distributions for the EM4 F2 combination (Figs. 4 and 5) result in a similarly unrealistic oscillatory export time series. The negative export fluxes obtained with the EM5 F3 combination are even more disturbing.

### 3.2.3. *Experiment 1: cost functions*

Post-assimilation model-data misfits are quantified by the cost function, $J$ (Table 1). The assimilation dramatically reduces the cost function (by 37–68%) for each model/forcing combination. As was the case for the pre-assimilation results, for any given ecosystem model post-assimilation cost functions are lowest for F3. When either F1 or F2 is used EM4 produces the lowest cost function, whereas EM5 produces the lowest cost function when F3 is used. For any given forcing, however, the values of $J$ obtained for each ecosystem model are generally very similar, and in some cases are within the $\pm 0.5$ uncertainty estimates.

As discussed in Section 2.5.2, a more rigorous test of the models involves examining how well each can reproduce unassimilated data. Whereas the cost function, $J$, quantifies how well each model reproduces assimilated data, the predictive cost function, $J_P$, is defined to represent how well each model reproduces unassimilated data and is computed through a series of cross-validation model runs. As one would expect, values of $J_P$ are always greater than $J$ (Table 1), since reproducing unassimilated data is clearly a more challenging test of a model than is simply reproducing assimilated data. In this experiment, however, not only are values of $J_P$ greater than $J$, but they are also often greater than the analogous values of $J_0$, indicating that the original models without assimilation often reproduce the unassimilated data better than if a subset of the data are assimilated. In these cross-validation experiments, the inability of these models to reproduce unassimilated data suggests that they

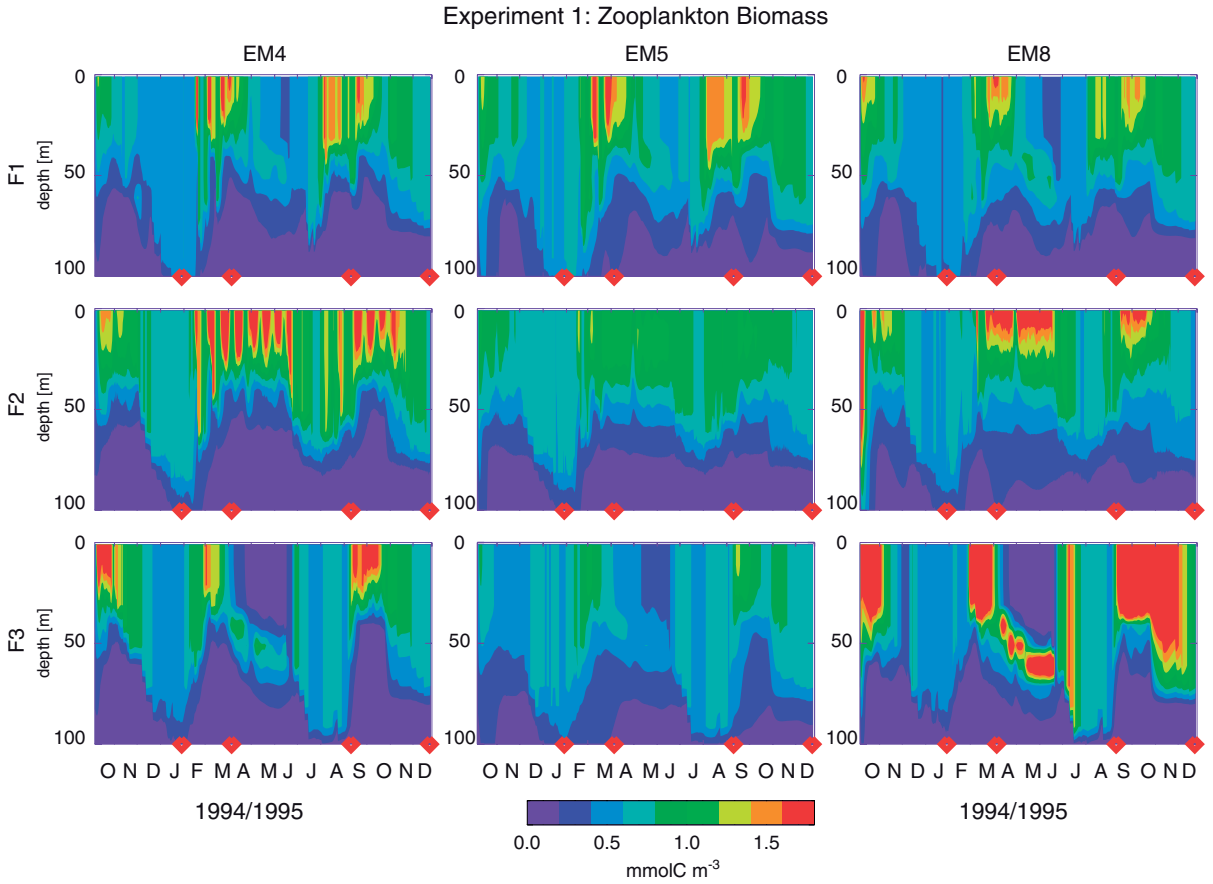


Fig. 5. As in Fig. 2, but for the post-assimilation (Experiment 1) distributions of zooplankton biomass.

have low predictive ability. Overall, the most complex model, EM8 with 19 control variables, produces the largest values of $J_P$, whereas the simplest model, EM4 with 10 control variables, always produces values of $J_P$ that are less than the analogous $J_0$ values.

### 3.3. Experiment 2: optimization of control variable subset

In the second assimilation experiment an objectively and systematically chosen subset of ecosystem parameters (Table 1) are selected as control variables in the adjoint analysis (see Section 2.5.3).

#### 3.3.1. Experiment 2: post-assimilation distributions

The resulting simulated distributions of phytoplankton chlorophyll (Fig. 11) and zooplankton biomass (Fig. 12) are very similar to those obtained in Experiment 1 (Figs. 4 and 5). One of the most noteworthy differences lies in the magnitude of the oscillatory behavior for F2. Results from Experiment 2 still show oscillations for EM4; however, optimizing only a subset of uncorrelated ecosystem parameters clearly dampens the magnitude of these oscillations, as compared to Experiment 1. In general, Experiment 2 produces slightly reduced phytoplankton and zooplankton concentrations for EM4 and increased concentrations for EM5 as compared to the results for Experiment 1. EM8 simulations show very few differences between the two experiments.

#### 3.3.2. Experiment 2: post-assimilation model-data fit

The second assimilation experiment produces phytoplankton and zooplankton distributions that fit the data nearly as well as those from the first assimilation experiment. Profiles (not shown) of phytoplankton and zooplankton are very similar to those shown in Figs. 6 and 7, while profiles of

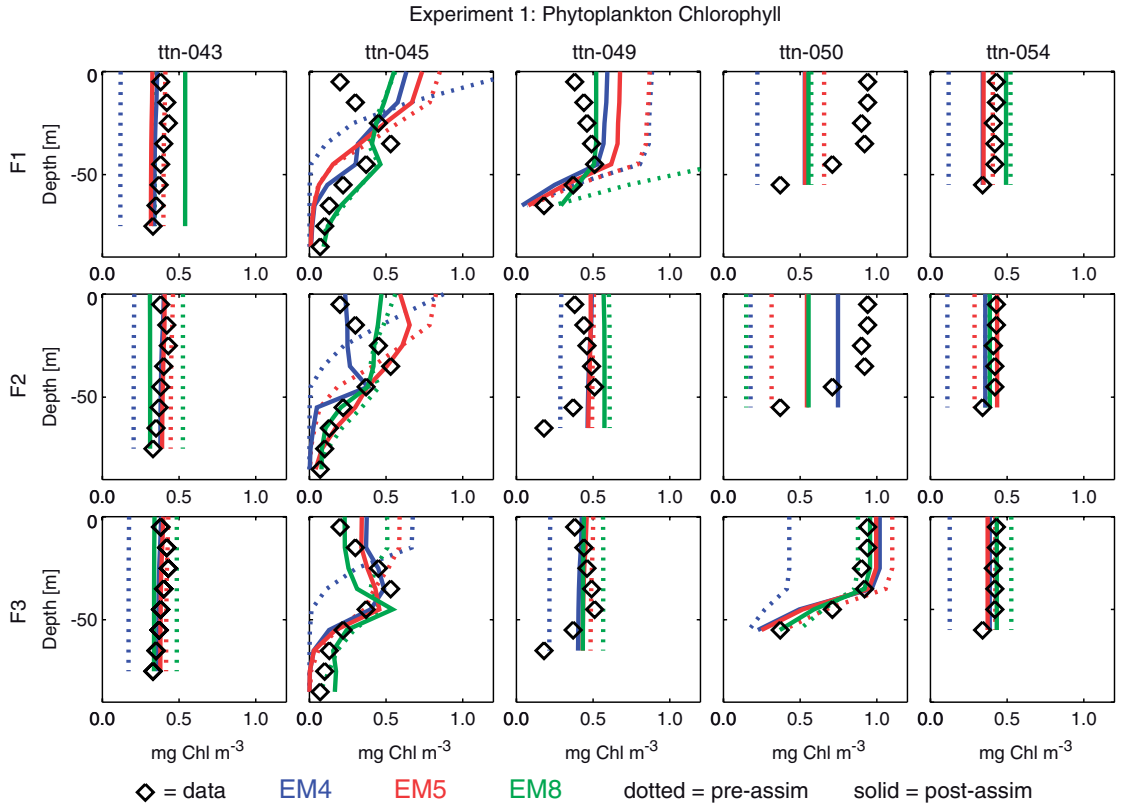


Fig. 6. Pre-assimilation (dotted lines) and post-assimilation Experiment 1 (solid lines) profiles of phytoplankton chlorophyll overlaid on the data (diamonds) from the Arabian Sea Process Study cruises.

nitrate are indistinguishable from those shown in Fig. 8. Primary productivity profiles are similar to those of Fig. 9, but have somewhat reduced surface rates. EM4, however, still produces the highest rates, most closely aligned with the observations.

In contrast to the other model-data comparisons, export flux comparisons for Experiment 2 show some improvement over those obtained for Experiment 1 (Fig. 10). Specifically, the negative fluxes obtained for the EM5 F3 combination are rectified here, and although the unrealistic oscillations produced by the EM4 F2 combination still exist, their magnitude has been reduced considerably.

### 3.3.3. *Experiment 2: cost functions*

As expected, the assimilation significantly reduces the cost function ($J<J_0$ by 24–57%) for each of the model/forcing combinations, which is only a slightly smaller reduction in the magnitude of $J$ than that obtained in Experiment 1. Even more importantly, the values of $J_P$ obtained from Experiment 2 are always less than those obtained from Experiment 1. In this experiment, on average $J_P<J_0$ by 21% as compared with the 1% average increase ($J_P>J_0$) for Experiment 1. Thus, although Experiment 2 produces simulations that do not fit the assimilated data quite as well as in Experiment 1, these simulations reproduce unassimilated data much better than the analogous Experiment 1 simulations. For example, when data from the first, second and fourth quarters of 1995 are assimilated using F3 and the resulting parameters are used to estimate the model-data misfit for the withheld third quarter of 1995 (ttn-050), Experiment 1 produces unrealistically high chlorophyll concentrations for EM4, and unrealistically low concentrations for EM8 (Fig. 13; dashed lines). Similarly, EM8 results in extremely high concentrations of zooplankton ($>3$ mmol C $m^{-3}$) during ttn-045 when data from the other cruises are assimilated. On the contrary, Experiment 2 always produces realistic concentrations of all the state variables (Fig. 13; solid lines).

As was the case for Experiment 1, F3 again produces values of both $J$ and $J_P$ that are

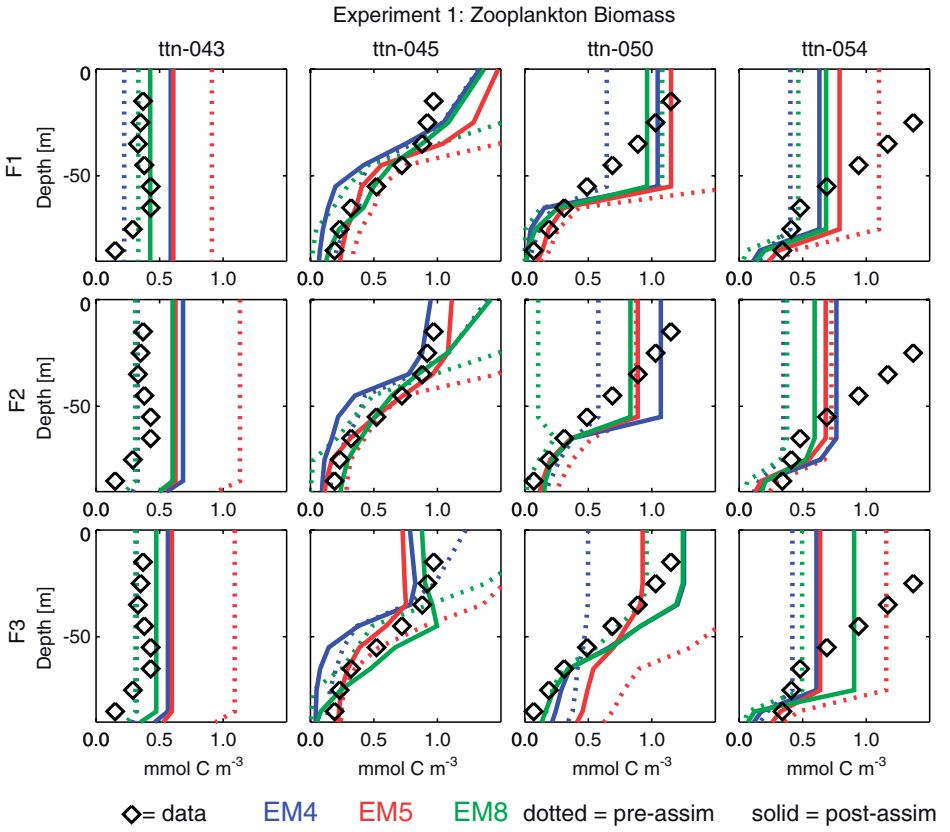


Fig. 7. As in Fig. 6, but for zooplankton biomass.

significantly lower than those obtained with F1 and F2. The predictive cost functions for the F3 EM4 and F3 EM8 combinations are the lowest obtained, and are not significantly different from each other.

### 3.3.4. *Experiment 2: optimized parameter values*

In Experiment 1 where all possible parameters are optimized, the resulting uncertainties in the parameter values computed from the Hessian matrix are high (often greater than the magnitude of the parameters themselves) because of correlations between many of the ecosystem model parameters. In Experiment 2 where only an objectively chosen subset of parameters is optimized for each of the nine modeling realizations, all optimized parameters are well constrained and are associated with relatively small uncertainties (typically less than 10% of the magnitude of the parameters).

The specific parameters that are included in the optimized parameter subset varies as a function of the ecosystem model and forcing fields; however, certain parameters are always well constrained and others are never well constrained. For example, the remineralization and phytoplankton growth rates are well constrained in each of the nine modeling realizations because the cost functions, specifically the chlorophyll, production and export components, are very sensitive to these parameters. The cost functions are also sensitive to the half-saturation constants for nutrient uptake ($k_N$), but the sensitivity is lower than it is for phytoplankton growth rate, i.e. a certain percent change in growth rate will cause a greater change in the cost function than an equivalent percent change in $k_N$. Because phytoplankton growth rate and $k_N$ typically appear in the same term in the time rate of change of phytoplankton equation, they are highly correlated and cannot both be constrained; thus $k_N$ is not included in any of the optimized parameter subsets.

Assimilation efficiency is another parameter that typically is very well constrained. The cost functions are typically somewhat sensitive to the maximum grazing rate, but this sensitivity is lower than it is for assimilation efficiency. Since both parameters

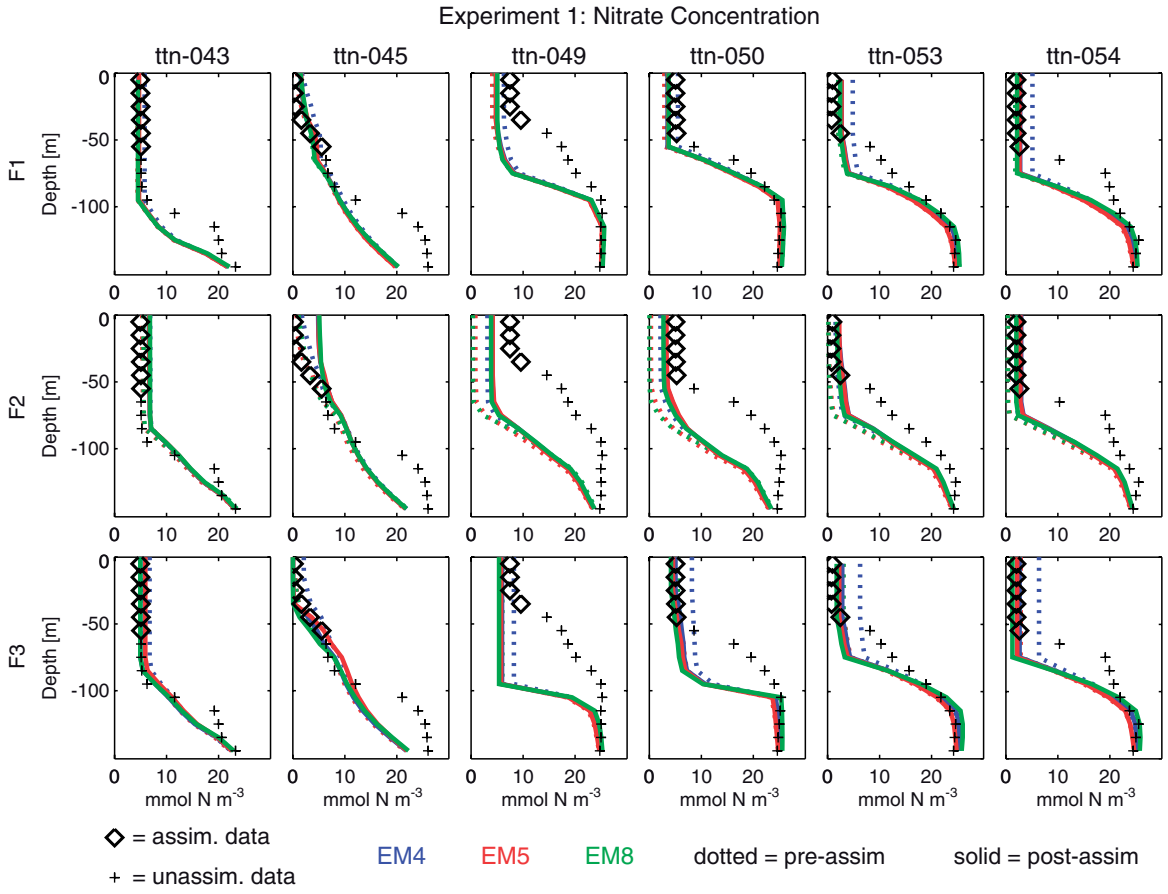


Fig. 8. As in Fig. 6, but for nitrate concentration.

appear in the same terms in each model and are highly correlated, only assimilation efficiency and not maximum grazing rate can be constrained.

Very little is known quantitatively about the magnitude of phytoplankton and zooplankton mortality rates. However, because these parameters are typically correlated with other optimized parameters such as growth rate and assimilation efficiency, they rarely appear in the optimized parameter subsets.

## 4. Discussion

### 4.1. Ecosystem model structure

Objectively and quantitatively comparing ecosystem model performance is not a straightforward proposition. If models are compared without first optimizing each one to the same data set, then it is largely the degree of tuning that is being compared, not the specific ecosystem structural characteristics. Without data assimilation, comparison results are also likely to be a strong function of the particular physical forcing fields that are selected. For example, prior to assimilation our five-component ecosystem model produces the lowest model-data misfit when the data-derived MLD and vertical velocities are used. Alternatively, when forcing time series are derived from the McCreary et al. (2001) or the Murtugudde et al. (1996) model output, the simpler and the more complex models, respectively, yield the best model-data fits. This is primarily because, prior to this study, the simplest model had already been tuned for use with the McCreary model (McCreary et al., 2001), whereas the most complex model was initially tuned for the Murtugudde model (Christian et al., 2002).

When the identical data are assimilated into these models and all parameters in each model are optimized (Experiment 1), the three ecosystem

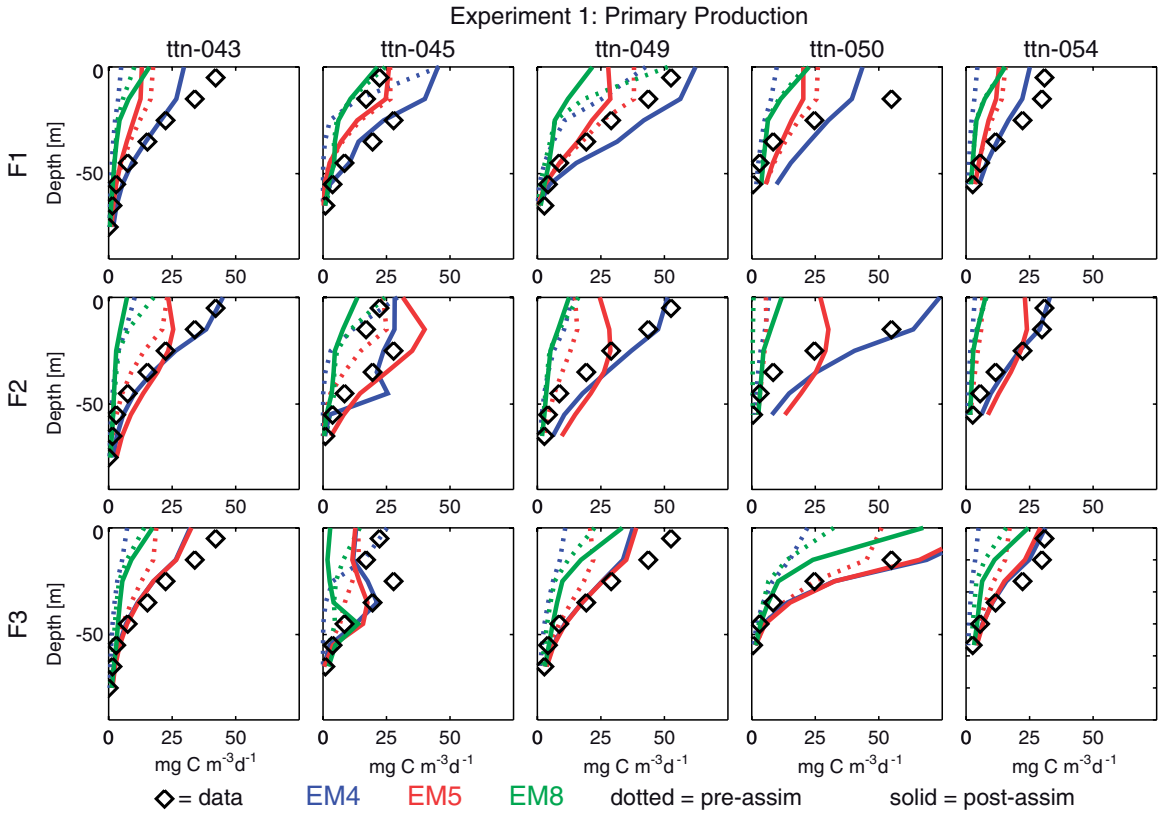


Fig. 9. As in Fig. 6, but for primary production.

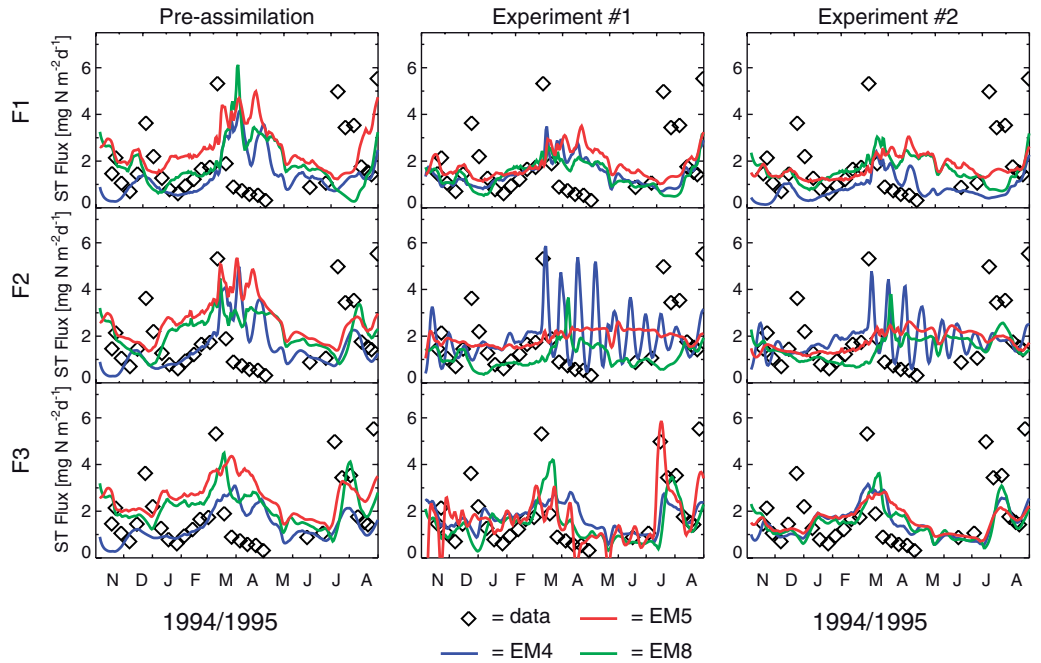


Fig. 10. Time series of particle flux export (black diamonds) from the MS-4 800 m sediment trap (Honjo et al., 1999), and model equivalents of these data obtained using EM4 (blue), EM5 (red) and EM8 (green) together with the three sets of forcing fields. Results shown are for the pre-assimilation simulation, as well as the Experiments 1 and 2 post-assimilation simulations.

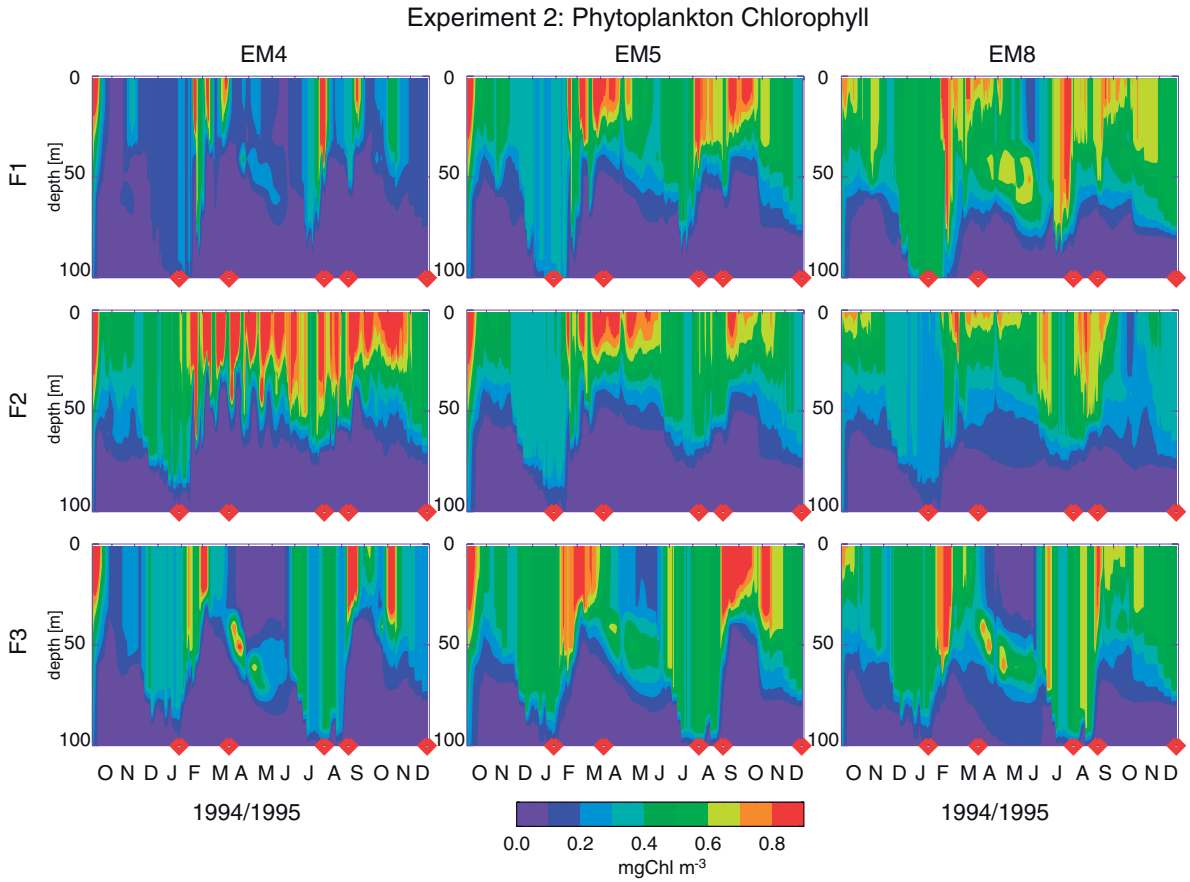


Fig. 11. As in Fig. 2, but for the post-assimilation (Experiment 2) distributions of phytoplankton chlorophyll.

models perform much more similarly (Figs. 6–9). Surprisingly, despite the fact that the most complex ecosystem model (EM8) has more degrees of freedom (19 control variables) than the simplest model (EM4 with 10 control variables), the more complex model does not fit the data better than the simpler model (Table 1). Export time series (Fig. 10; Experiment 1), also demonstrate that neither the EM4 nor the EM5 models produce reasonable sediment fluxes: an unrealistic oscillating export flux results in one case, and negative fluxes in another.

Although the cost function is an objective and quantitative measure of how well each model reproduces the assimilated data, a more general and challenging test of model performance involves examining how well the models reproduce a second unassimilated data set. It is possible for models to be tuned to a given data set and be able to reproduce that data set extremely well, yet not be able to match data from any other time or location, i.e. have almost no predictive ability. For example, suppose six bimonthly measurements of primary production are available for a given region. A fifth-order polynomial relating primary production to a given environmental variable could be generated to fit these data precisely, but such a model would be unlikely to have any predictive ability. In this case, the model would be fitting noise in the observational data rather than the underlying functional relationship. Since only the underlying functional relationship would be common to both the original data set and a second data set used to test predictive power, a second-order polynomial fit to the first data set would not provide as low a model-data misfit, but it would very possibly yield better estimates of future production than those derived from a high-order polynomial. The noise in the two data sets would be independent. Thus tuning the higher-order polynomial such that it gave a good fit to the noise in the first data set would not ensure a good fit to the second data set. Analogously, a complex ecosystem model with many unconstrained parameters can

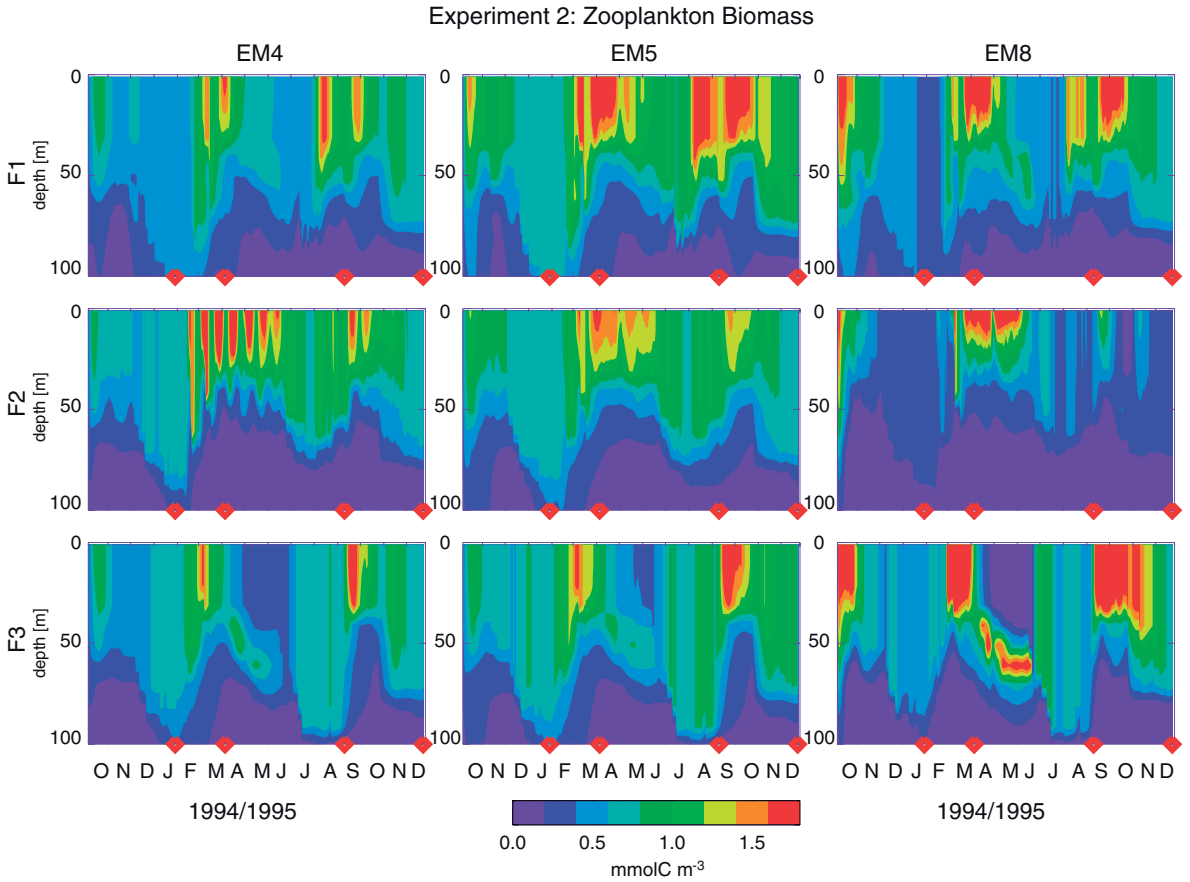


Fig. 12. As in Fig. 2, but for the post-assimilation (Experiment 2) distributions of zooplankton biomass.

often be tuned to fit a single data set very well, but simultaneously be unable to reproduce unassimilated data, and therefore be characterized by very low predictive ability.

In an attempt to assess the predictive abilities of our models, we define the predictive cost function, $J_P$, and compute this based on a series of cross-validation experiments. When all parameters are optimized, the value of $J_P$ can be a strong function of ecosystem model complexity: the most complex model yields predictive cost functions that are 57% (for F2) and 70% (for F3) higher than those of the simplest model (Table 1). In these cases, the most complex ecosystem model has many more control variables than the data are able to constrain, and many of these are highly correlated. In assimilation and parameter optimization analyses, correlated control variables by definition can never be constrained, since multiple combinations of values for these parameters will result in identical model simulations. As a result, very high uncertainty estimates are associated with these control variables. These uncertainties lead to the high $J_P$ values associated with the most complex model. Thus when all model parameters are optimized, the three ecosystem models fit the assimilated data equally well; however, the more complex model cannot reproduce unassimilated data as well as the simpler models. On the contrary, when the method outlined in Section 2.5.3 is used to objectively identify a subset of uncorrelated control variables (Experiment 2), the predictive cost functions for the most complex model are significantly lower, and are very similar to those of the other two models. In other words, the models now exhibit an equal ability to reproduce unassimilated data.

These results highlight the fact that care must be taken when selecting the control variables to be included in an assimilation analysis. Simultaneously optimizing too many parameters in a complex model, especially correlated parameters, can lead to high uncertainty estimates associated with the

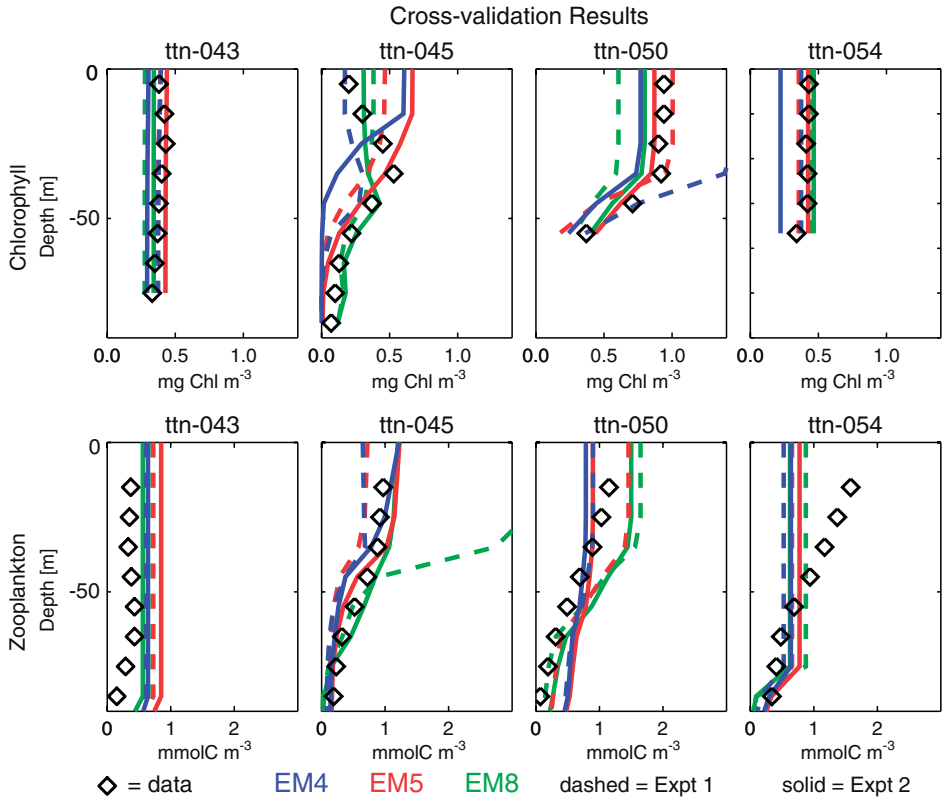


Fig. 13. Phytoplankton and zooplankton profiles obtained from cross-validation experiments overlaid on Arabian Sea Process Study cruise data (diamonds). Each set of profiles was obtained using parameters resulting from the assimilation of the remaining data, e.g., parameters used to generate the ttn-043 simulations were derived from the assimilation of ttn-045, 050 and 054 data. Results from both Experiment 1 (dashed) and Experiment 2 (solid) for all the ecosystem models are shown: EM4 (blue), EM5 (red), and EM8 (green).

resulting parameter estimates, high predictive cost functions and poor predictive ability. Counter to expectations, optimizing only a critical subset of model parameters reveals that models of higher complexity, i.e. those with multiple plankton size classes or other state variables such as DON, do not yield significantly better estimates of bulk properties such as primary production, export, and chlorophyll concentrations. The simplest and most complex models reproduce assimilated and unassimilated data equally well.

### 4.2. *Physical forcing fields*

Changes in the physical forcing fields applied to the ecosystem models produce far greater changes in plankton distributions than do changes in ecosystem model complexity. In both assimilation experiments average model-data misfits ($J_P$) obtained using F3 are significantly lower than those obtained with either F1 or F2 (Table 2). This highlights the fact that biogeochemical variability is, to a large extent, determined by physical forcing. If a key physical process is missing or incorrectly modeled, no change in ecosystem structure will be able to overcome this physical model deficiency. While other analyses (e.g., Hood et al., 2003) have reached similar conclusions concerning the importance of physical processes, this is the first study to have used data assimilation techniques to rigorously assess the relative importance of accurately modeling key physical processes versus including additional levels of ecosystem complexity.

The importance of correctly modeling physical processes in data assimilative marine ecosystem modeling studies cannot be overstated. If the time scales or space scales of the physics in which an ecosystem model is embedded are not consistent with the ecosystem itself, unrealistic parameter values can be obtained from the assimilation analysis as it attempts to compensate for the inadequate physical forcing fields. In this study,

Table 2
Means and standard deviations of predictive cost functions for the three ecosystem models using the same physical forcing (first three rows) and for the three physical forcing fields using the same ecosystem models (second three rows)

| | Experiment #1 | Experiment #2 |
|---|---|---|
| F1 | 20.6±1.0 | 18.7±1.1 |
| F2 | 27.6±6.1 | 19.9±1.4 |
| F3 | 13.9±4.0 | 10.8±1.4 |
| EM4 | 17.8±6.0 | 15.7±5.3 |
| EM5 | 20.6±8.6 | 17.8±4.7 |
| EM8 | 23.7±7.7 | 15.9±4.9 |

For Experiment 1, all control variables within each model are estimated via the variational adjoint method. For Experiment 2, an objectively chosen subset of uncorrelated model parameters is estimated.

the broad spatial scales of the McCreary et al. (2001) forcing fields (only four layers in the vertical) are not consistent with the sharp vertical gradients characteristic of many biogeochemical observations (e.g., plankton and nutrient concentrations). This inconsistency may be at least partially responsible for the unrealistically high grazing rates that cause the spurious phytoplankton/zooplankton oscillations apparent in Figs. 4, 5 and 10. Interestingly, although the very high temporal resolution of these forcing time series (resolution of diurnal changes in MLD) was expected to be advantageous for F2, additional experiments using daily averaged mixed layers demonstrate that identical results are obtained whether or not the MLDs resolve diurnal variability. This is not to say, however, that including diurnal variability is not important. Indeed, recent analyses demonstrate that the inclusion of a diurnal cycle has very important influences on MLD variability as well as the intraseasonal and interannual supply of nutrients (McCreary et al., 2001; Wiggert et al., 2002).

Although one might logically assume that the mooring data-derived MLDs and vertical velocities (F1) would be most consistent with biogeochemical cruise observations and thus produce the lowest model-data misfits, this is not the case. In fact, the distance between the mooring and the S7 station location (~80 km) is large enough to cause significant differences between the MLDs obtained from the mooring and those obtained from the cruise CTD casts. For example, compare the 63–65 m MLD estimates derived from the mooring data (15.5°N, 62.5°W), with the 50 m MLD observed at S7 (16.0°N, 62.0°W; YD248) and the 100 m MLD observed at station S9 (15.25°N, 63.5°W; YD249).

A comparison of TOPEX-Poseidon images for this time and location reveals that mesoscale features are moving roughly 8–10 km per day (Dickey et al., 1998) and are thus consistent with a 10-day time difference between an event occurring at the S7 site and at the mooring location. This time lag may be largely responsible for the poorer performance of the models forced by the data-derived MLD time series (F1), as compared with the models forced by the Murtugudde model (F3).

While the success of F3 results from a number of factors, the excellent model-data fit for the ttn-050 chlorophyll values (Fig. 6) plays the largest single role. Models forced by F1 or F2 are unable to capture these high (O (1.0 mg chl $m^{-3}$)) chlorophyll values, whereas models forced by F3 reproduce these high concentrations even when these data are omitted from the assimilation (Fig. 13; Experiment 2).

In the models these high chlorophyll concentrations are produced by the rapid shoaling of the F3 mixed layers from 105 m on YD239 to 30 m on YD249 (Fig. 1B). Prior to YD239, the very deep mixed layer mixes high nutrient waters all the way to the surface, but it also decreases the average light supply by mixing phytoplankton below the critical depth. As the mixed layer shoals, nutrient concentrations remain high at the surface, while daily integrated irradiance experienced within the mixed layer increases substantially. The result is a relatively short-lived "detrainment bloom," i.e. the classic spring bloom mechanism of Sverdrup (McCreary et al., 1996), which lasts until the surface nitrate is consumed.

Despite much smaller changes in MLD, a detrainment bloom of similar magnitude is also produced by the F2-forced models. Both the 57-m shoaling (F2) and the 75-m shoaling (F3) produce chlorophyll concentrations of roughly 1.0 mg chl $m^{-3}$. However, because the shoaling begins 10 days earlier for F3 (YD239) than it does for F1 and F2 (YD249) (Fig. 1B), surface chlorophyll concentrations obtained with F3 are high (O (1 mg chl $m^{-3}$)) by the time the data were collected on (YD248), whereas surface chlorophyll concentrations obtained with the other forcing fields id not reach O (1 mg chl $m^{-3}$) until after YD254 (Fig. 14).

To investigate the idea that the F3-forced models produce lower cost functions because of the timing of this entrainment bloom, Experiment 2 is performed with two additional physical forcing fields (F4 and F5), which, except for the following modifications, are both identical to those of F3. For F4, during the time period YD220-242 the depth of the mixed layer is raised from below 100 to 65 m, so that it is roughly equal to the depth of the F2 mixed layer. For F5, an alternate mixed-layer time series is used from YD220-255 (Fig. 1B). This alternate time series is derived from the same Murtugudde et al. (1996) model, but in this case output from a higher spatial resolution simulation is used. Essentially, F4 tests the models' sensitivity to a change in the magnitude of mixed-layer shoaling associated with the termination of the SWM, whereas F5 tests the models' sensitivity to a change in the timing of this shoaling.

Results from this simple experiment demonstrate that changing the depth of the mixed layer during the SWM produces only a small change in model-data misfit. When the mixed-layer depth is raised from >100 to 65 m, the average value of $J_P$ for the three ecosystem models (F4: $J_P = 12.5 \pm 1.3$) is not significantly changed (F3: $J_P = 10.8 \pm 1.4$; Table 2). On the contrary, when the timing of the mixed-layer shoaling is altered, the average $J_P$ value (F5: $J_P = 17.2 \pm 1.4$) is outside the standard deviation of the F3 values, and is in much better agreement with the other values (F1: $J_P = 18.7 \pm 1.1$; F2: $J_P = 19.9 \pm 1.4$; Table 2). These results suggest that the significant differences obtained using the different physical forcing fields are largely attributable to differences in the timing of the mixed-layer shoaling and detrainment bloom associated with the termination of the SWM.

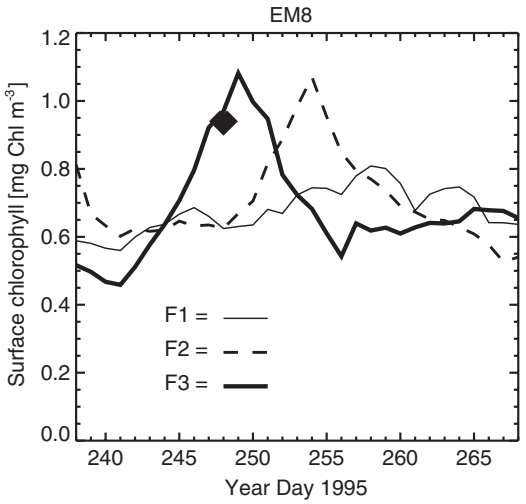


Fig. 14. Simulated (EM8) surface chlorophyll concentrations obtained using Forcing 1 (thin solid line), Forcing 2 (dashed line), and Forcing 3 (thick solid line.) Black diamond represents data from the S7 Arabian Sea Process Study cruise ttn-050.

This example also demonstrates the critical importance of having physical models that not only resolve the correct dynamics, but also reproduce the correct timing of events. This is particularly crucial when data assimilation and parameter estimation techniques are applied. Simulated plankton distributions obtained from an ecosystem model forced by a mixed-layer time series that is uniformly too shallow can be made to fit the available data by adjusting parameters such as uptake and growth rates; however, an ecosystem model forced by a mixed-layer time series that shoals at the wrong time will never produce plankton distributions in agreement with the data, no matter how many ecosystem parameters are adjusted. Developing physical models that reproduce the correct processes with the correct timing is a difficult task and is one that can probably best be accomplished by the assimilation of physical data.

Although the F3 forcing successfully reproduces the high chlorophyll observations from ttn-050, it is not completely clear whether the processes behind these high chlorophyll concentrations are correctly simulated. As discussed above, the models all react quickly to the sharp shoaling of the mixed layer and detrainment blooms are produced. However, several lines of evidence suggest that the high chlorophyll concentrations at S7 during ttn-050 were associated with a filament extending from the coastal upwelling zone just north of the southern transect, that advected coastal diatom communities well offshore (Flagg and Kim, 1998; Barber et al., 2001; Manghnani et al., 1998; Latasa and Bidigare, 1998; Garrison et al., 1998). Thus, these high chlorophyll concentrations at S7 may not result from in situ growth due to a detrainment bloom, but rather from the offshore advection of a coastal bloom. Furthermore, this mesoscale activity does not appear to impact the mooring location, as mixed-layer time series from a 1-D mixed-layer model forced with heat and momentum fluxes measured by the mooring's surface buoy show excellent correspondence with the F1 mixed layer throughout the SWM (data not shown). Therefore, while the models forced by F3 manage to reproduce

the high ttn-050 chlorophyll observations, they may be doing so for the wrong reasons. This demonstrates that application of data assimilation to achieve the best-fit model solution can be misleading if the additional context derived from all available ancillary data is not taken into account.

## 5. Conclusions

As knowledge regarding the complex components of marine ecosystems continues to grow, the models developed to examine these systems are correspondingly becoming more complex as they include increasing numbers of organisms and biological processes. Unfortunately, our knowledge of ecosystem complexities is growing faster than the available validation data sets, leading to the development of models with a greater number of unconstrained parameters. As the number of degrees of freedom for these models continues to grow, these models may become better able to reproduce specific data sets. However, as illustrated by this analysis, the fact that a model is better able to reproduce a given data set does not necessarily imply that this model will be better able to reproduce a second data set, taken from a slightly different time or location. Nor does it imply that it will have greater predictive ability.

Data assimilation techniques are becoming increasingly popular, as ocean-color time series such as Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and Moderate Resolution Imaging Spectroradiometer (MODIS) and in situ time series (e.g., the Bermuda Atlantic Time-Series Study and Hawaii Ocean Time-Series) continue to grow in length, availability and comprehensiveness. Applying formal data assimilation techniques, such as the variational adjoint method, is a useful way to avoid difficulties associated with subjectively tuning the many parameters typically associated with marine ecosystem models. By design, data assimilation reduces model-data misfit and typically results in a best-fit model solution for a given data set. As demonstrated in this paper, these techniques can also be used as a tool for quantitatively and objectively comparing the performance of multiple ecosystem models of varying complexity.

In all of these data assimilation applications, care must be taken when selecting the specific parameters to be optimized. Simultaneously optimizing too many unconstrained parameters, especially partially correlated ones, can lead to very high parameter uncertainties and unrealistic parameter values. In this study, the estimation of too many unconstrained parameters results in excessively high grazing rates that are the likely cause of the spurious phytoplankton/zooplankton oscillations that appeared most prominently in the simplest model. Although this may not have a deleterious effect on model-data fit for the assimilated data set, it can significantly impair a model's ability to reproduce, or 'predict' unassimilated data. Fortunately such oscillations can likely be avoided by assimilating time series data, e.g., ocean color or bio-optical data, in addition to the in situ profile snapshots assimilated here.

Here we present a method for selecting a subset of model parameters for optimization, which tends to remove correlated parameter pairs and maximize the predictive ability of a model. In short, by using the inverse of the Hessian matrix to determine the sensitivity of the cost function to each model parameter, and by optimizing only those parameters to which the cost function is most sensitive, greatly improved results were obtained for a set of cross-validation experiments and predictive ability was enhanced.

To our knowledge this is the first study to use data assimilation to rigorously demonstrate that the inclusion of additional state variables (e.g., DON or multiple plankton or detrital size classes) does not necessarily improve an ecosystem model's ability to reproduce unassimilated bulk biogeochemical distributions and rates, such as primary production, export, plankton concentrations and nutrients. Specifically we find that if a complex model is associated with too many unconstrained parameters, it fails to reproduce any data successfully except those that are assimilated. In this case the model may be describing noise in the original data set and the price associated with fitting noise is a loss of predictive ability. On the contrary, when only an objectively and systematically chosen subset of parameters is optimized, the models all demonstrate enhanced predictive ability, with the complex model exhibiting the greatest improvement. By choosing to optimize only uncorrelated parameters to which the cost function is most sensitive, tuning models to describe noise can be avoided or at least minimized. Hopefully, as the collection of additional observations beyond the standard suite applied herein becomes commonplace, it will be possible to expand the subset of uncorrelated parameters such that the potential of more complex ecosystem models can be better realized.

This analysis also demonstrates that small changes in physical forcing fields can produce greater changes in plankton distributions than substantial changes in ecosystem model complexity. The fact that biological distributions are very sensitive to the physical environment is not a new result; however, this study is uniquely able to quantify rigorously the importance of the physical processes, and compare this to the importance of ecosystem model complexity. Here we find that it is crucial to correctly capture the timing of major events, such as the shoaling of the mixed layer at the termination of the SWM. In an assimilative analysis such as that described here, a mixed layer that is slightly too deep can be compensated for by slight changes in growth and grazing parameters; however, no reasonable change in ecosystem parameters can correct for a mixed layer that shoals at the wrong time. Deficiencies in physical forcing fields and inconsistencies between temporal or spatial scales resolved by the physical and biological models also can cause the data assimilation process to yield inappropriate model parameters, resulting in unrealistic ecosystem behavior and low model predictive capability. The simultaneous assimilation of both physical and biological data is the most reliable means for ensuring that physical and biological models are consistent, and the timing of major physical events is correct.

An extension of the analysis presented here that is currently underway includes the simultaneous assimilation of data from three very distinct locations: the Arabian Sea, the North Atlantic and the equatorial Pacific. Through this effort we are investigating whether a complex model is more capable than a simple model of reproducing observed behavior when a constant parameter set is applied at all three locations. The data assimilation framework and methods established here will be used to address this question.

## Acknowledgements

This research was supported by the US National Science Foundation, through the JGOFS Synthesis and Modeling Project (OCE-0097285) and the National Aeronautics and Space Agency (NAG5-11259). Additionally, support for RRH was provided by NSF (OCE-9818708) and for JDW was provided by the NASA Oceanography Program (NAG-58595). This is University of Maryland Center for Environmental Science contribution number 3890. Computer facilities and support were provided by the Commonwealth Center for Coastal Physical Oceanography at Old Dominion University. We thank Jay McCreary, Kevin Kohler and Ragu Murtugudde for providing 3-D model output fields and Bob Weller for providing data from the WHOI mooring, all of which were used as forcing fields for our 1-D ecosystem models. The comments provided by Katja Fennel, Ed Laws, Jim Christian, Scott Doney, and an anonymous reviewer were very helpful. We are especially grateful for the hard work of the many JGOFS scientists who helped acquire the excellent Arabian Sea biogeochemical data set used in our analysis.

## References

Anderson, T.R., 1993. A spectrally averaged model of light penetration and photosynthesis. Limnology and Oceanography 38 (7), 1403–1419.

Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Marine Ecological Progress Series 271, 13–26.

Barber, R.T., Marra, J., Bidigare, R.C., Codispoti, L.A., Halpern, D., Johnson, Z., Latasa, M., Goericke, R., Smith, S.L., 2001. Primary productivity and its regulation in the Arabian Sea during 1995. Deep-Sea Research II 48, 1127–1172.

Bissett, W.P., Walsh, J.J., Dieterle, D.A., Carder, K.L., 1999. Carbon cycling in the upper waters of the Sargasso Sea. I: Numerical simulation of differential carbon and nitrogen fluxes. Deep-Sea Research Part I 46, 205–269.

Chen, D., Busalacchi, A., Rothstein, L., 1994. The roles of vertical mixing, solar radiation, and wind stress in a model simulation of the sea surface temperature seasonally cycle in the tropical Pacific Ocean. Journal of Geophysical Research 99, 20,345–20,359.

Christian, J.R., Verschell, M.A., Murtugudde, R., Busalacchi, A.J., McClain, C.R., 2002. Biogeochemical modelling of the tropical Pacific Ocean. I: Seasonal and interannual variability. Deep-Sea Research Part II 49, 509–543.

Coles, V.J., Hood, R.R., Pascual, M., Capone, D.G., 2004. Modeling the effects of *Trichodesmium* and nitrogen fixation in the Atlantic Ocean. Journal of Geophysical Research 109(C6), C06007, doi:10.1029/2002JC001754.

Denman, K.L., 2003. Modelling planktonic ecosystems: parameterizing complexity. Progress in Oceanography 57, 429–452.

Denman, K.L., Pena, M.A., 2002. The response of two coupled one-dimensional mixed layer/planktonic ecosystem models to climate change in the NE subarctic Pacific Ocean. Deep Sea Research II 49, 5739–5757.

Dickey, T., Marra, J., Sigurdson, D.E., Weller, R.A., Kinkade, C.S., Zedler, S.E., Wiggert, J.D., Langdon, C., 1998. Seasonal variability of bio-optical and physical properties in the Arabian Sea: October 1994–October 1995. Deep Sea Research Part II 45, 2001–2025.

Fennel, K., Losch, M., Schroter, J., Wenzel, M., 2001. Testing a marine ecosystem model: sensitivity analysis and parameter optimization. Journal of Marine Systems 28, 45–63.

Fischer, A.S., Weller, R.A., Rudnick, D.L., Eriksen, C.C., Lee, C.M., Brink, K.H., Fox, C.A., Leben, R.R., 2002. Mesoscale eddies, coastal upwelling, and the upper-ocean heat budget in the Arabian Sea. Deep-Sea Research Part II 49, 2231–2264.

Flagg, C.N., Kim, H.-S., 1998. Upper ocean currents in the northern Arabian Sea from shipboard ADCP measurements collected during the 1994–1996 U.S. JGOFS and ONR programs. Deep-Sea Research Part II 45, 1917–1960.

Friedrichs, M.A.M., 2001. A data-assimilative marine ecosystem model of the central equatorial Pacific: numerical twin experiments. Journal of Marine Research 59, 859–894.

Friedrichs, M.A.M., 2002. Assimilation of SeaWiFS and JGOFS EqPac data into a marine ecosystem model of the central equatorial Pacific. Deep-Sea Research Part II 49, 289–319.

Friedrichs, M.A.M., Hofmann, E.E., 2001. Physical control of biological processes in the central equatorial Pacific. Deep-Sea Research I 48, 1023–1069.

Garrison, D.L., Gowing, M.M., Hughes, M.P., 1998. Nano and microplankton in the northern Arabian Sea during the Southwest Monsoon, August–September 1995: a US-JGOFS study. Deep-Sea Research Part II 45, 2269–2300.

Gilbert, J.C., Lemarechal, C., 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. Mathematical Programming 45, 405–435.

Gregg, W.W., Ginoux, P., Schopf, P.S., Casey, N.W., 2003. Phytoplankton and iron: validation of a global three-dimensional ocean biogeochemical model. Deep Sea Research Part II 50, 3143–3169.

Gundersen, J.S., Gardner, W.D., Richardson, M.J., Walsh, I.D., 1998. Effects of monsoons on the seasonal and spatial distributions of POC and chlorophyll in the Arabian Sea. Deep Sea Research Part II 45, 2103–2132.

Honjo, S., Dymond, J., Prell, W., Ittekkot, V., 1999. Monsoon-controlled export fluxes to the interior of the Arabian Sea. Deep-Sea Research II 46, 1859–1902.

Hood, R.R., Bates, N.R., Capone, D.G., Olson, D.B., 2001. Modeling the effect of nitrogen fixation on carbon and nitrogen fluxes at BATS. Deep-Sea Research Part II 48, 1609–1648.

Hood, R.R., Kohler, K.E., McCreary, J.P., Smith, S.L., 2003. A four-dimensional validation of a coupled physical–biological model of the Arabian Sea. Deep-Sea Research Part II 50, 2917–2945.

Hood, R.R., Coles, V.J., Capone, D.G., 2004. Modeling the distribution of *Trichodesmium* and nitrogen fixation in the Atlantic Ocean. Journal of Geophysical Research 109, C06006, doi:10.1029/2002JC001753.

Kantha, L.H., 2004. A general ecosystem model for applications to primary productivity and carbon cycle studies in the global oceans. Ocean Modelling 6, 285–334.

Kraus, E.B., Turner, J.S., 1967. A one-dimensional model of the seasonal thermocline. II: the general theory and its consequences. Tellus 119, 98–106.

Latasa, M., Bidigare, R.R., 1998. A comparison of phytoplankton populations of the Arabian Sea during the Spring Intermonsoon and Southwest Monsoon of 1995 as described by HPLC-analyzed pigments. Deep-Sea Research Part II 45, 2133–2170.

Lawson, L.M., Spitz, Y.H., Hofmann, E.E., Long, R.B., 1995. A data assimilation technique applied to a predator–prey model. Bulletin of Mathematical Biology 57, 593–617.

Lee, C.M., Johns, B.H., Brink, K.H., Fischer, F.S., 2000. The upper-ocean response to monsoonal forcing in the Arabian Sea: seasonal and spatially variability. Deep-Sea Research II 47, 1177–1226.

Manghnani, V., Morrison, J.M., Hopkins, T.S., Bohm, E., 1998. Advection of upwelled waters in the form of plumes off Oman during the Southwest Monsoon. Deep-Sea Research Part II 45, 2027–2052.

Matear, R.J., 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. Journal of Marine Research 53, 571–607.

McCreary, J.P., Kohler, K.E., Hood, R.R., Olson, D.B., 1996. A four-component ecosystem model of biological activity in the Arabian Sea. Progress in Oceanography 37, 117–165.

McCreary Jr., J.P., Kohler, K.E., Hood, R.R., Smith, S., Kindle, J., Fischer, A.S., Weller, R.A., 2001. Influences of diurnal and intraseasonal forcing on mixed-layer and biological variability in the central Arabian Sea. Journal of Geophysical Research 106 (C4), 7139–7155.

McGillicuddy Jr., D.J., Lynch, D.R., Moore, A.M., Gentleman, W.C., Davis, C.S., Meise, C.J., 1998. An adjoint data assimilation approach to diagnosis of physical and biological controls on *Pseudocalanus* spp. in the Gulf of Maine-Georges Bank Region. Fisheries Oceanography 7, 205–218.

Moore, J.K., Doney, S.C., Glover, D.M., Fung, I.Y., 2002. Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. Deep-Sea Research Part II 49 (1–3), 463–507.

Morrison, J.M., Codispoti, L.A., Gaurin, S., Jones, B., Manghnani, V., Zheng, Z., 1998. Seasonal variation of hydrographic and nutrient fields during the US JGOFS Arabian Sea Process Study. Deep-Sea Research Part II 45, 2053–2101.

Murtugudde, R., Busalacchi, A.J., 1998. Salinity effects in tropical oceans. Journal of Geophysical Research 103, 3283–3300.

Murtugudde, R., Seager, R., Busalacchi, A., 1996. Simulation of the tropical oceans with an ocean GCM coupled to an atmospheric mixed-layer model. Journal of Climate 9, 1795–1815.

Pacanowski, R.C., Philander, S.G.H., 1981. Parameterization of vertical mixing in numerical models of tropical oceans. Journal of Physical Oceanography 11, 1443–1451.

Schartau, M., Oschlies, A., 2003. Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part I—Method and parameter estimates. Journal of Marine Research 61, 765–793.

Schartau, M., Oschlies, A., Willebrand, J., 2001. Parameter estimates of a zero-dimensional ecosystem model applying the adjoint method. Deep Sea Research Part II 48, 1769–1800.

Smith, S.L., Codispoti, L.A., Morrison, J.M., Barber, R.T., 1998. The 1994–1996 Arabian Sea Expedition: an integrated, interdisciplinary investigation of the response of the northwestern Indian Ocean to monsoonal forcing. Deep-Sea Research Part II 45, 1905–1915.

Spitz, Y.H., Moisan, Abbott, J.R., 2001. Configuring an ecosystem model using data from the Bermuda Atlantic Time Series (BATS). Deep Sea Research II 48, 1733–1768.

Tziperman, E., Thacker, W.C., 1989. An optimal-control/adjoint-equations approach to studying the oceanic

general circulation. Journal of Physical Oceanography 19, 1471–1485.

Vallino, J.J., 2000. Improving marine ecosystem models: use of data assimilation and mesocosm experiments. Journal of Marine Research 58, 117–164.

Weller, R.A., Baumgartner, M.F., Josey, S.A., Fischer, A.S., Kindle, J.C., 1998. Atmospheric forcing in the Arabian Sea during 1994–1995: observation and comparisons with climatology and models. Deep Sea Research Part II 45, 1961–1999.

Wiggert, J.D., Jones, B.H., Dickey, T.D., Brink, K.H., Weller, R.A., Marra, J., Codispoti, L.A., 2000. The Northeast Monsoon's impact on mixing, phytoplankton biomass and nutrient cycling in the Arabian Sea. Deep-Sea Research Part II 47, 1353–1385.

Wiggert, J.D., Murtugudde, R.G., McClain, C.R., 2002. Processes controlling interannual variations in wintertime (Northeast Monsoon) primary productivity in the central Arabian Sea. Deep-Sea Research II 49, 2319–2343.

Wiggert, J.D., Murtugudde, R.G., Christian, J.R., 2006. Annual ecosystem variability in the tropical Indian Ocean: results of a coupled bio-physical ocean general circulation model. Deep-Sea Research Part II, this issue [doi:10.1016/j.dsr2.2006.01.027].

Wishner, K.F., Gowing, M.M., Gelfman, C., 1998. Mesozooplankton biomass in the upper 1000 m in the Arabian Sea: overall seasonal and geographic patterns, and relationship to oxygen gradients. Deep-Sea Research II 45, 2405–2432.