Physical and biological modeling in the Gulf Stream region:
I. Data assimilation methodology

Laurence A. Anderson*, Allan R. Robinson, Carlos J. Lozano

Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

Received 4 October 1999; received in revised form 1 February 2000; accepted 2 February 2000

Abstract

Physical and biological data are assimilated into a time-evolving, mesoscale-resolution three-dimensional (3-D) ocean model using optimal interpolation. Simulations are conducted in the Gulf Stream region during the BIOSYNOP/Anatomy of a Meander Experiment in September–October of 1988. Physical data assimilation only or biological data assimilation only resulted in misalignment of the physical and biological fronts, causing spurious cross-frontal fluxes of biological quantities. Assimilation of both physical and compatible biological fields was necessary for adequate equilibration of the simulated fields. The resulting combined 4-D fields substantially extend the value of the observations alone. A technique is presented for deriving the necessary, dynamically consistent 3-D physical and biological field estimates from data for initialization and assimilation into time-evolving model simulations. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Data assimilation; Gulf Stream; Modeling

1. Introduction

Data assimilation in physical and biological ocean models is desirable for keeping model simulations consistent with observations. Data-assimilative simulations interpolate between the data and extrapolate from them dynamically, improving field estimation and allowing parameter estimation through model-data comparison. Data-assimilative simulations yield dynamically balanced fields that agree with the

---

*Corresponding author. Present address: Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA; Fax: + 1-508-457-2132.

E-mail address: landerson@whoi.edu (L.A. Anderson).

data, but that can be analyzed in ways that the raw data cannot (due to poor data resolution), e.g. derivatives taken and balance of terms computed. Multivariate data assimilation allows the synthesis of various types of available data with each other and with physical and biological principles (i.e. the model equations), to arrive at a consistent, best estimate of ocean fields, parameters and processes.

The objective of this paper is to present a multivariate data assimilation methodology that is able to make a complete and accurate reconstruction of four-dimensional (4-D) mesoscale physical and biological distributions and processes. This is done in two steps: (a) the estimation of dynamically consistent 3-D physical and biological fields from irregularly spaced 4-D data (i.e. interdisciplinary nowcasts), and (b) the assimilation of these multivariate data fields into a 4-D, mesoscale-resolution forward dynamical ocean model in a smooth way (i.e. with minimal artificial shock). An assimilation technique with low computational expense is chosen so that this method can be applied to interdisciplinary ocean forecasting and for analyzing interdisciplinary data sets in four dimensions with adequately complex models.

Physical ocean data assimilation into dynamical models is now relatively well developed (Haidvogel and Robinson, 1989; Anderson and Willebrand, 1989; Brasseur and Nihoul, 1994; Brasseur, 1995; Malanotte-Rizzoli, 1996). However, biological data assimilation into three-dimensional (3-D) ocean models is still in its infancy. Biological data assimilation is particularly challenging due to the complexity of the ocean ecosystem (i.e. model errors), its dependency on physical processes, and its significant (non-Gaussian) spatial and temporal variability.

Data assimilation techniques can be divided into two categories: "forward" techniques (i.e. estimation theory methods, e.g. direct insertion, nudging, optimal interpolation, Kalman filtering), which directly incorporate the data into a prognostic model simulation, and "inverse" techniques (i.e. control theory methods, e.g. generalized inverse, adjoint, conjugate gradient, simulated annealing, Monte Carlo) which seek the best parameter values, initial conditions or boundary conditions that lead to optimal agreement between the model and data. Combination of the two approaches is of course possible. For a review of ocean data assimilation techniques see Ghil and Malanotte-Rizzoli (1991), Evensen (1994), and Robinson et al. (1998).

Advantages of inverse techniques are that they optimally fit the model to the data for all space and time, and obey mass conservation. Previous work using inverse techniques to assimilate biological data have largely been limited to estimating zero-dimensional steady-state balances (e.g. Niquil et al., 1998; Jackson and Eldridge, 1992; Vézina and Pace, 1994) or fitting time-series data with zero spatial dimensions (Carpenter et al., 1994; Fasham and Evans, 1995; Matear, 1995; Lawson et al., 1996; Hurtt and Armstrong, 1996, 1999; Prunet et al., 1996a, b; Harmon and Challenor, 1997; Spitz et al., 1998; Evans, 1999). A few studies have used inverse methods with three-dimensional ocean models to assimilate biological data (Nihoul et al., 1994; Matear and Holloway, 1995; Bowen et al., 1995; Semovski and Woźniak, 1995; McGillicuddy et al., 1998) but rely on simplified dynamics or assume steady-state balance. At present, the use of inverse techniques with fully complex, 4-D mesoscale-resolution physical–biological models is primarily limited by computational expense, and has not yet been successfully demonstrated.

In this study we use the forward data assimilation method of optimal interpolation, commonly used in meteorological forecasting (Bengtsson et al., 1981; Phillips, 1982; Ghil, 1989; Lorenc et al., 1991) and physical ocean forecasting (Robinson and Leslie, 1985; Mellor and Ezer, 1991; Fox et al., 1992; Robinson, 1992; Lozano et al., 1996). Previous studies using forward techniques to assimilate biological data in three dimensions have used the simpler methods of direct insertion (Ishizaka, 1990) and nudging (Najjar et al., 1992; Anderson and Sarmiento, 1995; Armstrong et al., 1995; Moisan et al., 1996). Semovski and Woźniak (1995) used a simplified form of optimal interpolation to assimilate CZCS data into a coarse-resolution model of the North Atlantic, but there was no circulation model per se, i.e. no horizontal advection or diffusion, only vertical mixing. To our knowledge, this is the first study to assimilate both physical and biological data into a prognostic, eddy-resolving 3-D coupled model using optimal interpolation to reconstruct an observed mesoscale event.

From Ishizaka (1993), Johannessen et al. (1993) and the studies listed above, the current problems facing forward biological data assimilation can be summarized:

(1) Variability in the data or non-smooth assimilation methods (e.g. direct insertion) can generate discontinuities (spikes) and subsequently waves (assimilation shock) in the physical and biological fields, producing spurious behavior.
(2) Assimilating only one biological variable can cause unrealistic values in non-assimilated variables. This indicates values for all biological variables should be assimilated. Assimilating all biological variables is also important for keeping total nitrogen (or carbon, etc.) approximately conserved in the simulation. But how does one determine values to be assimilated for variables for which there is no data?
(3) Biological variables may quickly change from their assimilated values, for instance by returning to the pre-assimilation model values. While this is often an indication of inadequacy of the biological model (i.e. that the model state is being pulled toward a solution that does not agree with the data), it may also be due to boundary conditions, surface forcing or physical structures that are inconsistent with the biological data.
(4) A way is needed to interpolate and extrapolate data spatially. For example, ocean color data need to be extrapolated vertically. The assimilation of sparse physical data points that are not adequately interpolated can generate spurious eddies.
(5) It is difficult to assimilate multivariate data in a smooth way (i.e. with minimal artificial shock), because different data types often contain apparent discrepancies.

In addition to addressing these points, we also wished to investigate whether the previously used techniques of assimilating only biological data into a physical–biological model (e.g. Armstrong et al., 1995) or assimilating only physical data (to drive the biology) are reasonable for obtaining mesoscale biological distributions that are in good agreement with the data. Solutions to all the above issues have not been offered in previous studies. The originality of our result is that here we have developed a method that addresses each of these successfully. One of our main conclusions is therefore the methodology presented in Section 2. The general method is not limited to the study described here, but can be applied to analyzing many interdisciplinary

oceanographic data sets. As the model is prognostic, it need only be run forward to provide a forecast after assimilation of the most recent data.

This study focuses on the Gulf Stream frontal region for four reasons. The first is to study the physical and biological processes at the Gulf Stream front, the topic of our companion paper (Anderson and Robinson, 2000). The second reason is because of the relatively large amount of physical and biological data available in this region for model calibration, assimilation and validation (e.g. Halkin and Rossby, 1985; Bower and Rossby, 1989; Lohrenz et al., 1993; Hitchcock et al., 1993; Lindstrom and Watts, 1994). The third is because of our previous experience in physical modeling in the Gulf Stream region (e.g. Robinson et al., 1988; Spall and Robinson, 1990; Robinson and Gangopadhyay, 1997; Gangopadhyay and Robinson, 1997). The fourth is as a basis for future work in realtime physical–biological forecasting in the northwest Atlantic.

In this study we synthesize information from two main data sets. The first is the combined fall BIOSYNOP/Anatomy of a Meander data set, from cruises made in the Gulf Stream region in Sept–Oct 1988. The second is the GULFCAST data set for this same time period, which is weekly information on Gulf Stream axis and ring positions. These data sets are described in Section 2.3. In this study we use the data for model initialization, assimilation and verification.

We concentrate on a specific 17-day time period, 21 Sept–6 Oct 1988, which is Leg 1 of the Fall BIOSYNOP experiment. The challenge is to construct an accurate realization of the Gulf Stream frontal position and physical–biological fields for that specific period.

2. Method

2.1. Overview of modeling approach

The simulations in this study use a time-evolving 3-D physical model coupled to a biological model, which are described in Section 2.2 and the appendix. Physical and biological fields are needed for model initialization and assimilation. Because the state and position of the Gulf Stream front and rings at a specific time are desired, a long model spin-up period cannot be used to generate the initialization fields (as is often done for idealized or annual cycle simulations). Rather the initialization and assimilation fields must be generated prior to the main simulation as our best estimate of the fields on a specific date. Furthermore, the physical and biological fields must be consistent with each other to reduce the impact of model equilibration ("shock") within the main simulation. The general method we have developed to do this, shown in Figs. 1 and 2, will now be briefly overviewed, and described in greater detail in Section 2.4.

First, initial estimates of the temperature and salinity distributions in the domain, including mesoscale structures of interest, are made. This is done by means of a physical "feature model", described in Section 2.4.1, which is based on the Anatomy of a Meander/GULFCAST data described in Section 2.3. Next, temperature and salinity data are melded into these fields to bring the field estimates as close to observed as possible. The melding scheme is described in Section 2.4.2. A first
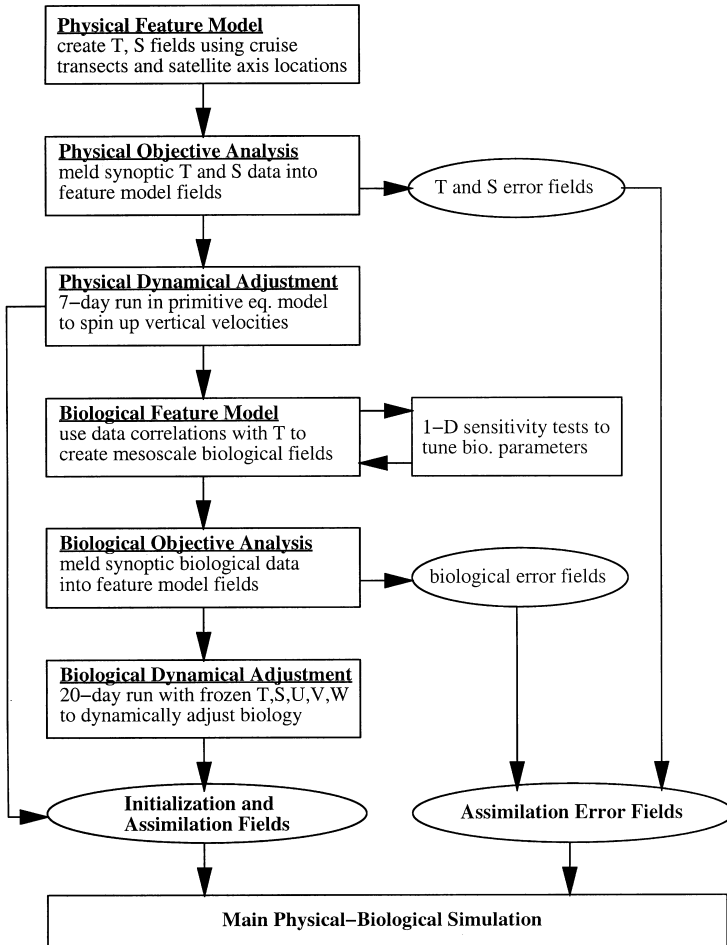
Fig. 1. Flowchart for creating compatible physical and biological initialization and assimilation fields.

approximation to the velocity fields is then made using the thermal wind relationship. Ageostrophic vertical velocities are generated by running the fields for several days in a primitive equation model, as described in Section 2.4.3. This completes the determination of the physical fields (temperature, salinity and velocity). Initial estimates are then made for the biological fields. As the biological data are sparse but relatively well-correlated with temperature, and as the biological distributions are desired to be consistent with the mesoscale physical features, initial estimates for the biological fields are made using the temperature distribution and biological-temperature correlations based on the BIOSYNOP data. This procedure is described in Section 2.4.4. 1-D sensitivity tests are carried out at this time in order to tune the biological model parameters to values that give the best agreement between the model and observations. Biological data are then melded into the initial estimate fields, as described in
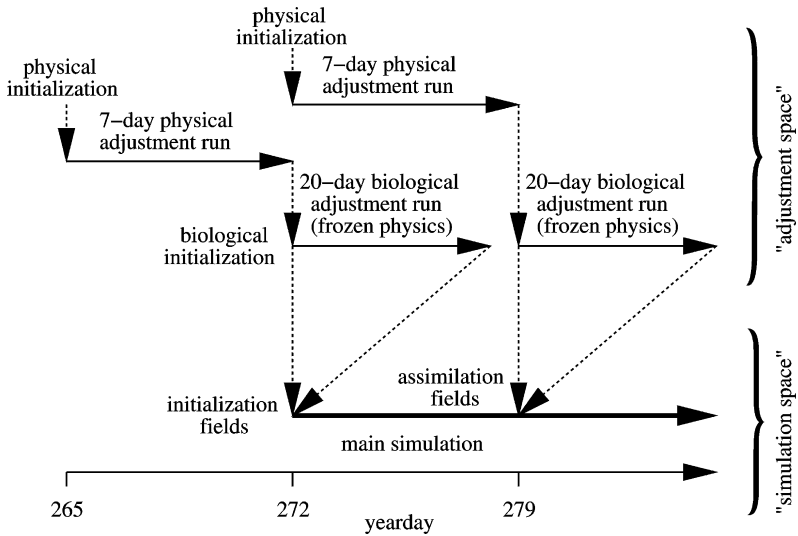
Fig. 2. Time chart of compatible physical and biological field generation and assimilation into the main simulation.

Section 2.4.5. Finally, the biological fields are dynamically adjusted to each other and the synoptic velocity field, as described in Section 2.4.6. This completes the procedure necessary to generate consistent physical and biological fields for initialization or assimilation. The temporal evolution of this procedure and the assimilation of these fields into the main simulation is illustrated in Fig. 2. The simulations conducted to generate the initialization and assimilation fields are done in what we call "adjustment space", outside of the main simulation of interest, which is in "simulation space".

The method used for data assimilation is an optimal interpolation scheme, described in Section 2.5. This method utilizes weights based on error fields. As indicated on the right-hand column in Fig. 1, these error fields are determined during the objective analysis procedure prior to the main simulation.

2.2. The model

The Harvard Ocean Prediction System (HOPS) is used (Lozano et al., 1996; Robinson, 1996). The physical and biological models are described in detail in the appendix. The physical model is a primitive equation model, run in a domain with 15 km horizontal resolution and 30 vertical levels. The biological model (Fig. 3, Tables 1–3) is of the nitrogen cycle. The five state variables are nitrate (N), ammonium (A), phytoplankton (P), zooplankton (Z), and DON plus suspended PON (D). These are advected and diffused in the physical model in the same manner as temperature and salinity:

$$\frac{\partial C}{\partial t} + \mathbf{V} \cdot \nabla C = \kappa_h \, \nabla_h^2 C + \frac{\partial}{\partial z} K_v \left(\frac{\partial C}{\partial z} \right) + S,$$
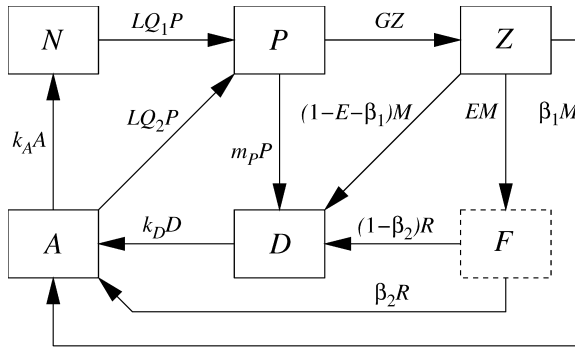
Fig. 3. Five-component biological model. The five state variables are nitrate (N), phytoplankton (P), zooplankton (Z), ammonium (A), and dissolved and suspended particulate organic nitrogen (D). Fast-sinking detritus (F) is included, but sinks and remineralizes instantaneously, and thus is not an explicit state variable.

Table 1
Biological model variables

i	depth indici of the current model level (nondimensional)
z	depth of the current model level (m)
Δz	vertical thickness of the current model level (m)
A	ammonium concentration (μM)
C	phytoplankton carbon-to-chlorophyll ratio (g C/g Chl)
D	semi-labile DON + suspended PON concentration (μM N)
E	fraction of zooplankton loss rate going into F (nondimensional)
F	fast-sinking detritus (implicit variable)
G	zooplankton grazing rate (day^{-1})
I	PAR light intensity (W m^{-2})
$J(i)$	fast-sinking detritus flux at the bottom of level i (mmol N m^{-2} day^{-1})
L_0	light limitation factor (nondimensional)
L	phytoplankton doubling rate (day^{-1})
M	zooplankton loss rate (μM N day^{-1})
N	nitrate concentration (μM)
P	phytoplankton concentration (μM N)
Q_1	nitrate limitation factor (nondimensional)
Q_2	ammonium limitation factor (nondimensional)
R	fast-sinking detritus remineralization rate (μM N day^{-1})
T	temperature (C)
Z	zooplankton concentration (μM N)

where C is an arbitrary biological concentration (in micromoles of nitrogen per liter), V is the 3-D velocity vector, κ_h is the horizontal diffusivity coefficient, K_v is the spatially and temporally variable vertical diffusivity coefficient, and S represents the biological source and sink terms shown in Fig. 3. Fast-sinking PON (F) is also included in the model, but does not have an explicit distribution, as it is parameterized as sinking and remineralizing instantaneously, as explained in the appendix.

Table 2
Biological Model Parameters

	Value	Description[a]	Reference[b]
k_w	0.04	light attenuation coefficient of seawater (m^{-1})	1
k_P	0.04	light attenuation coefficient of phytoplankton ($m^{-1}\,\mu M^{-1}$)	2
μ_0	5.0	maximum photosynthetic rate (g C (g Chl h)$^{-1}$)	3
α	0.142	initial slope of P–I curve (g C (g Chl h)$^{-1}$ (W m^{-2})$^{-1}$)	3
β	0.0	photoinhibition factor (g C (g Chl h)$^{-1}$ (W m^{-2})$^{-1}$)	3
λ	3.5	factor for NO_3 uptake inhibition by NH_4 (μM^{-1})	4
k_1	0.01	nitrate half-saturation constant (μM)	4
k_2	0.01	ammonium half-saturation constant (μM)	4
r_m	1.5	maximum zooplankton grazing rate (day^{-1})	4
Λ	1.5	Ivlev constant for zooplankton grazing (μM^{-1})	4
k_A	0.25	nitrification rate (day^{-1})	4
m_P	0.10	phytoplankton mortality rate (day^{-1})	4
γ	0.20	fraction of zooplankton grazing excreted (nondim.)	4
n_1	0.10	linear zooplankton loss rate (day^{-1})	4
n_2	1.0	quadratic zooplankton loss rate (day^{-1} μM^{-1})	4
β_1	0.25	fraction of zooplankton loss to NH_4 (nondim)	4
β_2	0.25	fraction of fast-sinking detritus loss to NH_4 (nondim.)	4
k_D	0.02	remineralization rate of D (day^{-1})	4

[a]All molar units refer to moles of nitrogen per liter of seawater
[b]References: 1 = Sarmiento et al. (1993); 2 = McGillicuddy et al. (1995); 3 = Lohrenz et al. (1993); 4 = tuned in this study

Table 3
Biological auxiliary equations

$$I(i) = I(i-1)\exp(-(k_w + k_P P)\Delta z) \tag{1}$$

$$L_0 = (1 - \exp(-\alpha I/\mu_0))\exp(-\beta I/\mu_0) \tag{2}$$

$$Q_2 = A/(k_2 + A) \tag{3}$$

$$Q_1 = \min(e^{-\lambda A}N/(k_1 + N), 1 - Q_2) \tag{4}$$

$$C = \max(20, \exp(4.62 - 0.014|z|)) \tag{5}$$

$$L = \mu_0 \min(1, L_0/(Q_1 + Q_2))1.066^{T-22}/C \tag{6}$$

$$G = r_m(1 - \exp(-\Lambda P)) \tag{7}$$

$$M = \gamma G Z + n_1 Z + n_2 Z^2 \tag{8}$$

$$E = \min(\max(0.01, \min(0.50, 4.9P - 0.235)), 0.99 - \beta_1) \tag{9}$$

$$J(i) = J(i-1)\exp(-\Delta z/\max(100, 1.15|z|)) + EM\Delta z \tag{10}$$

$$R = (J(i-1) - (J(i) - EM\Delta z))/\Delta z \tag{11}$$

2.3. The data

Two primary data sets were used for model initialization, assimilation and verification: the combined BIOSYNOP/Anatomy of a Meander data set and the

GULFCAST data set, which will now be described. In additon, other data sets (e.g. BATS, TTO/NAS) were used for model calibration purposes, as noted in other sections, but were not assimilated directly.

The Fall BIOSYNOP cruise was made aboard the *R/V Cape Hatteras* between 21 Sept and 21 Oct 1988 (Lohrenz et al., 1993); observations that we utilize include 122 XBT stations, 124 CTD stations and 32 nutrient (NO_3 and NO_2) stations. The Fall Anatomy of a Meander cruise was made aboard the *R/V Endeavor* between 17 Sept and 13 Oct 1988 (Hitchcock et al., 1993; Mariano et al., 1996); the data we use consist of 160 XBT stations and 92 CTD stations, the latter including fluorometer (chlorophyll) measurements. An additional 38 AXBT stations were made on 13 Sept and 4 Oct. The BIOSYNOP/Anatomy of a Meander CTD station locations between yeardays 261 and 290 are shown in Fig. 4 as crosses. During this period the *R/V Endeavor* made 4 triangular sections around a Gulf Stream meander, while the *R/V Cape Hatteras* wove back and forth across the stream axis. As can be seen, this is an interdisciplinary data set of good mesoscale spatial and temporal resolution. For brevity, we will hereafter refer to the combined BIOSYNOP/Anatomy of a Meander cruise data as BIOSYNOP.
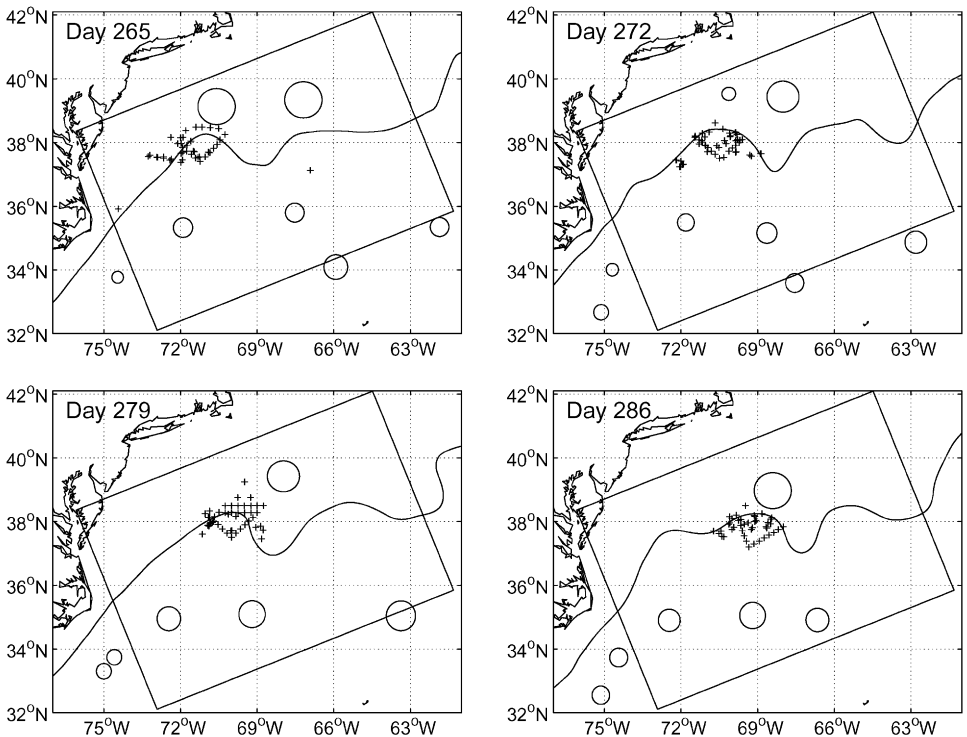

Fig. 4. Solid lines: Weekly estimates of Gulf Stream axis and ring locations from GULFCAST for 1988. Crosses: CTD locations taken by the *R/V Endeavor* and the *R/V Cape Hatteras* during BIOSYNOP between yeardays 261 and 290, binned weekly. The model domain is also shown.

The GULFCAST data set consists of weekly analyses of Gulf Stream axis and ring positions, derived from AVHRR sea surface temperature, GEOSAT altimetry and AXBTs (Robinson et al., 1989a, b; Glenn and Robinson, 1995; Gangopadhyay and Robinson, 1997; A. Gangopadhyay, unpublished data). Stream axis locations were further updated using the Anatomy of a Meander transect data. The GULFCAST data set is also shown in Fig. 4. Two major events occurred in the vicinity of the data during the observational period: the absorption of a warm-core ring between yeardays 272 and 279, and the formation of a cold-core ring following yearday 286.

As discussed in Sections 2.4.2 and 2.4.5, the data objectively analyzed into initialization and assimilation fields are temperature, salinity, nitrate and chlorophyll. Zooplankton biomass estimates from BIOSYNOP (Ashjian, 1993; Ashjian et al., 1994) were used for model calibration (Section 2.4.4), and primary production estimates (Lohrenz et al., 1993) for model verification.

During Leg 1 of BIOSYNOP (yearday 261–280) surface nitrate values were low, suggesting stratified conditions typical of late summer. During Leg 2 high surface nitrate values were encountered, suggesting the beginning of autumn mixed layer deepening. Since in our study (Anderson and Robinson, 2000) we wished to examine the nutrient transport fluxes during summer-like conditions, we restricted our simulation to the time period of Leg 1.

2.4. Estimation of consistent physical and biological fields

The following subsections describe the method used to generate synoptic, consistent physical and biological fields as outlined in Fig. 1. This uses a combination of data, parameterizations, objective analysis techniques and model simulations carried out in "adjustment space" (Fig. 2).

2.4.1. Initial estimates of physical fields: the physical feature model
We wish to initialize the model with temperature and salinity fields for the Gulf Stream density structure, meander positions and ring locations as they were on 21 Sept 1988 (yearday 265). Toward this end we cannot directly use climatological data atlases such as Levitus (1982), which has a greatly broadened Gulf Stream and no rings or meanders. We used instead a data-based "feature model" (Gangopadhyay et al., 1997) which estimates the Gulf Stream structure by spatially extending (horizontally and vertically) the realtime data through parameterized structures. From the eight cross sections of the Gulf Stream made by the R/V Endeavor during BIOSYNOP, mean temperature and salinity cross sections are constructed. These cross sections are placed in the model domain along the location of the Gulf Stream axis as determined from GULFCAST (Fig. 4) for the date in question. To the north and south of the Gulf Stream, the last profile of the cross section is persisted; thus, the Sargasso and Slope waters are horizontally homogeneous. Below 840 m (the depth of the cross-sections), all waters decay exponentially with depth to the same properties ($T = 4.26°C$, $S = 35.01$). The deep western boundary current is not included. Idealized ring structures of sizes determined from the AVHRR imagery are placed at observed

locations. The result is an idealized but synoptic 3-D estimate of the Gulf Stream structure and location for the date in question, as shown in Fig. 5.

Feature models are also necessary for creating assimilation fields. For instance, while it is desirable to assimilate SST data into the model, SST is insufficient to direct the Gulf Stream northward if the deeper structure is directing it southward. Sea surface information needs to be extrapolated with depth, and the feature model does this, using the GULFCAST SST information on Stream axis and ring locations to estimate the deeper structure.

2.4.2. Objective analysis of physical data

Into the idealized temperature and salinity fields described above, we meld in the BIOSYNOP T and S data. Incorporating the available data gives us the best estimate of the fields at that time and is used to verify the feature model, i.e. to confirm that the initial field estimates agree well with the data at their locations in space and time.

The melding procedure is based on Carter and Robinson (1987) with modifications. We first used the Carter and Robinson (1987) scheme, but found it to overestimate the range of observed values near the sharp Gulf Stream frontal gradient. A simplified scheme that does not overestimate was therefore used instead.

All observations are first vertically interpolated to flat (model level) depths, and objective analyses performed in the x–y plane. The value of the newly estimated field at a specific time and grid location is determined by

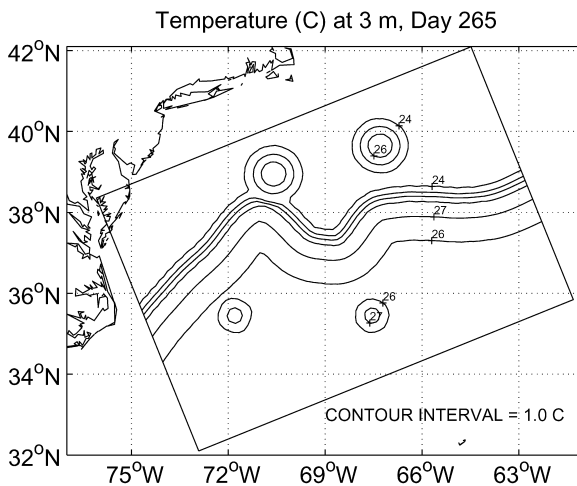$$T_j^{oa} = \sum_{i=1}^{i_{max}} w_{ij} T_i^{obs} / \sum_{i=1}^{i_{max}} w_{ij},$$

Fig. 5. Feature model estimate of sea surface temperature for yearday 265, 1988.

where T_j^{oa} is the estimated value at (x_j, y_j, t_j), T_i^{obs} is an observation at (x_i, y_i, t_i), and the weight w_{ij} is

$$w_{ij} = \exp(-((t_j - t_i)/t_d)^2 - ((x_j - x_i)/x_d)^2 - ((y_j - y_i)/y_d)^2),$$

where t_d, x_d and y_d are temporal and spatial correlation scales. We use $t_d = 4$ days and $x_d = y_d = 20$ km, the latter approximately corresponding to the spacing of the BIO-SYNOP observations. Although water parcels in the Gulf Stream are themselves advected very quickly (up to 150 km/day^{-1}), properties associated with meanders (in the Eulerian framework) change much more slowly, such that these correlation scales are reasonable (Hitchcock et al., 1993; Mariano et al., 1996). In this way, observations made over several days are collected into a single field on a specific day (with the data weighted by its distance in time and space).

The feature model described in Section 2.4.1 is used as an initial estimate. The initial estimate at a given location is treated as an observation with $w_{ij} = 0.2$. The value 0.2 was determined empirically so that at the location of an observation, the observation has 5 times the weight of the initial estimate; 25 km from an observation, it has equal weight with the initial estimate; at 50 km, the observation weight is 1% of the initial estimate weight. Thus, the maximum radius of influence of an observation is approximately 50 km. Away from all observations the objectively analyzed field reverts to the feature model field.

In addition, associated with each estimated field is a normalized error field, $e_j = 1 - \max(w_{ij})$ (excluding the weight of the feature model field). Therefore, at a time and place where an observation is present, the error value is zero, and far from the data the error value is one. This error field is used during data assimilation, as described in Section 2.5. Poor data accuracy can be accounted for by increasing the error values.

An objectively analyzed temperature field and its corresponding error field are shown in Fig. 6.

2.4.3. Physical dynamical adjustment

From the T and S fields, geostrophic velocity fields are then computed using the thermal wind relationship and a level of no motion of 1210 m. The resulting horizontal velocity cross section of the Gulf Stream agrees well with Halkin and Rossby (1985).

The Gulf Stream, however, is not purely geostrophic. If a simulation is started from a geostrophic initialization, large vertical velocities associated with Gulf Stream meanders and rings are immediately generated, and it takes several days for the vertical velocities to equilibrate. It is critical to determine these ageostrophic vertical velocities because of their significant impact on biological fields. Therefore, all physical fields to be used for initialization or assimilation are first run in the primitive equation model for 7 days, to allow this dynamical adjustment to occur. By the 7th day the adjustment is essentially over, with no further qualitative and little quantitative change in vertical velocities. Dynamically adjusted temperature, horizontal velocity and vertical velocity fields are shown in Fig. 7. As discussed in Anderson and Robinson (2000), the vertical velocities agree well with observations (e.g. Lindstrom and Watts, 1994) both qualitatively and quantitatively.
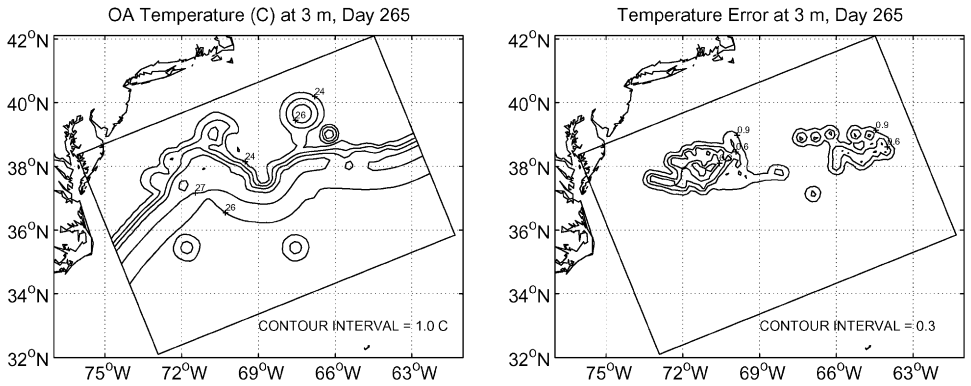
Fig. 6. Estimate of sea surface temperature (in C) for yearday 265, 1988, after melding in the BIOSYNOP temperature observations. The error field indicates where data have been incorporated. Away from the data the error value is 1.0.
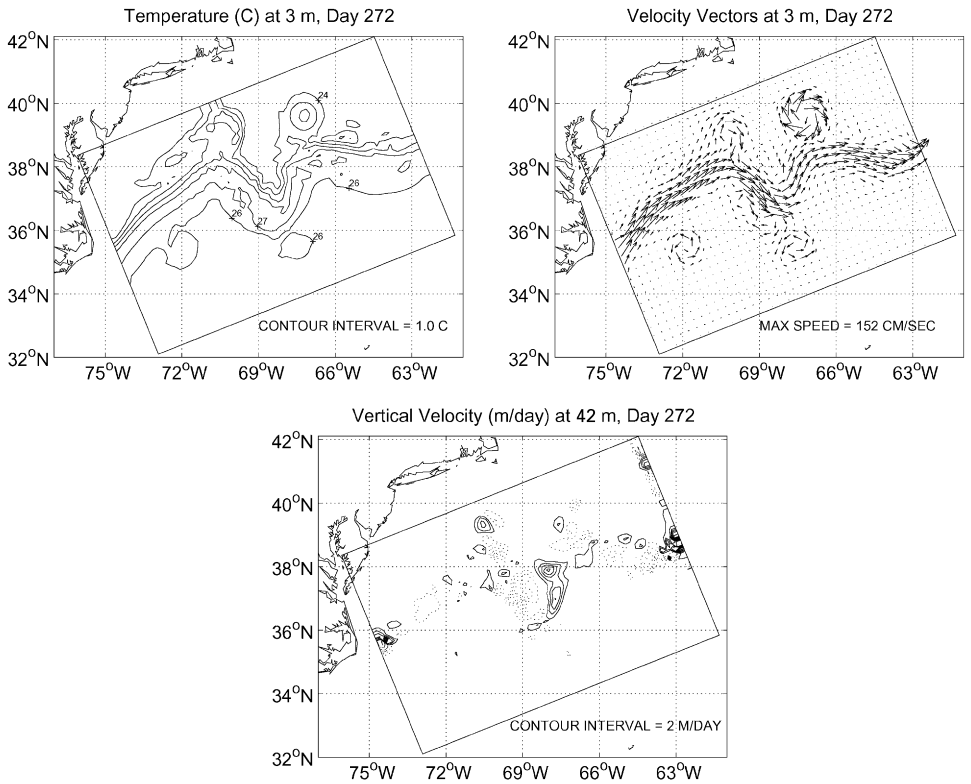

Fig. 7. Temperature (C), horizontal velocity (cm/sec) and vertical velocity (m/day) after the 7-day physical dynamical adjustment run, i.e. for yearday 272. Horizontal velocity is subsampled every other grid point.

In addition, preliminary simulations were carried out at this point; minor adjustments were made to initial ring and stream positions to ensure that the physical evolution was consistent with later data, to ensure smooth assimilation of the later physical fields.

This dynamical adjustment completes the preparation of the physical fields. Since the physical fields were adjusted for 7 days, the initial realization is now no longer for day 265 but for day 272. This is used to initialize the main simulation. A field for day 279 is constructed in the same manner (Fig. 2), to be assimilated into the main simulation.

2.4.4. Initial estimates of biological fields

The three-step procedure used to generate the physical fields (making an initial estimate, melding in available data, dynamical adjustment; see Fig. 1) is also used to generate the biological fields. Because we wish the biological distributions to coincide with the mesoscale physical features (e.g. the boundary in Slope versus Sargasso phytoplankton to coincide with the location of the physical Gulf Stream front), initial estimates for the biological fields cannot be directly based on climatological biological atlases or even parameterized spatial structures. Instead we use a "biological feature model", which uses the temperature distribution derived in Section 2.4.3 and data-based correlations between temperature and biological properties to make initial estimates of the biological distributions.

Let us begin with nitrate. First, model temperature profiles at grid points ($i = 70$, $j = 35$) and ($i = 70, j = 5$) were assumed to be representative of Slope and Sargasso waters. Corresponding nitrate profiles were then determined using the piecewise-linear NO^{-3}-temperature relationship of the BIOSYNOP and TTO/NAS data (Fig. 8). However, this relationship breaks down near the sea surface, and was
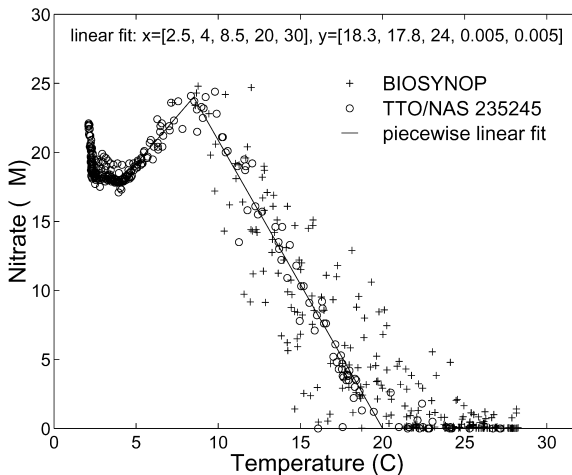

Fig. 8. Nitrate versus temperature trend from the BIOSYNOP data and TTO/NAS stations 235–245. The solid line is the piecewise linear fit used to make an initial estimate for nitrate (see text).

therefore applied only below 200 m. Nitrate was set to 0.005 μM in both Sargasso and Slope waters from the surface to 80 and 30 m, respectively; a linear gradient with depth was used between these depths and 200 meters. Lastly, these two nitrate profiles were horizontally interpolated as a linear function of temperature across the model domain, temperature serving as an indicator of watermass.

The phytoplankton distribution was initialized by first determining typical chlorophyll profiles for Slope and Sargasso waters during BIOSYNOP (Fig. 9), and converting these from mg Chl m^{-3} to μM N using a depth-dependent C/Chl ratio (Table 3, Eq. (5); from Malone et al., 1993), and a C/N ratio of 6.625 mol/mol (Redfield et al., 1963). These two phytoplankton profiles were then assumed to be associated with the Slope and Sargasso temperature profiles previously described, and were horizontally interpolated linearly with temperature.

The BIOSYNOP zooplankton data made available to us were relatively sparse. The spatial distribution of zooplankton essentially mimics that of phytoplankton, its prey, i.e. decreasing with depth, and with higher biomass in Slope than Sargasso water (Ashjian et al., 1994). As such, the zooplankton distribution was initialized as proportional to the phytoplankton field. A proportionality constant of 1.0 (mol N/mol N) was used, based on estimates from the BIOSYNOP data (Lohrenz et al., 1993; Hitchcock et al., 1993; Ashjian et al., 1994; Mariano et al., 1996).

Measurements of NH_4 (unpublished) were made during BIOSYNOP but appear uncalibrated (negative values, high positive values). In the open ocean NH_4 has been observed to have a subsurface maximum near the Chl maximum (Brzezinski, 1988). Thus, NH_4 was also initialized as proportional to the phytoplankton distribution, using a factor of 0.2 so that a concentration of 0.01 μM NH_4 is obtained for the Sargasso mixed layer (Brzezinski, 1988).

There is no BIOSYNOP data for (non-living) suspended PON or DON. Other data sets however indicate that their vertical distribution is similar to plankton, i.e.
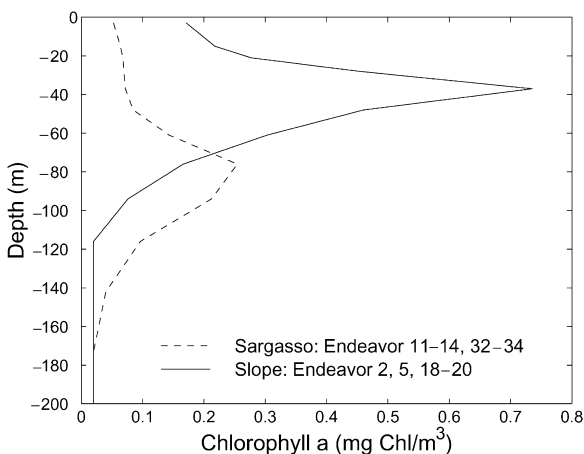

Fig. 9. Chlorophyll profiles for Sargasso water based on *R/V Endeavor* stations 2, 5, 18–20 and for Slope water stations 11–14, 32–34.

decreasing with depth, and higher in Slope than Sargasso water. The BATS data (Michaels et al., 1994b; http://www.bbsr.edu/) indicate that in the Sargasso in late summer POC values in the euphotic zone are on average 25 mg C/m^3. For Slope waters, Joyce et al. (1992) estimate suspended particulate matter in the upper 90 m in summer at typically 165 mg dry wt/m^3, at most 50% of which could be organic carbon. We therefore estimate non-living suspended PON is approximately 0–5 times the phytoplankton concentration in both Sargasso and Slope waters in summer, depending upon what fraction of the observed PON is "suspended" (effectively, sinking rate < 1 m day^{-1}) and labile on a timescale < 100 days. DON is also included in state variable D. Surface DON concentrations in September near Bermuda are near 5 μM N (Hansell and Carlson, 2000; BATS data at http://www.bbsr.edu/). However, this pool changes very little during summer, such that only a small fraction is expected to be labile on short timescales. If we consider the labile fraction to be 10% (i.e. 0.5 μM N), then the total concentration of D should be approximately 10–15 times the phytoplankton concentration.

   We cannot presume to be able to make a more accurate estimate for D, in part because of the arbitrariness in the definitions of "suspended" PON (sinking rate < 1 m day^{-1}? < 0.1 m day^{-1}?) and "semi-labile" DON (remineralization time-scale = 10 days? 100 days?), and because of the uncertainty in observed concentrations for specific definitions. Fortunately, an accurate estimate of D is not vital. Firstly, it is not one of our variables of interest (since we have no data for assimilation or validation). Secondly, it is not an independent variable in the ecosystem equations (Fig. 3 and Table 3). That is, the only term in which it is involved is $k_D D$ (see Fig. 3), which supplies NH$_4$ to fuel recycled production. The values of both D and k_D are poorly known, which means there are two degrees of freedom (their values), but only one constraint (the observed production rate). We therefore are allowed to "choose" one, as long as the other is "tuned" so that enough NH$_4$ is supplied to match the observed production rate. The most sensible choice of D is at its lower estimate, which allows a higher k_D and a faster dynamical adjustment (allowing the D distribution to reach near-equilibrium more quickly), which is important for reasons given in Section 2.4.6. We therefore set the initial estimate for D as 10 times the phytoplankton concentration. It turns out the best k_D value corresponding with this D concentration is (50 days)$^{-1}$, in good agreement with estimates for semi-labile DOM (Archer et al., 1997), which gives us confidence in these D and k_D estimates.

   It should be noted that these rough, initial estimates are not crucial because during the dynamical adjustment (Section 2.4.6) the biological tracers will adjust to be consistent with each other, the biological equations and the physical circulation. We do however make the best possible estimates based on the data (rather than starting from completely arbitrary initial values) because it (a) allows a shorter biological dynamical adjustment period, and consequently better agreement with the data (see Section 2.4.6), and (b) minimizes the danger of the state variables in phase space being attracted towards a local minimim that is not the one closest to the observed state.

   Values for the biological model parameters (Table 2) are also needed. First, values from the literature were chosen as starting points. The parameters were then adjusted (within their ranges of uncertainty) so that the size of the terms in Fig. 3 are near the

observed rates and approximately balance. 1-D sensitivity tests were then carried out in Sargasso and Slope waters to further tune the biological model parameters, parameterizations and initial conditions. The goal was for the biological variables after 20 days of simulation to have values near observed and a small time rate of change. Bermuda Atlantic Time-series Study (BATS) data (Michaels and Knap, 1996) were used to help calibrate the model in Sargasso waters. Variables not adjusting to values near observed indicated bad parameter values or parameterizations. These sensitivity tests therefore led to the refinement of the biological model. It was found that the variables at day 20 were relatively insensitive to the uncertainty in their initial estimates, except for D, which changes slowly due to its slow remineralization rate. In future work, these 1-D sensitivity tests will be replaced with automated data-fitting (inverse) methods (e.g. Fasham and Evans, 1995) to determine optimal parameter values and evaluate parameterizations.

2.4.5. Objective analysis of biological data

Into these initial estimate biological fields we meld the BIOSYNOP biological data. The objective analysis scheme for biological data is the same as that used for physical data, described in Section 2.4.2. The only biological data available are nitrate and phytoplankton biomass, the latter determined from chlorophyll using the depth-dependent C/Chl ratio and constant C/N ratio as described in Section 2.4.4. Zooplankton data were sparse and therefore used only for model calibration. The correlation time- and length-scales used for nitrate and phytoplankton are identical to those used for the physical fields (Steele and Henderson, 1992; Denman and Abbott, 1994).

As the Chl distribution in nature is typically log-normal (Campbell, 1995), a few studies (e.g. Denman and Freeland, 1985; Denman and Abbott, 1994) objectively analyze $\log_{10}(\text{Chl})$ rather than Chl concentration itself in order to properly satisfy requirements in the gridding algorithm based on statistical interpolation (e.g. for the OA estimate to equal the maximum likelihood estimate). In this study we chose to objectively analyze phytoplankton concentration itself because the simulated phytoplankton fields are not log-normal but primarily bimodal; in the simulations the phytoplankton scales are essentially the same as those of temperature and NO_3. It remains to be explored how log-normal data are best assimilated into a simulation with finite grid resolution.

2.4.6. Biological dynamical adjustment

The final step is to conduct a biological adjustment simulation (see Fig. 1). That is, while we have made initial estimates for the biological fields and parameters (Section 2.4.4) and melded in data (Section 2.4.5) the biological fields are not necessarily consistent with each other, with the model parameters or with the 3-D physical transports. Consequently, simulations typically undergo an immediate bloom or decline solely as an adjustment to the initial conditions. We expect the biological fields to be primarily in balance, except possibly along the Gulf Stream itself, because the biological concentrations in the Slope and Sargasso waters are not observed to change greatly during late summer. In addition, in this study we want to examine

changes due to Gulf Stream processes alone, and therefore do not want the background fields in the Slope and Sargasso waters to be changing greatly. Like with the physical initialization, a dynamical adjustment is necessary, and this is done by conducting a simulation with the physical–biological model using frozen (fixed) temperature, salinity and velocity fields. In this way, the biological fields are given time to achieve greater balance with the desired 3-D physical fields and each other.

It is necessary for all the biological state variables to dynamically adjust simultaneously. For example, it was also tried to keep the NO_3 field fixed while allowing the other biological fields to adjust. However it was found that, once the main simulation was started, the NO_3 field would rapidly change. This is because, when NO_3 is held fixed, the biological variables do not go to the true steady-state solution that is determined by the parameter values. Also, the initial NO_3 field estimate may not be in balance with the physical transports.

It is desirable for the adjustment simulation to be as short as possible because (a) after the rapid initial adjustment the biological state often has a slow drift (see below), (b) the NO_3 gradient across the front weakens as the run progresses (see below), and (c) in reality the biology at the front is not in steady state. Each of these cause the biological fields to depart from observed (the initial conditions) as the adjustment run progresses. Nevertheless, the adjustment run must be long enough to get through the initial phase of rapid adjustment.

The biological adjustment run is conducted for 20 days. The biological fields adjust rapidly in the first few days and by day 20 have largely equilibrated. Phytoplankton adjust quickly (approx. 7 days), but it takes 20 days for the more slowly-changing fields (viz. D) to settle down. It is not necessary that the rate of change on the final day be zero; it should be the size observed. The time rate of change on day 20 is very small (<0.01 μM N/day) except for D in Slope waters, whose change per day is still less than 2% of its total concentration, and for NO_3, which is decreasing at the top of the nutricline due to entrainment into the mixed layer. Both N and D are effectively being redistributed to greater depths. As DOM does slowly decrease during summer (Michaels et al., 1994a) and the nutricline does deepen (Strass and Woods, 1991), this slow drift is consistent with observations. For a simulation during a more active period (e.g. the spring bloom), a much shorter adjustment period should be used.

The final, adjusted biological fields are shown in Fig. 10. The characteristics of these fields are discussed in detail in Anderson and Robinson (2000). Fields for day 272 are used for biological intialization and fields for day 279 for assimilation.

Fig. 11a shows how the dynamically adjusted NO_3 values along $i = 60$ have changed from their initial estimates. In both Slope and Sargasso waters NO_3 stays near the initial-estimate T-NO_3 trend. In Slope waters, the increased NO_3 at 20°C are within the range of observations (Fig. 8). In the Gulf Stream, horizontal and vertical advection and diffusion cause the T-NO_3 values to start to fall off the line. This illustates why a short dynamical adjustment period is desirable: advection and diffusion gradually deteriorate the NO_3 gradients across the front, particularly in the deep ocean, because the large-scale physical and biological processes that maintain the front are not included in the subdomain.
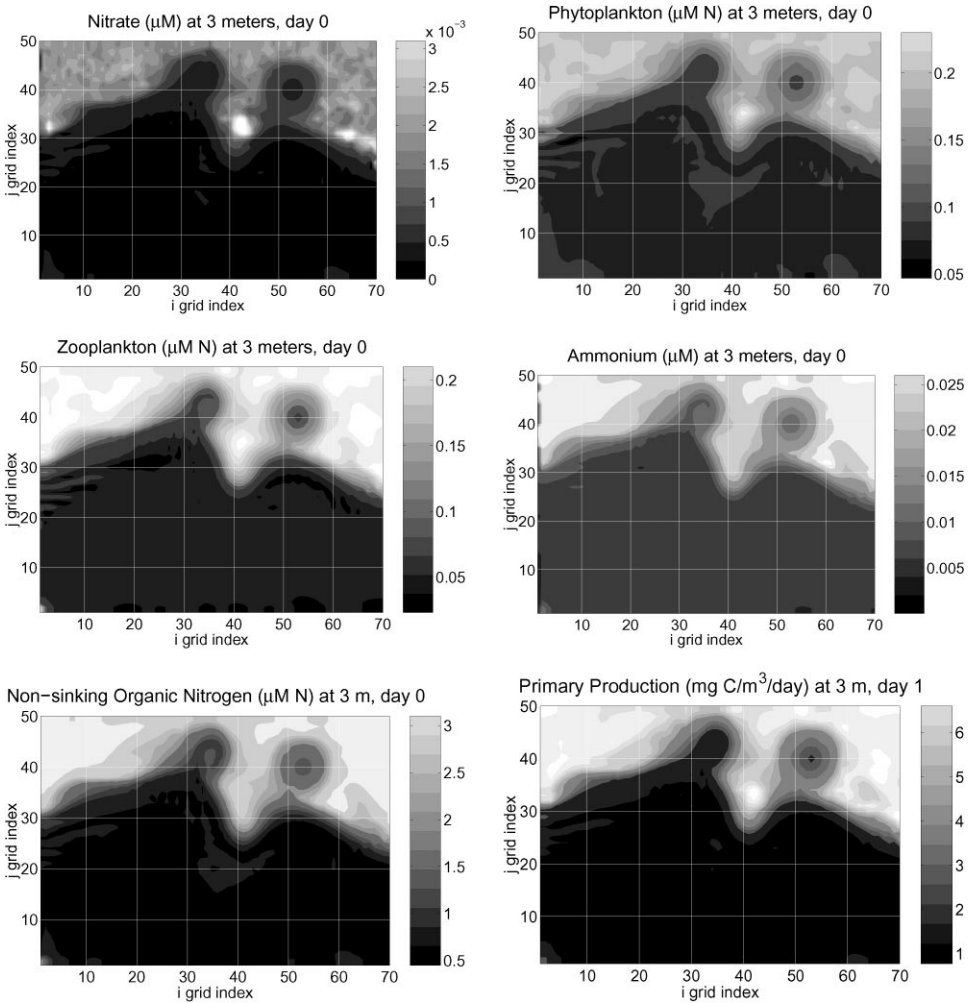
Fig. 10. Dynamically adjusted biological fields for yearday 272, used to initialize Runs 6 and 7.

In Slope water, phytoplankton at the sea surface remain close to their initial estimate (Fig. 11b). The subsurface phytoplankton maximum is low by a factor of two, but this is within the range of observed profiles for Slope water, which were highly variable. In addition, the initial-guess P maximum may be an overestimate (see Anderson and Robinson, 2000). The depth of the subsurface maximum agrees well with the observations. Zooplankton in Slope water mimic phytoplankton with approximately a 1 : 1 ratio, consistent with the initial assumptions. In Sargasso water, phytoplankton agree well with the initial estimate (data). Zooplankton mimic phytoplankton, but are lower by almost a factor of two. This is acceptable given that the initial assumption of a 1 : 1 ratio was very approximate, and is within the uncertainty bounds of the data.
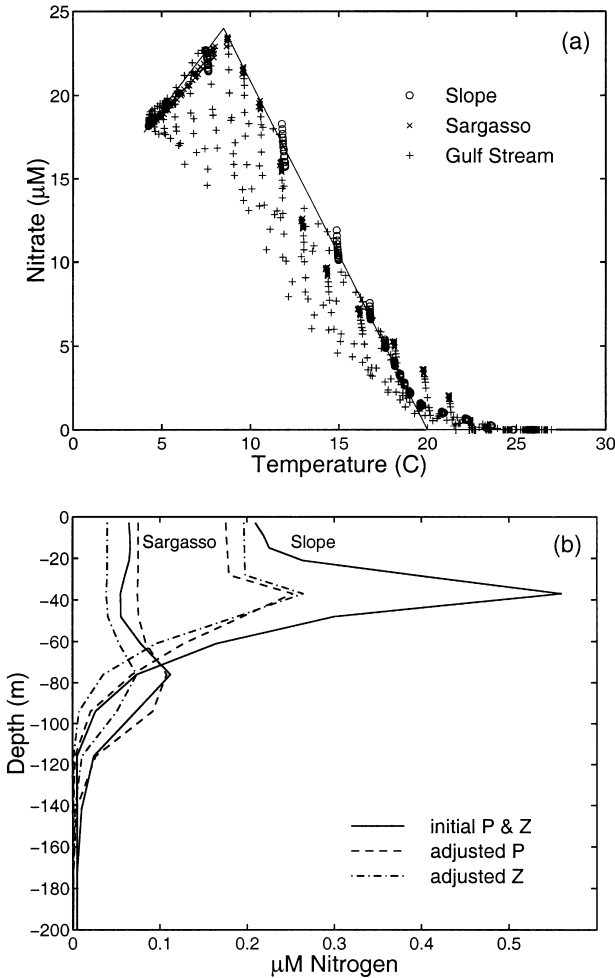
Fig. 11. (a) Dynamically adjusted NO_3 values along $i = 60$ for Slope (o), Sargasso (x) and Gulf Stream (+) water, compared to the initial-estimate T-NO_3 trend (straight lines). (b) Dynamically adjusted phytoplankton (dashed) and zooplankton (dot-dashed) profiles for Slope and Sargasso water, relative to their initial estimates (solid lines).

2.5. Data assimilation scheme

Finally, we describe the data assimilation method, which is applied to dynamically adjusted assimilation fields derived in Section 2.4. The assimilation method used is optimal interpolation (Lozano et al., 1996). On an assimilation timestep (which is once per day) the model fields are updated using the following equation:

$$T_{new}^{model} = wT^{oa} + (1 - w)T_{old}^{model},$$

where T^{model} is the model value, T^{oa} is the field being assimilated, and

$$w = s(t)*(w_{max} - (w_{max} - w_{min})*e^{oa}),$$

where e^{oa} is the error field of the assimilated field described in Section 2.4.2, w_{min} and w_{max} are minimum and maximum weight values ($0 \leqslant w_{min} \leqslant w_{max} \leqslant 1$), and $s(t)$ is the time dependent ramping factor ($0 \leqslant s(t) \leqslant 1$) shown in Fig. 12. When $s(t) = 1$, the value of w at observation locations ($e^{oa} = 0$) is w_{max} and at feature model locations ($e^{oa} = 1$) is w_{min}. In simulations assimilating fields into which data have not been melded (see Section 3), we use $w_{max} = 1$ and $w_{min} = 1$ to allow full and homogeneous assimilation of the fields. In simulations assimilating fields into which data have been melded, we use $w_{max} = 1$ and $w_{min} = 0.5$ to allow full assimilation of fields where there is data and partial assimilation of the background estimates. The value 0.5 was arbitrarily chosen to reflect partial confidence in the feature model estimates. The error fields used are those determined by the objective analysis technique (see Section 2.4.2).

The assimilated physical variables include temperature, salinity, and velocity (in the form of zonal and meridional baroclinic velocities and barotropic streamfunction). Although no observations for velocity are used in this study, the dynamically balanced velocity is assimilated with an error field identical to that of the temperature field. All biological variables are assimilated.

The value of $s(t)$ (shown in Fig. 12) allows an assimilation field to be ramped in gradually in order to minimize shock to the model. For example, the observed field for day 7 is gradually introduced into the model over days 4–7. Ramping down $s(t)$ on the 3 days following maximum assimilation (day 7) was also tried, but had adverse effects. It is desired for the correlation between the model and the day-7 assimilation fields to be a Gaussian (bell-shaped) curve centered on day 7. It was found that ramping down (i.e. re-assimilating) the day-7 assimilation field on days 8, 9 and 10 persists the model fields as they were on day 7. The result is that the correlation between the model and the assimilation field remains very high on days 8–10, and the correlation curve is not bell-shaped in time. (This persistence also makes the model less prepared on day 11 for the assimilation of the day-14 field than if the model were allowed to evolve freely, given the physical model has forecast skill; Gangopadhyay and Robinson, 1997.) Using a saw-toothed shape for $s(t)$ as shown in Fig. 12 allows the correlation to be bell-shaped, because the model state will naturally decay away from the assimilation field after day 7. Another advantage of this $s(t)$ shape is that the three days following
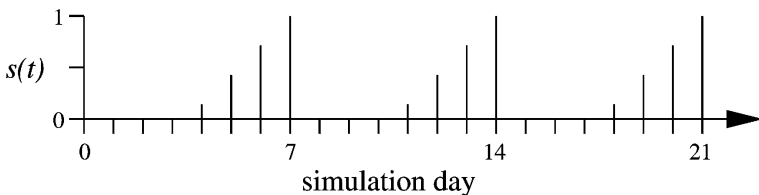

Fig. 12. Time-dependent ramping factor $s(t)$ for data assimilation. The factor increases linearly from 0 at day 3.5 to 1.0 at day 7.0, with one assimilation timestep per day.

day 7 can be examined to see how the model dynamics are responding to the assimilated field and assimilation procedure.

3. Results

The first series of simulations (Table 4, Runs 1–5) were tests conducted to examine the effects of the physical and biological data assimilation methodology on the simulations. Each of these runs were initialized with dynamically adjusted feature model fields (i.e. without BIOSYNOP data melded in), in order to keep them more idealized so that processes and relationships could be examined more clearly. Runs 6–8 are discussed in Anderson and Robinson (2000). All simulations were run for 10 days. In this paper we focus on the success of the data assimilation scheme; Anderson and Robinson (2000) discusses the physical and biological processes in detail.

3.1. Runs 1 and 2: No assimilation

Run 1 was initialized with the dynamically balanced fields for yearday 272 and run for 10 days with no data assimilation. Fig. 13 shows sea surface temperature, vertical velocity at 42 m, surface nitrate concentrations, and vertically integrated phytoplankton concentrations on days 7 and 10 of the simulation. Temperature indicates significant interaction with both warm core rings and cold core rings. Vertical velocities are most intense at the trough in the center of the domain; high vertical velocities are also associated with ring-stream interactions. Although nitrate concentrations are not particularly enhanced in the central trough, phytoplankton concentrations are slightly enhanced there.

Run 2 was initialized with a dynamically balanced realization for yearday 279, i.e. 7 days later than Run 1. As such, days 0 and 3 of Run 2 (shown in Fig. 14) can be

Table 4
Model simulations

Run	Start day	Fields[a]	Rings[b]	Assimilation[c]	Wind forcing[b]
1	272	FM	Y	N	N
2	279	FM	Y	N	N
3	272	FM	Y	P	N
4	272	FM	Y	B	N
5	272	FM	Y	P + B	N
6	272	FM + data	Y	P + B	N
7	272	FM + data	Y	P + B	Y
8	279	FM	N	N	N

[a] FM = feature model fields, dynamically adjusted; FM + data = feature model fields with data melded in and dynamicaly adjusted.
[b] Y = yes; N = none
[c] P = physical data assimilation; B = biological data assimilation; N = none.
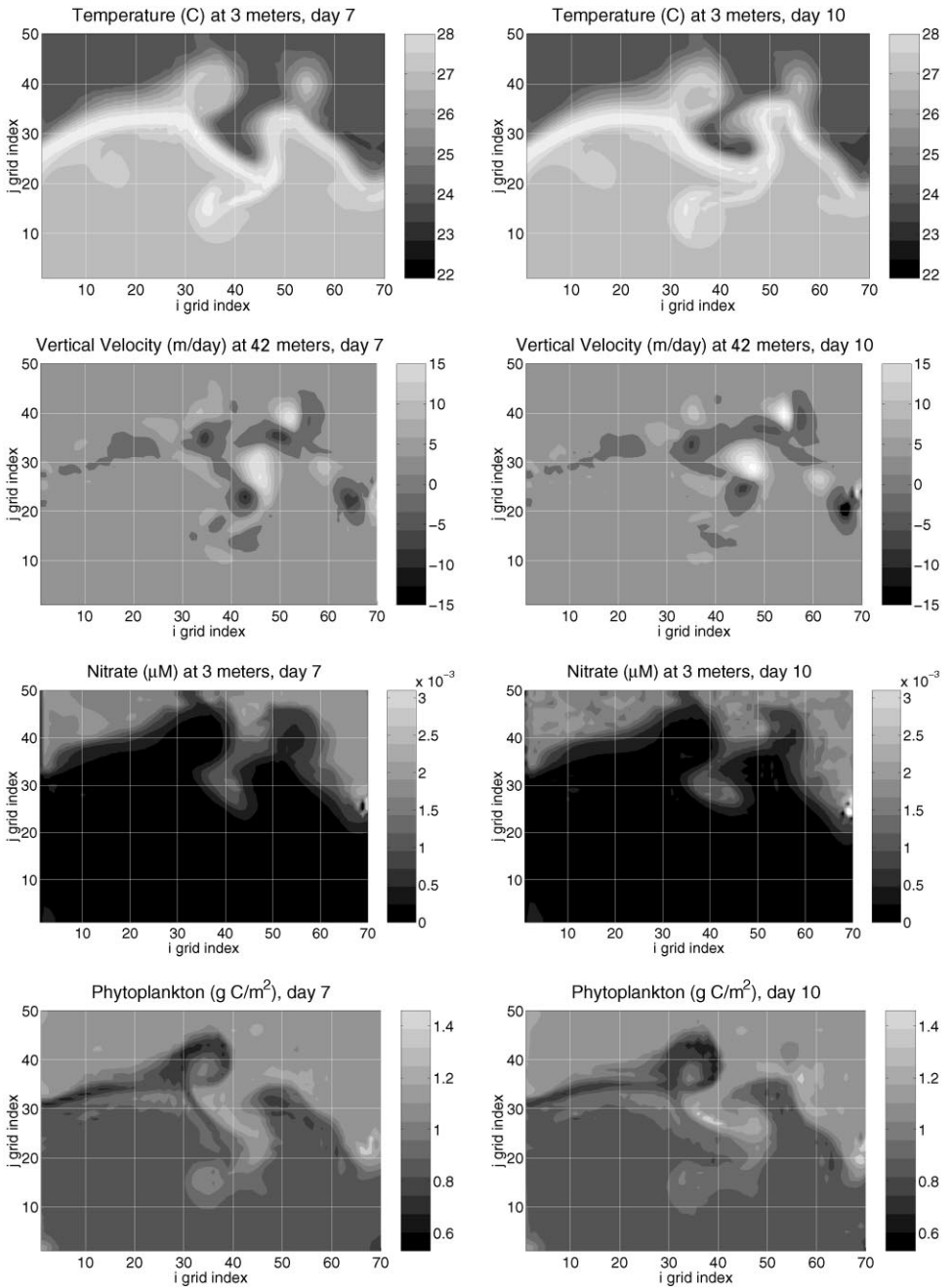
Fig. 13. Temperature, vertical velocity, nitrate and vertically integrated phytoplankton for days 7 and 10 (yeardays 279 and 282) of Run 1.
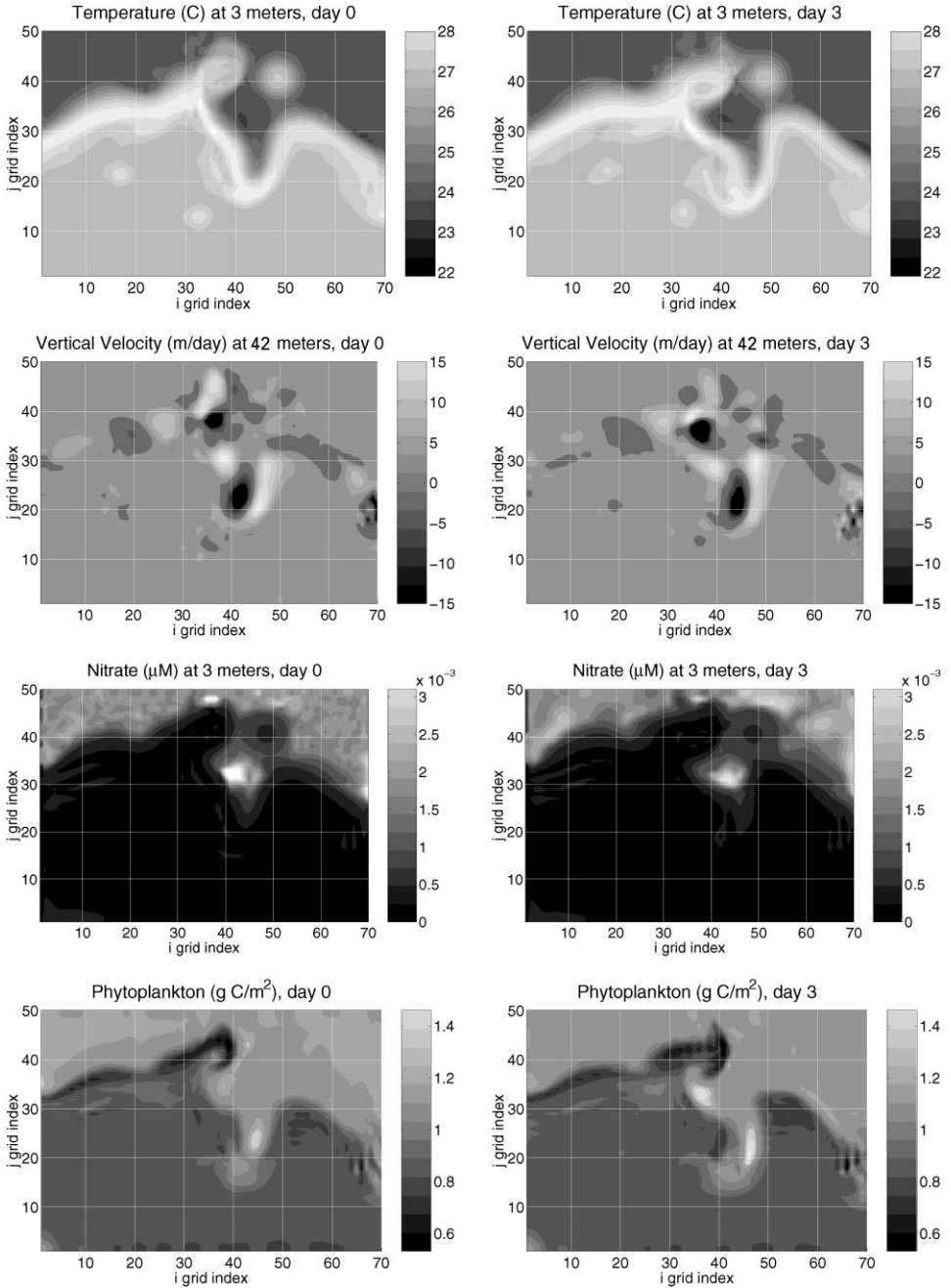
Fig. 14. Temperature, vertical velocity, nitrate and vertically integrated phytoplankton for days 0 and 3 (yeardays 279 and 282) of Run 2.

compared with days 7 and 10 of Run 1, with Run 1 being considered the "forecast" and Run 2 being the "validation" suggested by later data. Comparing the Run 1 day-7 temperature field with Run 2 day 0 shows that the Gulf Stream in Run 1 is interacting too strongly with the eastern cold core ring (at $i = 34, j = 14$), although the interaction with the western warm core ring (at $i = 35, j = 35$) is expected. Also, the central trough in Run 1 is pointing more eastward than suggested by the AVHRR data manifest in Run 2 day 0, and the eastern warm core ring (at $i = 55, j = 40$) did not drift westward as observed.

While Run 1 does a reasonably good job at predicting most of the meander evolution and the general behavior of the biological fields, in some places it does a rather poor job at predicting the location of the Gulf Stream axis, some of the jet-ring interactions, and their related biological enhancements. For this reason, simulations with no data assimilation can be deficient; they can depart from the observed state. Data assimilation is desirable to keep the simulation on track. In the next three simulations, the consequences of physical and biological data assimilation are investigated.

3.2. *Run 3*: *Physical assimilation only*

Run 3 is initialized identically with Run 1 (i.e. on day 272), but assimilates the day-279 feature model fields (Run 2 day 0) as indicated in Section 2.5, i.e. with assimilation starting on day 4 and ramping up to 100% on day 7. Only physical fields (temperature, salinity, and velocity) are assimilated.

Fig. 15 shows the results of Run 3. The temperature and vertical velocity fields of Run 3 day 7 correspond exactly with Run 2 day 0 (Fig. 14), because these fields were fully assimilated. The nitrate and phytoplankton values however are higher in certain places on day 7 in Run 3 as compared to Run 2, viz. phytoplankton in the central trough at ($i = 44, j = 23$) and nitrate near the warm core ring at ($i = 50, j = 46$). This is because the adjustment of the physical fields during data assimilation makes the physical front temporarily out of alignment with the biological front, particularly in places where the physical adjustment is strongest (i.e. where Run 1 day 7 temperature disagrees with Run 2 day 0). The net result is enhanced cross-frontal transport of biological quantities at these locations, and consequently falsely enhanced nitrate and phytoplankton concentrations.

These results suggest that the less frequent the physical data assimilation, the more out of alignment the biological field will be with the updated physical field on an assimilation timestep. This raises the question whether more frequent physical data assimilation will allow a smoother physical evolution and consequently reduce the spurious biological response. To test this, a simulation was conducted identical to Run 3 but with physical data assimilation every 0.5 days, instead of every 1.0 days. Because the assimilation frequency was doubled, the $s(t)$ values (Fig. 12) were halved (except for day 7), so that the day-7 fields would not be brought in too quickly. The results of this simulation were virtually identical to Run 3, showing no reduction of the spurious nitrate or phytoplankton concentrations. A second simulation was then conducted with assimilation every 0.25 days, with the values of $s(t)$ halved again
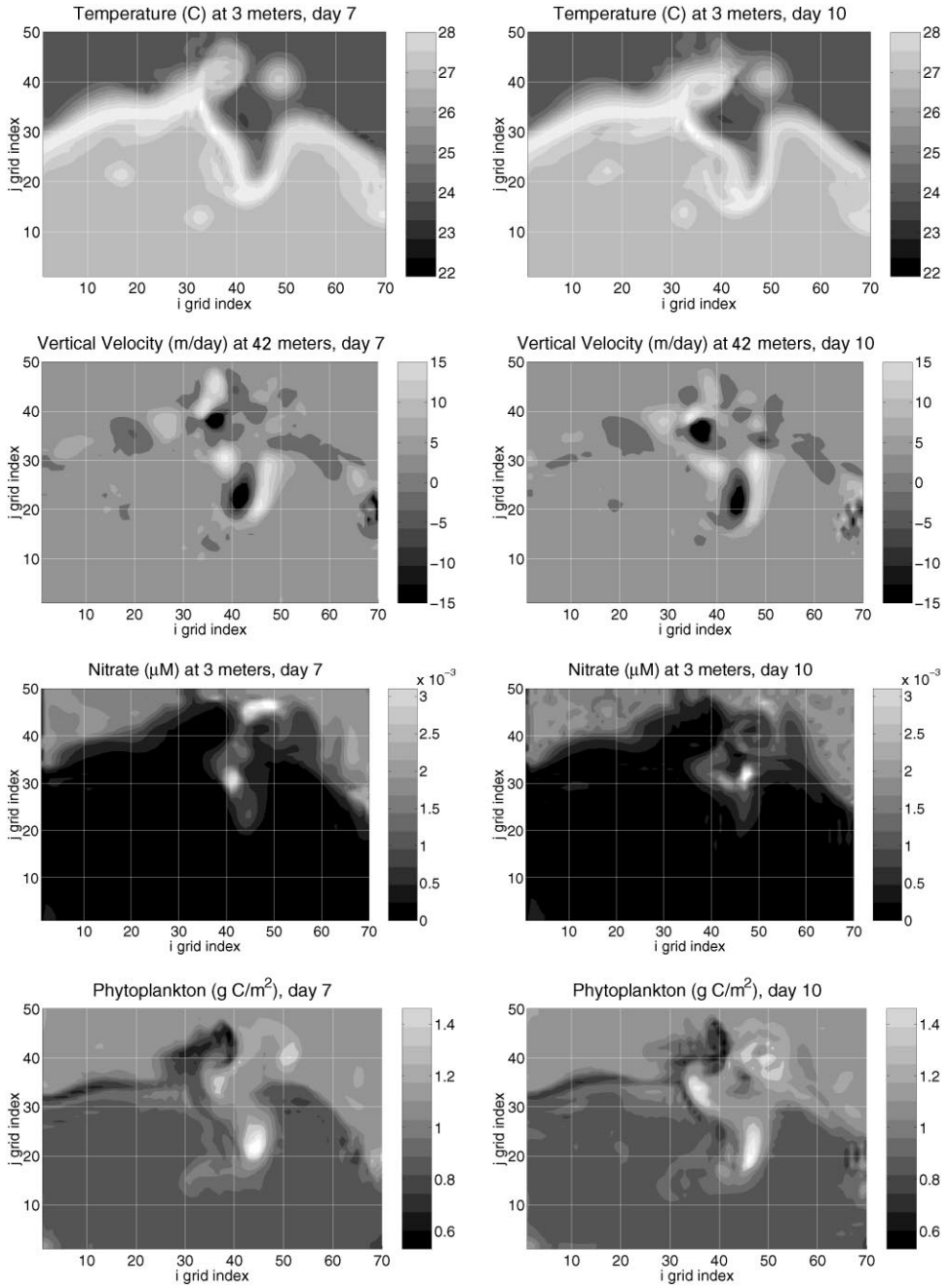
Fig. 15. Temperature, vertical velocity, nitrate and vertically integrated phytoplankton for days 7 and 10 (yeardays 279 and 282) of Run 3.

(except day 7). This simulation showed a significant reduction in the spurious NO_3 values, although the spurious phytoplankton values were only slightly lower. To adequately investigate higher assimilation frequencies (e.g. 0.1 days or less), assimilation fields would need to be created more frequently e.g. every 3 days or less, rather than every 7 days. Because of the significant additional amount of investigation required, this remains an objective of future work. It therefore remains an open question whether very high physical data assimilation frequency ($\leqslant 0.1$ days) can reduce the spurious biological response to negligible levels. Clearly an assimilation frequency of 0.5 days or longer is inadequate in this active region.

3.3. Run 4: Biological assimilation only

Run 4 is similar to Run 3 except with biological rather than physical data assimilation. All day-279 biological fields are assimilated (N, P, Z, A and D).

Fig. 16 shows the results of Run 4. The nitrate and phytoplankton fields of Run 4 day 7 agree exactly with those of Run 2 day 0 (Fig. 14) because these were assimilated. However, the temperature and vertical velocity fields of Run 4 day 7 agree with Run 1 day 7 (Fig. 13), because they evolve in the same way. The result is that in Run 4 on day 7 the physical and biological fronts are not aligned. Consequently in the post-assimilation period, i.e. on day 10 in Run 4, the nitrate and phytoplankton fields are spuriously enhanced, due to the spurious cross-frontal transport of biological quantities. Unlike Run 3, this false enhancement cannot be reduced by more frequent assimilation. Biological data assimilation without the assimilation of associated physical features clearly leads to a discrepancy in physical and biological frontal positions, with disastrous consequences.

3.4. Run 5: Physical and biological assimilation

The final simulation in this series is Run 5, which includes both physical and biological data assimilation. Days 1–4 of this run agree with Run 1, while days 7–10 agree with Run 2 (Fig. 14) as all the fields are fully assimilated. No significant spurious effects on days 4–7 of the simulation were found resulting from the assimilation procedure. Thus Run 5 is the ideal simulation, evolving smoothly from our best estimate of the day-272 fields to our best estimate of the day-279 fields.

4. Discussion

4.1. Compatible physical and biological fields

The net lesson of Runs 1–5 is that simulations with no data assimilation can easily depart from observed, while having only physical or biological data assimilation can lead to a misalignment of the physical and biological fronts, with consequently enhanced cross-frontal transport of biological quantities and spuriously enhanced biological productivity. The best method is to have both physical and
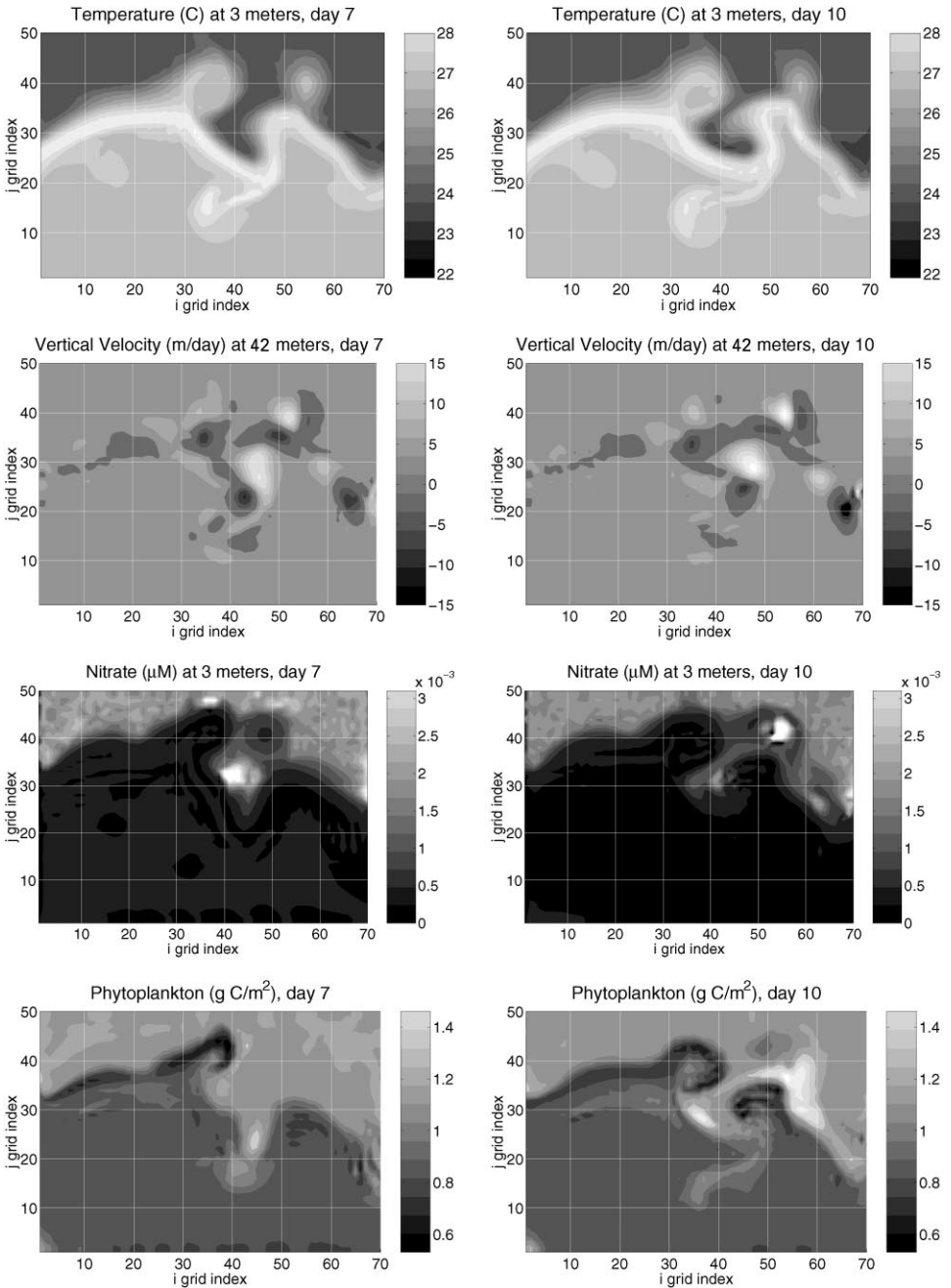
Fig. 16. Temperature, vertical velocity, nitrate and vertically integrated phytoplankton for days 7 and 10 (yeardays 279 and 282) of Run 4.

biological data assimilation. This makes the critical issue deriving from the available data physical and biological fields for assimilation that are dynamically balanced and consistent with each other. The topic of deriving such fields was addressed in Section 2; the importance of the methodology is why it was stressed at such length.

There are two additional problems with assimilating physical data only to drive a physical–biological model. Physical data assimilation that is not smooth (due to variability in the assimilated field or its error field, or disagreement between the assimilated field and the model forecast) may generate spurious vertical velocities (see Anderson and Robinson, 2000), which may cause a spurious biological response if biological data are not also assimilated (to keep the biology near observed). Secondly, good forecast skill of the biological model is unlikely due to uncertainty in parameter values, parameterizations, initial and boundary conditions, and nonlinearities in the biological equations.

When only physical data are available, it is therefore desirable to go through the procedure detailed in Fig. 1 to generate a compatible biological field for assimilation (with weaker confidence weight). This is reasonable to do because mesoscale biological features are generally correlated with physical features (Denman and Abbott, 1994). Similarly, when only biological data are available, it is desirable to create a compatible physical field to assimilate. This is reasonable because biological distributions are often due to physical processes (e.g. advection). One way would be to first generate the biological field, and then use local correlations with physical properties (from data or previous model output) to generate the physical field. Another way would be to use the biological information on frontal positions as input to the physical feature model. A crucial property is for the physical and biological fronts to agree. Yet as physical fronts are not always associated with biological fronts and vice versa (e.g. Fasham et al., 1985; Srokosz, 1997), it is left to the discretion of the modeler to judge when and where such correlations may be applied.

Note that these conclusions on the necessity of assimilating consistent physical and biological fields apply only to forward assimilation techniques (insertion, nudging, optimal interpolation). The equivalent statement in inverse modeling is to optimize not only the biological parameters, but also the uncertain physical parameters and the initial and boundary conditions.

Interestingly the assimilation studies of Najjar et al. (1992) and Anderson and Sarmiento (1995), who nudged surface ocean phosphate to annual mean concentrations, were largely successful using biological data assimilation only because the coarse-resolution models had broad property gradients and no temporal variability. They did not encounter as severe problems with mismatching physical and biological fronts as we do with the evolving Gulf Stream. This illustrates that data assimilation is easier at coarse resolution than at mesoscale resolution.

4.2. Generality of scheme

The method outlined in Fig. 1 is a general procedure with details that can be modified. For example, the objective analysis modules can be excluded (if there is no

data) and the physical and biological dynamical-adjustment periods can be shortened (for regions with weak velocities or periods of rapid biological change, respectively). Other criteria can be used to generate the initial field estimates (e.g. climatological data) and different objective analysis or data assimilation schemes used. Nevertheless, the basic structure of this procedure for producing consistent physical and biological fields from available data should remain the same for most applications and data types, including satellite-derived data. The methods described here are those we have found to be most successful for our Gulf Stream simulations.

4.3. Error field propagation

The physical error fields used for assimilation were those determined by the objective analysis technique (see Section 2.4.2). This ignores the fact that during the 7-day physical dynamical adjustment run (Section 2.4.3), the physical fields move and mix, and hence our confidence in the fields changes. Future studies will use the Error Subspace Statistical Estimation (ESSE) method (Lermusiaux, 1997; Robinson et al., 1998; Lermusiaux and Robinson, 1999), which uses a Monte Carlo method to propogate the model field errors, and then combines model fields and assimilation fields with optimal weights that minimize error in the combined fields. Here we simply use the objective analysis error fields as a first approximation, as our initial objective was primarily to test the impact of the data assimilation in the simulations.

The biological error fields used were also those determined by the objective analysis technique (Section 2.4.5). As with the physical error fields, we should actually allow the error fields to propagate during the dynamical adjustment run. Again, a good way to estimate the change in the error fields would be to use a Monte Carlo method, particularly one that takes into account biological model parameter uncertainty. Alternatively, the errors of the dynamically adjusted biological fields could be estimated through a recomparison with the data (i.e. the objective analysis fields). All of this is the focus of future work. For the time being, a simple way to include the uncertainty due to biological model error is to decrease the assimilation weights w_{max} and w_{min} for biological fields, so that they are assimilated with less confidence (i.e. weakly).

4.4. Dynamically adjusted biological fields

Ideally the dynamically adjusted biological fields should match the data (within measurement and sampling error). That is, biological model parameters and parameterizations should be adjusted until such agreement is achieved, particularly in places where the time rate of change is slow. However, it may not be possible to match the data (1) if physical structures are absent due to lack of physical data or coarse model resolution (each of which should be addressed appropriately) or (2) in places where the time rate of change is fast (e.g. as a consequence of transient physical events, such as recently outcropped isopycnals that increase surface nutrients initially but cannot sustain them throughout a 20-day dynamical adjustment run).

To include such transient data in the final dynamically adjusted fields, two dynamical adjustment runs may be necessary, first a longer one to adjust the initial feature model estimates, and then a shorter one to dynamically adjust the fields after the data have been melded in. Alternatively, the assimilation weights could be adjusted so that the dynamically adjusted field estimates that do not agree with the data are not assimilated; this would allow the model transient behavior to dominate.

In the end, it may not be possible to incorporate all the data values within their observational error bounds. As the goal of the main simulation is to reach a compromise between all the available types of data, assimilation of dynamically adjusted values is preferable to assimilating exact data values that are not consistent with other data types.

5. Summary of conclusions

Regarding forward data assimilation techniques:

In physical–biological models using physical data assimilation only, the adjustment of the physical field on an assimilation timestep can lead to temporary misalignment between the physical and biological fronts, causing spurious cross-frontal fluxes and consequently spurious biological responses (e.g. enhanced productivity). If only physical data are available, it would therefore be desirable to generate compatible biological fields and assimilate these as well (although at lower confidence).

In physical–biological models using biological data assimilation only, there is little or no feedback to the physics. Physical and biological fronts will generally become misaligned, causing spurious cross-frontal fluxes and consequently spurious biological responses (e.g. enhanced productivity). More frequent biological assimilation cannot correct this. What is needed is to generate and assimilate physical fields that can support the biological features. For example, the assimilation of satellite ocean color data should be accompanied by the assimilation of SST or altimetry data.

Having both physical and biological data assimilation is the best scenario. To do this requires generating consistent physical and biological fields for assimilation. Here we describe a generally applicable six-step method: (a) initial estimation of synoptic physical features, (b) melding physical data into these fields to obtain the best realtime estimates, (c) physical dynamical adjustment to generate vertical velocities, (d) initial estimation of mesoscale biological fields based on physical–biological correlations, (e) melding biological data into these fields, and (f) biological dynamical adjustment with frozen physical fields to balance the biological fields with each other, the model parameters, and the 3-D physical transports. The generation of these fields is done in "adjustment space", outside of the simulation of interest ("simulation space").

This study demonstrates successful assimilation of synoptic physical and biological data to reconstruct an observed mesoscale physical–biological event. Assimilation into a model dynamically adjusts the data, allowing a more powerful process study

than analysis of the data alone, or simulation alone. This method is generally applicable for realtime mesoscale physical–biological forecasting (e.g. Srokosz, 1997; Lozano et al., 1998).

Acknowledgements

Appendix A

A.1. The physical model

The Harvard Ocean Prediction System (HOPS) primitive equation model (Lozano et al., 1996; Robinson, 1996) was used with a 75 by 51 horizontal grid of 15 kilometer resolution. Hybrid vertical coordinates were used (Lozano et al., 1994; Haley, 1997); 16 flat levels extended to a depth of 336 m, with 14 terrain-following levels beneath. Level thicknesses increased with depth from 6 m at the sea surface to 796 m for the thickest bottom level. Realistic bathymetry was used, with the exception that depths less than 1300 m were deepened to 1300 m and depths greater than 4600 m were shallowed to 4600 m. The upper limit was imposed so that the terrain-following levels would not go to extremely small thicknesses upon the shelf; the bathymetry was shallowed to 4600 m for consistency with previous simulations (e.g. Gangopadhyay et al., 1997). A timestep of 15 min was used.

The dynamics are governed by the primitive equations under the hydrostatic, Boussinesq and rigid-lid approximations. Subgridscale horizontal mixing of tracers (temperature, salinity and biological quantities) and momentum was parameterized using Laplacian diffusion with a coefficient of 250 m^2/s, appropriate for cross-Gulf Stream diffusivity at this scale (Bower et al., 1985). Horizontal mixing of vorticity was parameterized by applying a fourth-order Shapiro (1970) filter once every timestep, which removes any unresolved gravity waves and the small scales created by the geostrophic turbulence enstrophy cascade (Rhines, 1979; Robinson and Walstad, 1987). Vertical mixing of momentum and tracers was parameterized by the Richardson-number dependent scheme of Pacanowski and Philander (1981). Convective mixing was handled by using a local vertical diffusivity of 50 cm^2/s in instances of local gravitational instability.

Persisted boundary conditions were used for momentum, vorticity and streamfunction. Persisted boundary conditions were also used for temperature and salinity, except on the outflow (eastern) boundary, where values were set equal to that of the first interior point each timestep.

Regarding surface forcing, in Runs 1–6 and 8 no surface heat flux or wind stress was applied, because we wanted to isolate changes associated solely with Gulf Stream meandering and the data assimilation scheme. In Run 7, 12-hourly ECMWF winds of $2.5°$ spatial resolution were applied for the appropriate dates of the simulation (21 Sept–6 Oct 1988). An idealized shortwave radiation flux was applied, which was in the shape of a half-sinusoid with a noontime maximum of $707\ W/m^2$. In addition to this a constant cooling of $-225\ W/m^2$ was applied, so that the daily net heat flux integrated to zero, with heating during the day and cooling at night. No surface water (salt) flux was applied.

In simulations without surface forcing (Runs 1–6 and 8), a constant mixed layer depth of 30 m, typical of late summer conditions, was prescribed over the entire domain. This was done by specifying a vertical diffusivity of $50\ cm^2/s$ between 0 and 30 m. In simulations with surface forcing (Run 7), the Large et al. (1994) mixed-layer scheme was first tried, but found to be computationally expensive. Instead, a simple mixed layer model was used, described as follows, which is based on the heart of the Large et al. (1994) scheme. First, at each location the mixed layer depth was estimated as that depth at which the change in density with depth became greater than $0.003\ kg/m^4$. Second, a "target" mixed layer depth was estimated at each location as

$$Z_{\text{target}} = \begin{cases} Z_{\text{Ekman}} & \text{if } B \leqslant 0, \\ \min(Z_{\text{Ekman}}, Z_{\text{Monin-Obukhov}}) & \text{if } B > 0, \end{cases}$$

where

$$Z_{\text{Ekman}} = ku_*/f,$$

$$Z_{\text{Monin-Obukhov}} = 2mu_*^3/B,$$

$$u_* = \sqrt{\tau/\rho_0},$$

where B is the surface buoyancy flux (positive for heating), τ is the local wind stress, ρ_0 is the density of seawater, f is the Coriolis parameter, k is a proportionality constant of 0.3 (tuned empirically), and m is a proportionality constant of 1.25 (Niiler and Kraus, 1977; Large et al., 1994). If Z_{target} is less than the mixed layer depth, a vertical diffusivity of $50\ cm^2/s$ was applied between 0 and Z_{target} meters. If Z_{target} is greater than the depth of the next level below the mixed layer depth, the vertical diffusivity of $50\ cm^2/s$ was applied to the depth of that next level in order to entrain it into the mixed layer. In this way, the mixed layer depth was primarily set as Z_{target}, with a retarding effect on mixed layer deepening depending upon the local stratification.

This simple mixed layer scheme, in combination with our vertical mixing scheme and convection scheme, has shown to give good results under a variety of forcing conditions.

A.2. The biological model

The biological model (Fig. 3, Tables 1–3) is of the nitrogen cycle. In the equations, greek and lower-case roman letters denote constants (parameters). Upper-case roman letters denote fields that vary in space and time (variables). The nitrogen cycle is closed, and the model is applied to the entire water column. It is important that remineralization processes be included even in the euphotic zone, because the majority of daily biological production is recycled there.

This biological model is an expansion of the four-component model of McGillicuddy et al. (1995). As such, several equations in Table 3 are the same (Eqs. (1)–(3) and (7) (8)). The primary changes were to modify the phytoplankton growth equations so that productivity in the summertime Sargasso could be well simulated (Eqs. (4)–(6)), and to explicitly include non-living organic nitrogen (Eqs. (9)–(11)).

Let us first discuss the phytoplankton growth equations. The nitrate limitation factor (Eq. (4)) was modified from its previous definition (Fasham et al., 1990) to ensure that $Q_1 + Q_2 \leq 1$, at the expense of nitrate uptake if ammonium is plentiful. A depth-dependent carbon-to-chlorophyll ratio was introduced (Eq. (5); Malone et al., 1993) as it is important for the productivity of the Sargasso deep chlorophyll maximum (Doney et al., 1996). It was found that in the Sargasso at 40–60 m depth, which in summer is both light- and nutrient-limited, multiplying the factors $(Q_1 + Q_2)L_0$ gives a primary production estimate much lower than observed; we therefore use a $\min(Q_1 + Q_2, L_0)$ formulation (Hurtt and Armstrong, 1996; McClain et al., 1996) which accurately represents Blackman's law of the minimum. The proper division of this into nitrate and ammonium uptake is then $\min(Q_1 + Q_2, L_0)Q_1/(Q_1 + Q_2)$ and $\min(Q_1 + Q_2, L_0)Q_2/(Q_1 + Q_2)$ respectively, which is reflected in Eq. (6) and Fig. 3. Eq. (6) also includes the maximum specific growth rate (g C (g Chl h)$^{-1}$), the carbon-to-chlorophyll ratio, and a temperature dependence factor (Sarmiento et al., 1993; normalized to 22°C) included on account of the strong temperature gradient across the Gulf Stream.

Let us now discuss the modeling of PON and DON. Preliminary simulations were carried out using a single detrital component, but this was found to be inadequate. Detrital sinking rates during a bloom can be up to 1000 m day^{-1} (Fowler and Knauer, 1986), while in oligotrophic regions detrital sinking rates are much lower and the dominant role may be played by DOM (Carlson et al., 1994). Previous studies (Fasham et al., 1993; Prunet et al., 1996a, b) have shown that a model's detrital sinking rate is of enormous importance, directly influencing recycled production rates. We therefore use two classes of non-living organic nitrogen, F which sinks fast (100 + meters per day) and which plays an important role in transporting bloom material, and D for which the sinking rate is essentially negligible relative to vertical mixing and biological rates, to represent the two ends of the spectrum. D includes semi-labile DON and non-living suspended PON, which are lumped together because (a) their function in the ecosystem is essentially the same (they are passively advected and diffused, and ultimately remineralize to nutrients), (b) there is no BIOSYNOP PON or DON data to validate these variables, and (c) there is significant uncertainty in what values to use for suspended PON and labile DON concentrations and re-

mineralization timescales (see Section 2.4.4). Consequently, modeling suspended PON and DON seperately would create mathematical redundancy in the equations, making the model difficult to tune, while combining them results in a better-constrained set of equations.

It was found critical to have the fraction of mortality and excretion going into F and D to vary with production. In oligotrophic regions approximately 5% of daily primary production goes into fast sinking particles while in high productivity regions 50% may be more typical (Eppley and Peterson, 1979). We parameterize this fractionation through the function $E(P)$ (Eq. (9)), where we use dependency on phytoplankton concentration rather than production rate to avoid values of zero at night. In retrospect, it may have been better to use a maximum value of 25% (Wefer, 1993) rather than 50%; see Anderson and Robinson (2000).

The next issue is how to model the remineralization of F and D. We use first-order kinetics for remineralization of D to A, and nitrification of A to N. However, a practical difficulty was found with modeling the F pool explicitly. As discussed in Section 2.4.6, we desire the biological components to reach near-equilibrium concentrations within a short spin-up period. Using F with a sinking rate of 100 m per day would require over 46 days to stabilize (because the water column is 4600 m deep), which is undesirably long. Instead F is modeled as sinking and remineralizing instantaneously (e.g. Anderson and Sarmiento, 1995), representing the extreme limit of the sinking rate spectrum.

The remineralization profile (R) of F is determined as follows. F is produced on each model level (as EM), and it is assumed that the sinking F flux decreases with depth with an e-folding scale. Observations (Martin et al., 1987) suggest that the e-folding scale is approximately 100 m in the upper ocean and 1000 m in the deep ocean, i.e. essentially proportional to depth itself. This suggested the form for Eq. (10), where the adjustment factor of 1.15 allows remarkable agreement with the Martin et al. (1987) flux profile. This equation has a distinct advantage over that of Martin et al. (1987) because it can be applied to material produced at an arbitrary depth, i.e. it does not depend upon the flux at 100 m. The remineralization of F sinking out of a given model box into the levels below is then determined as the vertical divergence of this flux (Eq. (11)). F reaching the seafloor is remineralized instantaneously in the bottom box.

Simulations were conducted that also included zooplankton grazing of D (i.e. suspended PON). This was found to be of no qualitative and little quantitative importance because the distributions of Z and D essentially mimic phytoplankton; the net effect of including this process was to modify the net remineralization rate of D. Therefore, we have simplified the model to control the D breakdown rates explicitly. Bacteria have been neglected, primarily because we have no data for assimilation. As with zooplankton, bacterial processes of remineralization are parameterized directly.

Recently, it has been suggested that nitrogen fixation may play an important role in the Sargasso Sea in summertime (Michaels et al., 1996; Gruber and Sarmiento, 1997). Estimates are still uncertain, ranging four-fold 0.20–0.88 mmol N m^{-2} d^{-1}. However, the proposed nitrogen fixation is not thought to occur in the ^{14}C bottle incubations that we use for model validation, but due to vertically migrating *Trichodesmium*. Furthermore, the nitrogen fixation is only a minor fraction of primary

production (2.5–3.8 mmol N m^{-2} d^{-1}; Michaels et al., 1994b); our study sets out to explain how this production is maintained despite negligible NO_3. In this study N_2 fixation is therefore ignored. As the various pools of nitrogen in the mixed layer are not changing rapidly during summer, any new nitrogen brought in by N_2 fixation must be balanced by an equivalent export. As such any production in the Sargasso due to nitrogen fixation and the balancing export is in addition to that estimated in this study.

Some biological model parameter values (Table 2) were taken from the literature, particularly those determined for BIOSYNOP by Lohrenz et al. (1993); others were tuned in sensitivity studies as described in Section 2.4.4. The low k_1 and k_2 values are supported by experiment (Harrison et al., 1996) and are necessary to obtain the observed growth rates given the low mixed-layer NO_3 and NH_4 concentrations ($\leqslant 0.01$ μM).

A PAR flux of 45% of the shortwave radiation flux was used to drive biological production. It was found important to include the daily cycle of PAR because of its complex impact on production rates and the depth of the chlorophyll maximum. Unfortunately available PAR data was sparse. In sensitivity tests (see Section 2.4.4), it was determined that the noontime PAR flux must be in the vicinity of 318 W/m^2 (which is in good agreement with the data) for the subsurface Chl maxima in the Slope and Sargasso waters to be at their observed depths (higher light intensity allowing deeper Chl maxima).

Orlanski-type boundary conditions (Stevens, 1990; Lermusiaux, 1997; Nicolas et al., 1997) were used for biological tracers along the inflow (western) boundary. On all other boundaries, biological tracer values were set equal to that of the first interior point. Spurious vertical velocities were often generated at the boundaries in places of high horizontal velocity (shear). To reduce the impact of these, the vertical advection term for biological tracers was set to zero along the boundaries, except for the outflow (eastern) boundary, where it was set equal to that of the first interior point. The boundary regions are not regarded as a believable part of our simulations; they are primarily buffer zones to protect the interior solution.

References

Anderson, D.L.T., Willebrand, J., 1989. Oceanic Circulation Models: Combining Data and Dynamics. In: NATO ASI Series C, 284. Kluwer Academic, Dordrecht.

Anderson, L.A., Robinson, A.R., 2000. Physical and biological modeling in the Gulf Stream region: II. Physical and biological processes. Deep-Sea Research I.

Anderson, L.A., Sarmiento, J.L., 1995. Global Ocean phosphate and oxygen simulations. Global Biogeochemical Cycles 9, 621–636.

Archer, D., Peltzer, E.T., Kirchman, D.L., 1997. A timescale for dissolved organic production in equatorial Pacific surface waters. Global Biogeochemical Cycles 11, 435–452.

Armstrong, R.A., Sarmiento, J.L., Slater, R.D., 1995. Monitoring ocean productivity by assimilating satellite chlorophyll into ecosystem models. In: Powell, T.M., Steele, J.H. (Eds.), Ecological Time Series. Chapman and Hall, London, pp. 371–390.

Ashjian, C.J., 1993. Trends in copepod species abundances across and along a Gulf Stream meander: evidence for entrainment and detrainment of fluid parcels from the Gulf Stream. Deep-Sea Research I 40, 461–482.

Ashjian, C.J., Smith, S.L., Flagg, C.N., Mariano, A.J., Behrens, W.J., Lane, P.V.Z., 1994. The influence of a Gulf Stream meander on the distribution of zooplankton biomass in the Slope Water, the Gulf Stream, and the Sargasso Sea, described using a shipboard acoustic Doppler current profiler. Deep-Sea Research I 41, 23–50.

Bengtsson, L., Ghil, M., Kallen, E. (Eds.), 1981. Dynamic Meteorology: Data Assimilation Methods. Springer, Berlin.

Bowen, A.J., Griffin, D.A., Hazen, D.G., Matheson, S.A., Thompson, K.R., 1995. Shipboard nowcasting of shelf circulation. Continental Shelf Research 15, 115–128.

Bower, A.S., Rossby, T., 1989. Evidence of cross-frontal exchange processes in the Gulf Stream based on isopycnal RAFOS float data. Journal of Physical Oceanography 19, 1177–1190.

Bower, A.S., Rossby, H.T., Lillibridge, J.L., 1985. The Gulf Stream — barrier or blender? Journal of Physical Oceanography 15, 24–32.

Brasseur, P.P. (Ed.), 1995. Data assimilation in marine science. Journal of Marine Systems 6(1–2), 175 pp.

Brasseur, P.P., Nihoul, J.C.J. (Eds.), 1994. Data Assimilation: Tools for Modelling the Ocean in a Global Change Perspective. NATO ASI Series, Vol. I 19. Springer, Berlin.

Brzezinski, M.A., 1988. Vertical distribution of ammonium in stratified oligotrophic waters. Limnology and Oceanography 33, 1176–1182.

Campbell, J.W., 1995. The lognormal distribution as a model for bio-optical variability in the sea. Journal of Geophysical Research 100, 13237–13254.

Carlson, C.A., Ducklow, H.W., Michaels, A.F., 1994. Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso Sea. Nature 371, 405–408.

Carpenter, S.R., Cottingham, K.L., Stow, C.A., 1994. Fitting predator-prey models to time series with observation errors. Ecology 75, 1254–1264.

Carter, E.F., Robinson, A.R., 1987. Analysis models for the estimation of oceanic fields. Journal of Atmospheric and Oceanic Technology 4, 49–74.

Denman, K.L., Abbott, M.R., 1994. Time scales of pattern evolution from cross-spectrum analysis of advanced very high resolution radiometer and coastal zone color scanner imagery. Journal of Geophysical Research 99, 7433–7442.

Denman, K.L., Freeland, H.J., 1985. Correlation scales, objective mapping and a statistical test of geostrophy over the continental shelf. Journal of Marine Research 43, 517–539.

Doney, S.C., Glover, D.M., Najjar, R.G., 1996. A new coupled, one-dimensional biological-physical model for the upper ocean: applications to the JGOFS Bermuda Atlantic Time-series Study (BATS) site. Deep-Sea Research II 43, 591–624.

Eppley, R.W., Peterson, B.J., 1979. Particulate organic matter flux and planktonic new production in the deep ocean. Nature 282, 677–680.

Evans, G.T., 1999. The role of local models and data sets in the Joint Global Ocean Flux Study. Deep-Sea Research I 46, 1369–1389.

Evensen, G., 1994. Inverse methods and data assimilation in nonlinear ocean models. Physica D 77, 108–129.

Fasham, M.J.R., Evans, G.T., 1995. The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47°N 20°W. Philosophical Transactions of the Royal Society of London B 348, 203–209.

Fasham, M.J.R., Platt, T., Irwin, B., Jones, K., 1985. Factors affecting the spatial pattern of the deep chlorophyll maximum in the region of the Azores front. Progress in Oceanography 14, 129–165.

Fasham, M.J.R., Ducklow, H.W., McKelvie, S.M., 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. Journal of Marine Research 48, 591–639.

Fasham, M.J.R., Sarmiento, J.L., Slater, R.D., Ducklow, H.W., Williams, R., 1993. Ecosystem behavior at Bermuda Station "S" and Ocean Weather Station "India": a general circulation model and observational analysis. Global Biogeochemical Cycles 7, 379–415.

Fowler, S.W., Knauer, G.A., 1986. Role of large particles in the transport of elements and organic compounds through the oceanic water column. Progress in Oceanography 16, 147–194.

Fox, D.N., Carnes, M.R., Mitchell, J.L., 1992. Characterizing major frontal systems: a nowcast/forecast system for the northwest Atlantic. Oceanography 5 (1), 49–54.

Gangopadhyay, A., Robinson, A.R., 1997. Circulation and dynamics of the western North Atlantic. Part III: Forecasting the meanders and rings. Journal of Atmospheric and Oceanic Technology 14, 1352–1365.

Gangopadhyay, A., Robinson, A.R., Arango, H.G., 1997. Circulation and dynamics of the western North Atlantic. Part I: multiscale feature models. Journal of Atmospheric and Oceanic Technology 14, 1314–1332.

Ghil, M., 1989. Meteorological data assimilation for oceanographers, I: description and theoretical framework. Dynamics of Atmospheres and Oceans 13, 171–218.

Ghil, M., Malanotte-Rizzoli, P., 1991. Data assimilation in meteorology and oceanography. Advances in Geophysics 33, 141–266.

Glenn, S.M., Robinson, A.R., 1995. Verification of an operational Gulf stream forecasting model. In: Lynch, D.R., Davies, A.M. (Eds.), Qantitative Skill Assessment for Coastal Ocean Models. Coastal and Estuarial Studies, 47. American Geophysical Union, Washington, DC, pp. 469–499.

Gruber, N., Sarmiento, J.L., 1997. Global patterns of marine nitrogen fixation and denitrification. Global Biogeochemical Cycles 11, 235–266.

Haidvogel, D.B., Robinson, A.R. (Eds.), 1989. Special Issue on Data Assimilation. Dynamics of Atmospheres and Oceans 13, 171–517.

Haley, P.J., 1997. GRIDS User's Guide, Harvard Open Ocean Model Report No. 54. Harvard University, Cambridge, MA.

Halkin, D., Rossby, H.T., 1985. The structure and transport of the Gulf Stream at 73 deg W. Journal of Physical Oceanography 15, 1439–1452.

Hansell, D.A., Carlson, C.A., 2000. Biogeochemistry of total organic carbon and nitrogen in the Sargasso Sea. Deep-Sea Research II.

Harmon, R., Challenor, P., 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. Ecological Modelling 101, 41–59.

Harrison, W.G., Harris, L.R., Irwin, B.D., 1996. The kinetics of nitrogen utilization in the oceanic mixed layer: nitrate and ammonium interactions at nanomolar concentrations. Limnology and Oceanography 41, 16–32.

Hitchcock, G.L., Mariano, A.J., Rossby, T., 1993. Mesoscale pigment fields in the Gulf Stream: observations in a meander crest and trough. Journal of Geophysical Research 98, 8425–8445.

Hurtt, G.C., Armstrong, R.A., 1996. A pelagic ecosystem model calibrated with BATS data. Deep-Sea Research II 43, 653–683.

Hurtt, G.C., Armstrong, R.A., 1999. A pelagic ecosystem model calibrated with BATS and OWSI data. Deep-Sea Research I 46, 27–61.

Ishizaka, J., 1990. Coupling of coastal zone color scanner data to a physical-biological model of the southeastern U.S. continental shelf ecosystem 3. Nutrient and phytoplankton fluxes and CZCS data assimilation. Journal of Geophysical Research 95, 20201–20212.

Ishizaka, J., 1993. Data assimilation for biogeochemical models. In: Evans, G.T., Fasham, M.J.R. (Eds.), Towards a Model of Ocean Biogeochemical Processes. NATO ASI Series I, 10. Springer, Berlin, pp. 295–316.

Jackson, G.A., Eldridge, P.M., 1992. Food web analysis of a planktonic system off southern California. Progress in Oceanography 30, 223–251.

Johannessen, J.A., Røed, L.P., Johannessen, O.M., Evensen, G., Hackett, B., Pettersson, L.H., Haugan, P.M., Sandven, S., Shuchman, R., 1993. Monitoring and modeling of the marine coastal environment. Photogrammetric Engineering and Remote Sensing 59, 351–361.

Joyce, T.M., Bishop, J.K.B., Brown, O.B., 1992. Observations of offshore shelf-water transport induced by a warm-core ring. Deep-Sea Research 39 (Suppl. 1), S97–S113.

Large, W.G., McWilliams, J.C., Doney, S.C., 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Reviews of Geophysics 32, 363–403.

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

Lermusiaux, P.F.J., 1997. Error subspace data assimilation methods for ocean field estimation: theory, validation and applications. Harvard Open Ocean Model Report No. 55, Ph.D. Thesis, Harvard University.

Lermusiaux, P.F.J., Robinson, A.R., 1999. Data assimilation via error subspace statistical estimation. Part I: theory and schemes. Monthly Weather Review 127, 1385–1407.

Levitus, S., 1982. Climatological Atlas of the World Ocean. NOAA Professional Paper 13. U.S. Government Printing Office, Washington, DC, pp. 173.

Lindstrom, S.S., Watts, D.R., 1994. Vertical motion in the Gulf Stream near 68 deg W. Journal of Physical Oceanography 24, 2321–2333.

Lohrenz, S.E., Cullen, J.J., Phinney, D.A., Olson, D.B., Yentsch, C.S., 1993. Distribution of pigments and primary production in a Gulf Stream meander. Journal of Geophysical Research 98, 14545–14560.

Lorenc, A.C., Bell, R.S., Macpherson, B., 1991. The Meteorological Office analysis correction data assimilation scheme. Quarterly Journal of the Royal Meteorological Society 117, 59–89.

Lozano, C.J., Haley, P.J., Arango, H.G., Sloan, N.Q., Robinson, A.R., 1994. Harvard coastal/deep water primitive equation model. Harvard Open Ocean Model Report No. 52, Harvard University, Cambridge, MA.

Lozano, C.J., Robinson, A.R., Arango, H.G., Gangopadhyay, A., Sloan, N.Q., Haley, P.J., Leslie, W.G., 1996. An interdisciplinary ocean prediction system: assimilation strategies and structured data models. In: Malanotte-Rizzoli, P. (Ed.), Modern Approaches to Data Assimilation in Ocean Modeling. Elsevier Oceanography Series. Elsevier Science, Amsterdam, pp. 413–452.

Lozano, C.J., Dusenberry, J.A., Haley, P.J., Anderson, L.A., Leslie, W.G., 1998. An interdisciplinary modeling study of the 1997 spring bloom in the eastern North Atlantic: P^2S^3 experiment. EOS, 1998 Ocean Sciences Meeting, American Geophysical Union, p. OS180.

Malanotte-Rizzoli, P. (Ed.), 1996. Modern Approaches to Data Assimilation in Ocean Modeling. Elsevier Oceanography Series. Elsevier Science, Amsterdam.

Malone, T.C., Pike, S.E., Conley, D.J., 1993. Transient variations in phytoplankton productivity at the JGOFS time series station. Deep-Sea Research I 40, 903–924.

Mariano, A.J., Hitchcock, G.L., Ashjian, C.J., Olson, D.B., Rossby, T., Ryan, E., Smith, S.L., 1996. Principal component analysis of biological and physical variability in a Gulf Stream meander crest. Deep-Sea Research I 43, 1531–1565.

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.

Matear, R.J., Holloway, G., 1995. Modeling the inorganic phosphorus cycle of the North Pacific using an adjoint data assimilation model to assess the role of dissolved organic phosphorus. Global Biogeochemical Cycles 9, 101–119.

Martin, J.H., Knauer, G.A., Karl, D.M., Broenkow, W.W., 1987. VERTEX: carbon cycling in the northeast Pacific. Deep-Sea Research 34, 267–285.

McClain, C.R., Arrigo, K., Tai, K.-S., Turk, D., 1996. Observations and simulations of physical and biological processes at ocean weather station P, 1951–1980. Journal of Geophysical Research 101, 3697–3713.

McGillicuddy, D.J., McCarthy, J.J., Robinson, A.R., 1995. Coupled physical and biological modeling of the spring bloom in the North Atlantic (I): model formulation and one dimensional bloom processes. Deep-Sea Research I 42, 1313–1357.

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 of *Pseudocalanus* spp. in the Gulf of Maine-Georges Bank region. Fisheries Oceanography 7, 205–218.

Mellor, G.L., Ezer, T., 1991. A Gulf Stream model and an altimetry assimilation scheme. Journal of Geophysical Research 96, 8779–8795.

Michaels, A.F., Knap, A.H., 1996. Overview of the U.S. JGOFS Bermuda Atlantic Time-series Study and the Hydrostation S program. Deep-Sea Research II 43, 157–198.

Michaels, A.F., Bates, N.R., Buesseler, K.O., Carlson, C.A., Knap, A.H., 1994a. Carbon-cycle imbalances in the Sargasso Sea. Nature 372, 537–540.

Michaels, A.F., Knap, A.H., Dow, R.L., Gundersen, K., Johnson, R.J., Sorensen, J., Close, A., Knauer, G.A., Lohrenz, S.E., Asper, V.A., Tuel, M., Bidigare, R., 1994b. Seasonal patterns of ocean biogeochemistry at the U.S. JGOFS Bermuda Atlantic time-series study site. Deep-Sea Research I 41, 1013–1038.

Michaels, A.F., Olson, D., Sarmiento, J.L., Ammerman, J.W., Fanning, K., Jahnke, R., Knap, A.H., Lipschultz, F., Prospero, J.M., 1996. Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. Biogeochemistry 35, 181–226.

Moisan, J.R., Hofmann, E.E., Haidvogel, D.B., 1996. Modeling nutrient and plankton processes in the California coastal transition zone 2. A three-dimensional physical-bio-optical model. Journal of Geophysical Research 101, 22677–22691.

Najjar, R.G., Sarmiento, J.L., Toggweiler, J.R., 1992. Downward transport and fate of organic matter in the ocean: simulations with a general circulation model. Global Biogeochemical Cycles 6, 45–76.

Nicolas, X., Traore, P., Mojtabi, A., Caltagirone, J.P., 1997. Augmented Lagrangian method and open boundary conditions in 2D simulation of Poiseuille–Bérnard channel flow. International Journal for Numerical Methods in Fluids 25, 265–283.

Nihoul, J.C.J., Adam, P., Brasseur, P., 1994. Mathematical visualisation of the northern Bering Sea's summer ecohydrodynamics. In: Brasseur, P.P., Nihoul, J.C.J. (Eds.), Data Assimilation: Tools for Modelling the Ocean in a Global Change Perspective. NATO ASI Series I, Vol. 19. Springer, Berlin, pp. 107–134.

Niiler, P.P., Kraus, E.B., 1977. One-dimensional models of the upper ocean. In: Kraus, E.B. (Ed.), Modelling and Prediction of the Upper Layers of the Ocean. Pergamon, Oxford, pp. 143–172.

Niquil, N., Jackson, G.A., Legendre, L., Delesalle, B., 1998. Inverse model analysis of the planktonic food web of Takapoto Atoll (French Polynesia). Marine Ecology Progress Series 165, 17–29.

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.

Phillips, N.A., 1982. On the completeness of multi-variate optimum interpolation for large-scale meteorological analysis. Monthly Weather Review 110, 1329–1334.

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

Prunet, P., Minster, J.-F., Echevin, V., Dadou, I., 1996b. Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean 2. Adjusting a simple trophic model to chlorophyll, temperature, nitrate, and pCO_2 data. Global Biogeochemical Cycles 10, 139–158.

Redfield, A.C., Ketchum, B.H., Richards, F.A., 1963. The influence of organisms on the composition of sea-water. In: Hill, M.N. (Ed.), The Sea, Vol. 2. Interscience, New York, pp. 26–77.

Rhines, P.B., 1979. Geostrophic turbulence. Annual Review of Fluid Mechanics 11, 401–441.

Robinson, A.R., 1992. Shipboard prediction with a regional forecast model. Oceanography 5 (1), 42–48.

Robinson, A.R., 1996. Physical processes, field estimation and an approach to interdisciplinary ocean modeling. Earth-Science Reviews 40, 3–54.

Robinson, A.R., Gangopadhyay, A., 1997. Circulation and dynamics of the western North Atlantic. Part II: dynamics of meanders and rings. Journal of Atmospheric and Oceanic Technology 14, 1333–1351.

Robinson, A.R., Leslie, W.G., 1985. Estimation and prediction of oceanic eddy fields. Progress in Oceanography 14, 485–510.

Robinson, A.R., Walstad, L.J., 1987. The Harvard open ocean model: calibration and application to dynamical processes, forecasting and data assimilation studies. Journal of Applied Numerical Mathematics 3, 89–131.

Robinson, A.R., Spall, M.A., Pinardi, N., 1988. Gulf Stream simulations and the dynamics of ring and meander processes. Journal of Physical Oceanography 18, 1811–1853.

Robinson, A.R., Glenn, S.M., Spall, M.A., Walstad, L.J., Gardner, G.M., Leslie, W.G., 1989a. Forecasting Gulf Stream meanders and rings. EOS 70 (45), 1464–1473.

Robinson, A.R., Spall, M.A., Walstad, L.J., Leslie, W.G., 1989b. Data assimilation and dynamical interpolation in GULFCAST experiments. Dynamics of Atmospheres and Oceans 13, 301–316.

Robinson, A.R., Lermusiaux, P.F.J., Sloan, N.Q., 1998. Data Assimilation. In: Brink, K.H., Robinson, A.R. (Eds.), The Sea. Vol. 10, The Global Coastal Ocean, Processes and Methods, Wiley, New York, pp. 541–594.

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

Semovski, S.V., Woźniak, B., 1995. Model of the annual phytoplankton cycle in the marine ecosystem-assimilation of monthly satellite chlorophyll data for the North Atlantic and Baltic. Oceanologia 37, 3–31.

Shapiro, R., 1970. Smoothing, filtering and boundary effects. Review of Geophysics and Space Physics 8, 359–387.

Spall, M.A., Robinson, A.R., 1990. Regional primitive equation studies of the Gulf Stream meander and ring formation region. Journal of Physical Oceanography 20, 985–1016.

Spitz, Y.H., Moisan, J.R., Abbott, M.R., Richman, J.G., 1998. Data assimilation and a pelagic ecosystem model: parameterization using time series observations. Journal of Marine Systems 16, 51–68.

Srokosz, M.A., 1997. Plankton patchiness studies by ship and satellite (P^2S^3), RRS Discovery Cruise 227, 15 April–16 May 1997. Cruise Report No. 12, National Oceanographic Library, Southampton Oceanography Centre.

Steele, J.H., Henderson, E.W., 1992. A simple model for plankton patchiness. Journal of Plankton Research 14, 1397–1403.

Stevens, D.P., 1990. On open boundary conditions for three dimensional primitive equation ocean circulation models. Geophysics Astrophysics and Fluid Dynamics 51, 103–133.

Strass, V.H., Woods, J.D., 1991. New production in the summer revealed by the meridional slope of the deep chlorophyll maximum. Deep-Sea Research 38, 35–56.

Vézina, A.F., Pace, M.L., 1994. An inverse model analysis of planktonic food webs in experimental lake. Canadian Journal of Fisheries and Aquatic Sciences 51, 2034–2044.

Wefer, G., 1993. Formation and composition of marine particulates. In: Heimann, M. (Ed.), The Global Carbon Cycle. NATO ASI Series I, Vol. 15. Springer, Berlin, pp. 505–530.