

Marine Biogeochemical Modeling: Recent Advances and Future Challenges

Scott C. Doney, Ivan Lima, Keith Lindsay and J. Keith Moore
National Center for Atmospheric Research • Boulder, Colorado USA

Stephanie Dutkiewicz
Massachusetts Institute of Technology • Cambridge, Massachusetts USA

Marjorie A.M. Friedrichs
Old Dominion University • Norfolk, Virginia USA

Richard J. Matear
CSIRO Division of Marine Research • Hobart, Australia

Introduction

One of the central objectives of the Joint Global Ocean Flux Study (JGOFS) is to use data from the extensive field programs to evaluate and improve numerical ocean carbon-cycle models. Substantial improvements are required if we are to achieve a better understanding of present-day biogeochemical properties and processes in the ocean and to predict potential future responses to perturbations resulting from human activities. We have made significant progress in this regard and expect even greater strides over the next decade as the synthesis of JGOFS data sets is completed and disseminated to the broader scientific community.

Marine biogeochemical modeling depends inherently upon field data. The data sets from the U.S. JGOFS process studies, global survey and time series programs, together with products derived from satellite-based observations, are invaluable in two ways. They serve as the basis for new and improved mechanistic parameterizations of specific biogeochemical processes. They also provide a resource for evaluating the overall skill of integrated system models through detailed model-data comparisons.

Ocean carbon-cycle models cover a variety of complexities and applications, ranging from simple box models to global four-dimensional coupled physical-biogeochemical simulations and from dedicated research tools to constructs able to generate climate-change projections with direct societal implications. We highlight some recent modeling advances and challenges for the future, drawing on results from the U.S. JGOFS Synthesis and Modeling Project, which are available at <http://usjgofs.whoi.edu/mzweb/syn->

[mod.htm](#) (see also Doney, 1999). Specific topics covered are phytoplankton production and community structure, interannual climate variability, mesoscale biological-physical interactions, data assimilation, export flux and subsurface carbon cycling, and responses to climate change.

Phytoplankton Production and Community Structure

Primary production in the surface ocean is the base for almost all marine food webs. Biological oceanographers devote considerable effort to developing conceptual and numerical models of the controls on phytoplankton production. Evidence accumulated over the last decade suggests that the micronutrient iron, rather than macronutrients such as nitrogen and phosphorus, governs primary production and phytoplankton community structure, especially the growth of the large diatoms, over much of the world ocean. Shifts between assemblages composed of nano- and pico-sized phytoplankton and ones dominated by diatoms, which are often responsible for seasonal and episodic blooms, affect the production and export of particulate organic carbon (POC), a critical component of the ocean "biological pump" (see Berelson, this issue). Other important groups of phytoplankton include nitrogen-fixing diazotrophs such as *Trichodesmium*, which provide a source of "new" nutrients to the subtropical gyres, and calcium carbonate-forming coccolithophores, which can significantly alter surface-water carbonate chemistry and therefore air-sea fluxes of carbon dioxide (CO₂).

The basic framework for most marine biogeochem-

Ecosystem Model

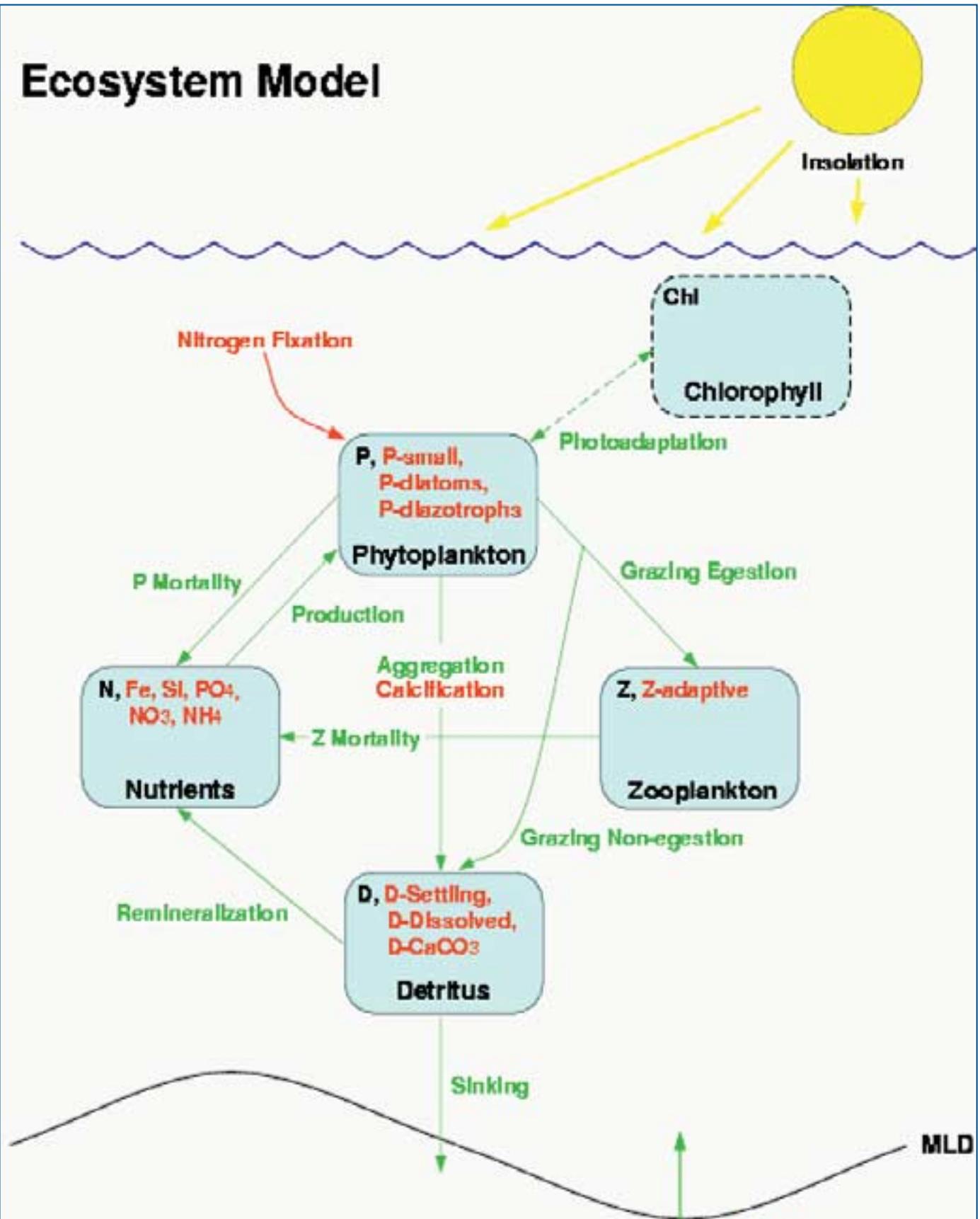


Figure 1. Diagram of a typical marine ecosystem model showing the simulated biomass compartments (boxes) and rates or fluxes (arrows) in terms of the concentration of nitrogen. In red are recent extensions of the model to incorporate multiple nutrient limitation, size structure and planktonic functional groups (Moore et al., 2001).

ical models (Figure 1) has been in use for several decades (Fasham et al., 1990). These models generally aggregate plankton populations into broadly defined trophic compartments (phytoplankton, zooplankton, detritus) and track the flow of a limiting element, such as the concentration of nitrogen or carbon, among the compartments. Biological systems have no analog to the Navier-Stokes equations of fluid dynamics; ecosystems models are by necessity highly empirical, non-linear and full of formulations based on poorly constrained parameters. The various terms for processes such as photosynthesis by phytoplankton, zooplankton grazing or detrital remineralization are calculated using standard, though not always well agreed-upon, sets of empirical functional forms derived either from limited field data or from laboratory experiments. A number of groups are working to expand this genre of model to include new concepts related to iron limitation and biogeochemical functional groups.

For example, Moore et al. (2001) present a global mixed-layer ecosystem model that explicitly accounts for multi-nutrient limitation (nitrogen, phosphorus, silica, iron), picoplankton, diatoms, nitrogen fixation, and calcification. The model has been tested against nine U.S. and international JGOFS time-series and process-study data sets drawn from a wide range of environments as well as with global ocean color data from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) instrument (see Yoder, this issue).

The marine iron cycle must be included for this model to reproduce accurately the observed high nutrient-low chlorophyll (HNLC) conditions observed in the Southern Ocean, and parts of the subarctic and equatorial Pacific (Figure 2). The atmospheric deposition of mineral dust is an important source of iron for the open ocean, especially near desert regions of North Africa, the Arabian Peninsula, northwest Asia and Australia (Figure 3). Subsurface iron contributes proportionally more in regions of upwelling, such as the equatorial Pacific, and deep winter convection, such as the subpolar North Atlantic and the Southern Ocean. The atmospheric contribution of iron input is generally low in the HNLC areas, which tend to be iron limited during summer months. The model study indicates that primary production in as much as half of the world ocean today may be iron limited.

Two main factors limiting progress on ecosystem modeling are our skill at conceptualizing key processes at a mechanistic level and our ability to verify model behavior through robust and thorough model-data comparisons. The phytoplankton iron limitation story offers an illuminating example. Atmospheric dust/iron deposition estimates vary considerably, by a factor of 10 or more in some areas, and the fraction of iron in dust that is biologically available is not well known. Surface and subsurface ocean iron measurements are limited, particularly from a global modeling perspective, and serious analytical and standardization issues

remain. Organic ligands may play a role in governing both bioavailability and subsurface iron concentrations. Not enough is known about the effect of iron limitation and variability on species competition at generally prevailing low iron levels. The same is true for a host of other processes, including iron release through photochemistry and zooplankton grazing, advection of iron from ocean margin sediment sources and iron remineralization from sinking particles.

Interannual Climate Variability

A key measure of the usefulness of numerical biogeochemical models is their ability to produce accurate descriptions of oceanic responses to natural climate forcing on interannual to interdecadal timescales. Ocean ecosystems exhibit significant variability associated with cyclical climate modes such as the El Niño-Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO) and the North Atlantic Oscillation (NAO). The response of ecosystems to physical conditions may be quite nonlinear; in the North Pacific, for example, a major

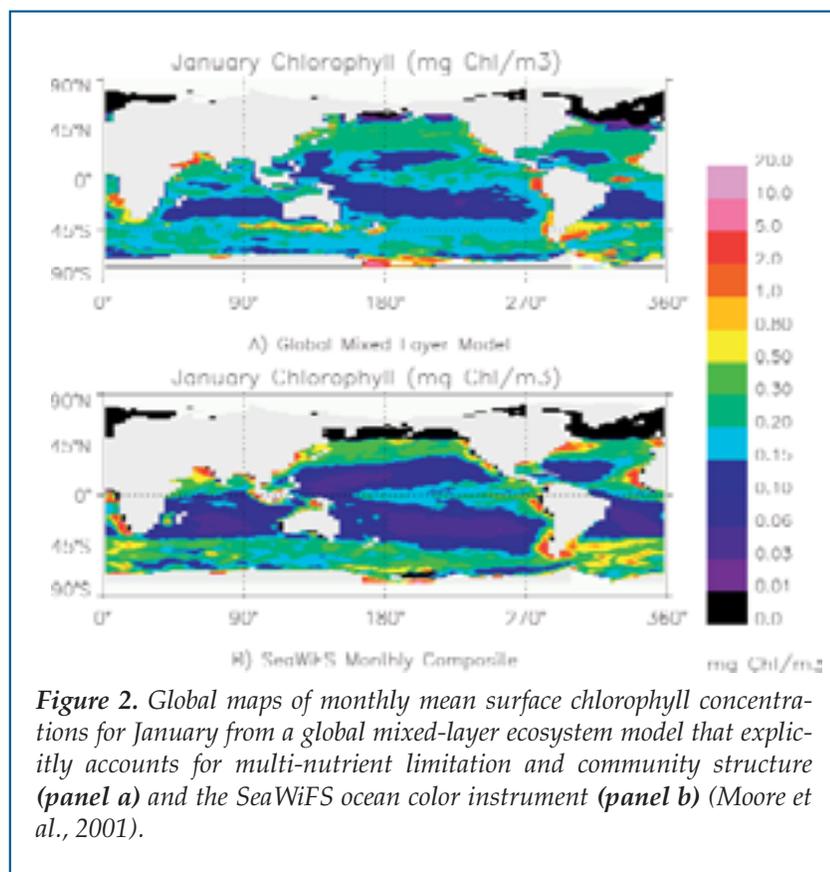


Figure 2. Global maps of monthly mean surface chlorophyll concentrations for January from a global mixed-layer ecosystem model that explicitly accounts for multi-nutrient limitation and community structure (**panel a**) and the SeaWiFS ocean color instrument (**panel b**) (Moore et al., 2001).

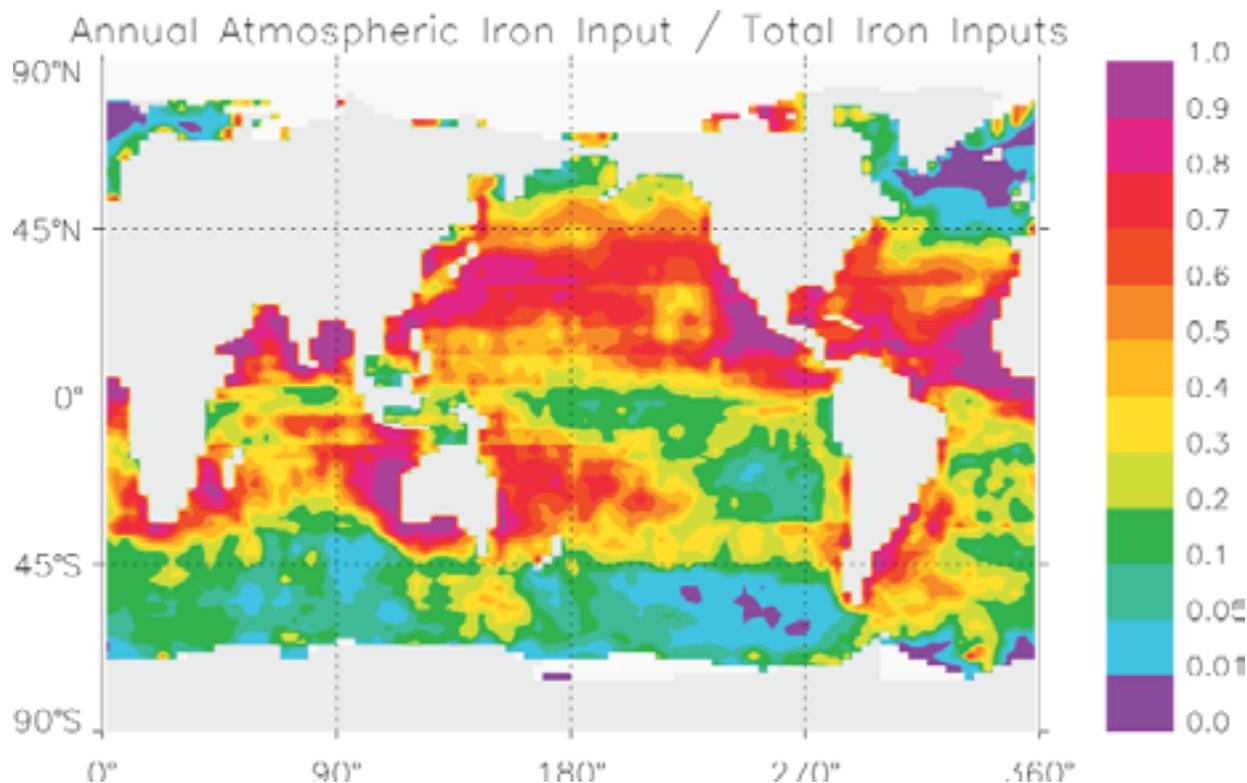


Figure 3. Global modeling simulation of the fractional contribution of atmospheric mineral dust deposition to total inputs of bioavailable iron in the surface ocean (Moore et al., 2001).

shift in biological regime took place in the mid-1970s in association with the onset of a warm phase of the PDO. Retrospective simulations can help explain the mechanisms that underlie these phenomena and provide clues as to how marine ecosystems could respond to future changes in climate.

Outside of the tropics, variability in boundary-layer mixing that accompanies changes in surface meteorological forcing from storms or winter heat loss affects spring and summer phytoplankton abundances. Simple theoretical models suggest that ecosystem response varies regionally, with increased mixing leading to increased subtropical productivity, which is primarily nutrient limited, and decreased subpolar productivity, which is primarily light limited. The boundaries of the two regimes can be roughly determined by comparing winter mixed-layer depth to the spring critical-layer depth, a measure of phytoplankton light limitation dating back to Harald Sverdrup in 1953.

Using an ecosystem model coupled to the Massachusetts Institute of Technology (MIT) three-dimensional ocean general circulation model, Dutkiewicz et al. (2001) examine this hypothesis in terms of geographical patterns and interannual variability of the spring bloom in the North Atlantic.

Simulations with the coupled model are able to reproduce the gross regional and seasonal distributions of chlorophyll over the basin. The three specific regions selected for investigation are subtropical, subpolar and “inter-gyre,” which includes the convergence of the Labrador and North Atlantic currents (Figure 4). The model simulations use interannually varying meteorological fields to force convection rates and ecosystem responses. Mean annual spring chlorophyll levels respond to variations in convection in a manner broadly consistent with that predicted by the simple conceptual model for the subtropics and the subpolar gyre, although the subpolar response is noticeably weaker. There is no obvious correlation in the model “inter-gyre” box, where other mechanisms such as horizontal nutrient transport may play a more important role (Follows and Dutkiewicz, 2001).

A corresponding diagnostic calculation has been carried out with chlorophyll concentrations derived from three years worth of SeaWiFS data and meteorological analysis from the National Center for Environmental Prediction to see whether the model results described above would be borne out by remote-sensing observations. The subtropical region shows a similarly good correlation between surface chlorophyll and turbulent mixing, estimated from the surface wind

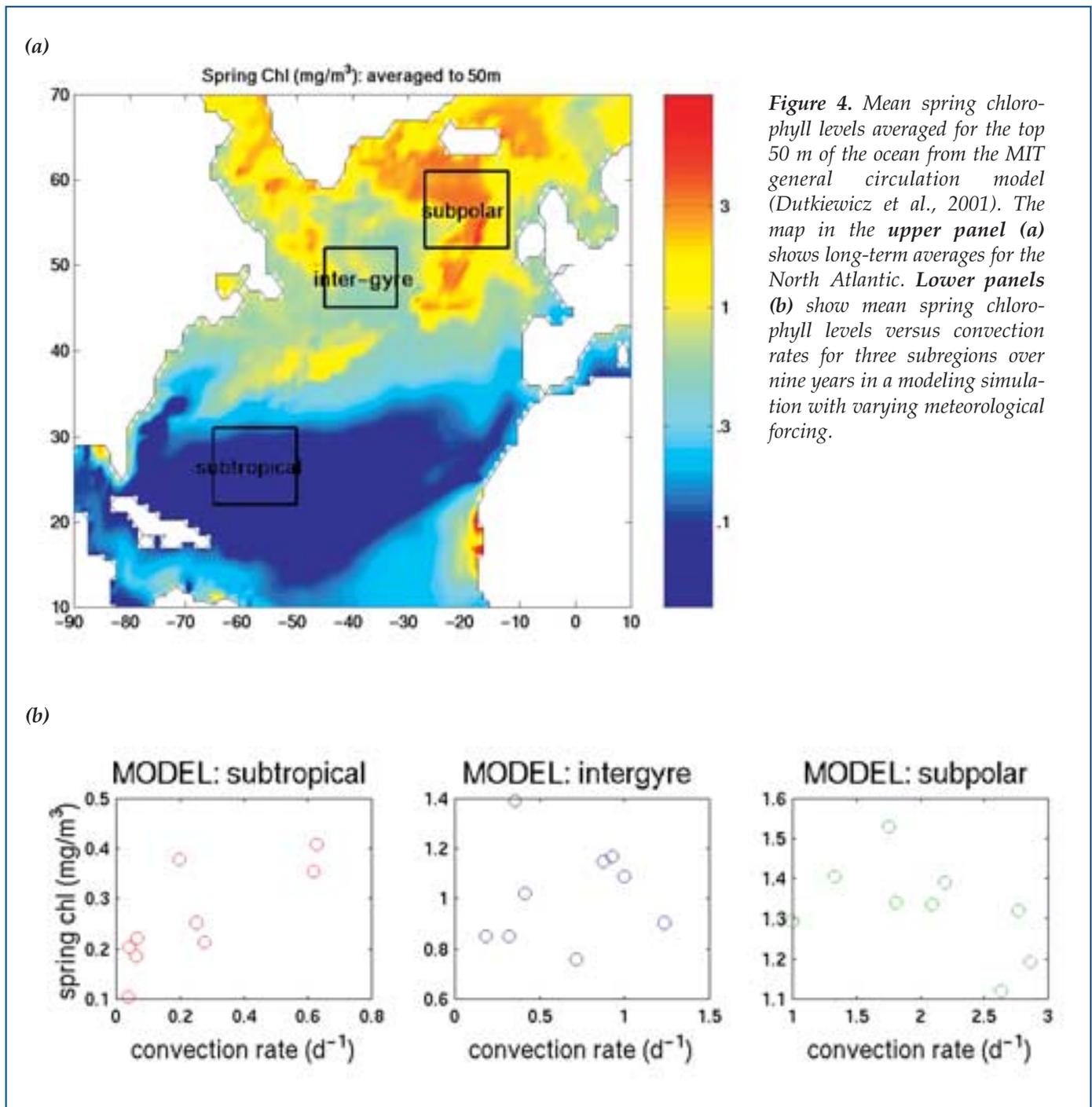


Figure 4. Mean spring chlorophyll levels averaged for the top 50 m of the ocean from the MIT general circulation model (Dutkiewicz et al., 2001). The map in the **upper panel (a)** shows long-term averages for the North Atlantic. **Lower panels (b)** show mean spring chlorophyll levels versus convection rates for three subregions over nine years in a modeling simulation with varying meteorological forcing.

stress and buoyancy fluxes, both between subregions and from year to year. The results of the diagnostic calculation for the subpolar region are less convincing, with essentially no correlation for interannual variability. The changes in chlorophyll predicted by the model that are associated with variations in surface mixing may be counterbalanced by other factors in the subpolar gyre. Relationships among physical forcing, more complex ecosystem interactions and competition among multiple species also must be addressed.

Such modeling studies offer a possible approach to projecting future ecosystem responses under various

climate change scenarios, at least for those regions where the key physical variables can be identified. Coupled ocean-atmosphere model simulations, while differing considerably in their details, generally show warming of the upper ocean and thermocline, weakening of the thermohaline circulation, and increased vertical stratification both in the low latitudes as a result of warming and in the high latitudes as a result of freshening surface waters. One might be so bold as to infer that subtropical production should decrease as nutrients become less available. This inference is contradicted, however, by evidence from the subtropical North

Pacific, where increased stratification associated with a warm phase of the ENSO cycle appears to have stimulated a shift in ecosystem structure toward organisms capable of nitrogen fixation, with resulting increases in primary production and surface chlorophyll levels (see Karl et al., this issue).

Mesoscale Biological-Physical Interactions

The sea is a turbulent medium in which variability on spatial scales of 10 to 200 km and temporal scales of days to weeks is a ubiquitous feature of the physical environment. These mesoscale variations are not typically resolved in ocean circulation models that focus on basin to global scales. New *in situ* measurements and high resolution models are providing evidence that mesoscale variability is a crucial factor governing the structure and functioning of pelagic ecosystems (see McGillicuddy, this issue). Eddies pump nutrients into the euphotic zone, stimulating rates of primary productivity and organic matter export and, potentially, shifts in community composition. The computational demands of basin- to global-scale calculations that resolve eddies accurately are significant; many of the recent advances arise from idealized process-model studies.

Figure 5 presents numerical results obtained with a Princeton Ocean Model simulation for an idealized oligotrophic subtropical frontal zone (Lima et al., 2001). As the front becomes unstable, regions of upwelling and downwelling form along the upstream and downstream sides of meander crests, leading to elevated surface nutrients and primary productivity, particularly on the north side of the front and in cyclonic meanders. Because of strong surface velocities and local patterns of convergence and divergence, regions of biomass accumulation are separated from production, with filaments of high phytoplankton concentration on the southern side of the front and around newly generated cyclonic eddies. The predicted spatial patterns of nutrients, phytoplankton and zooplankton from the model agree well in general with those reported from field observations.

The ecosystem model includes two size classes for both phytoplankton and zooplankton. In undisturbed areas, the smaller phytoplankton are abundant near the surface where nutrient concentrations are low and recycling is high; the larger, diatom-like phytoplankton are most numerous near the subsurface nutrient gradient or nutricline (50–75 m) and contribute disproportionately to new production in the simulation (Figure 5c and 5d). The rise of the nutricline inside cyclonic eddies results in near-surface blooms of both phytoplankton classes, vertical redistribution of the phytoplankton and a shift in the overall community structure towards the larger size class and thus higher export flux. Increased phytoplankton concentrations below 100 m indicate downwelling of near surface waters and an additional export flux associated with the downward advection of living biomass and detritus.

Based largely on such calculations, a semi-quantitative picture is emerging for specific processes and regions. An important next step is to undertake modeling simulations in the submesoscale range (roughly 1 km in space and a few hours in time), where vertical motions can be even stronger. Another is to expand simulations to the basin scale, where interactions with large-scale circulation and biogeography become important. It is also important to develop concerted remote sensing and field programs with data assimilation components to test and evaluate the results.

Data Assimilation

Marine ecosystem models typically include large numbers of poorly known oceanic parameters that are difficult or even impossible to measure with current instrumentation but are essential for numerical models. For example, while primary productivity measurements are common, plankton growth rates are estimated more infrequently, and mortality rates hardly ever. Improvement in consistency between data from field observations and modeling simulations can often be found by using different parameters or model structures. One possibility is to employ data assimilation techniques, which have been successful in fields such as meteorology and physical oceanography. Data assimilation is capable of providing solutions that are consistent with both observations and model equations, given estimated uncertainties. Under certain conditions these techniques provide a means for finding the best-fit or optimized set of parameters for a given model. They therefore may be a crucial part of the development and evaluation of ecosystem models.

Friedrichs (2001) describes a modeling project that focuses on a region of the equatorial Pacific Ocean that is characterized by physical variability on a wide range of time scales. ENSO cycles contribute to interannual variability, tropical instability waves (TIW) produce oscillations on monthly time scales, and equatorially trapped internal gravity waves affect the region on daily and weekly time scales. In order to capture the full range of variability in the region, daily observations of wind velocity, temperature and solar radiation from the Tropical Atmosphere Ocean (TAO) mooring array are used in the modeling simulations to force the ecosystem response. The base simulation adequately reproduces much of the variability in primary production, plankton and nutrient concentrations observed in 1992 during the two time-series cruises of the U.S. JGOFS Equatorial Pacific Process Study (EqPac).

Figure 6 shows time-series results from 50 simulations in which six model parameters (phytoplankton and zooplankton mortality, phytoplankton growth, zooplankton grazing, recycling, and the half-saturation coefficient for iron uptake) are randomly selected to be within a range of ± 30 –50% of their base values. The results clearly demonstrate that even modest changes in these parameters can lead to large differences in simulated fields. The solid line in each panel of Figure 6 is

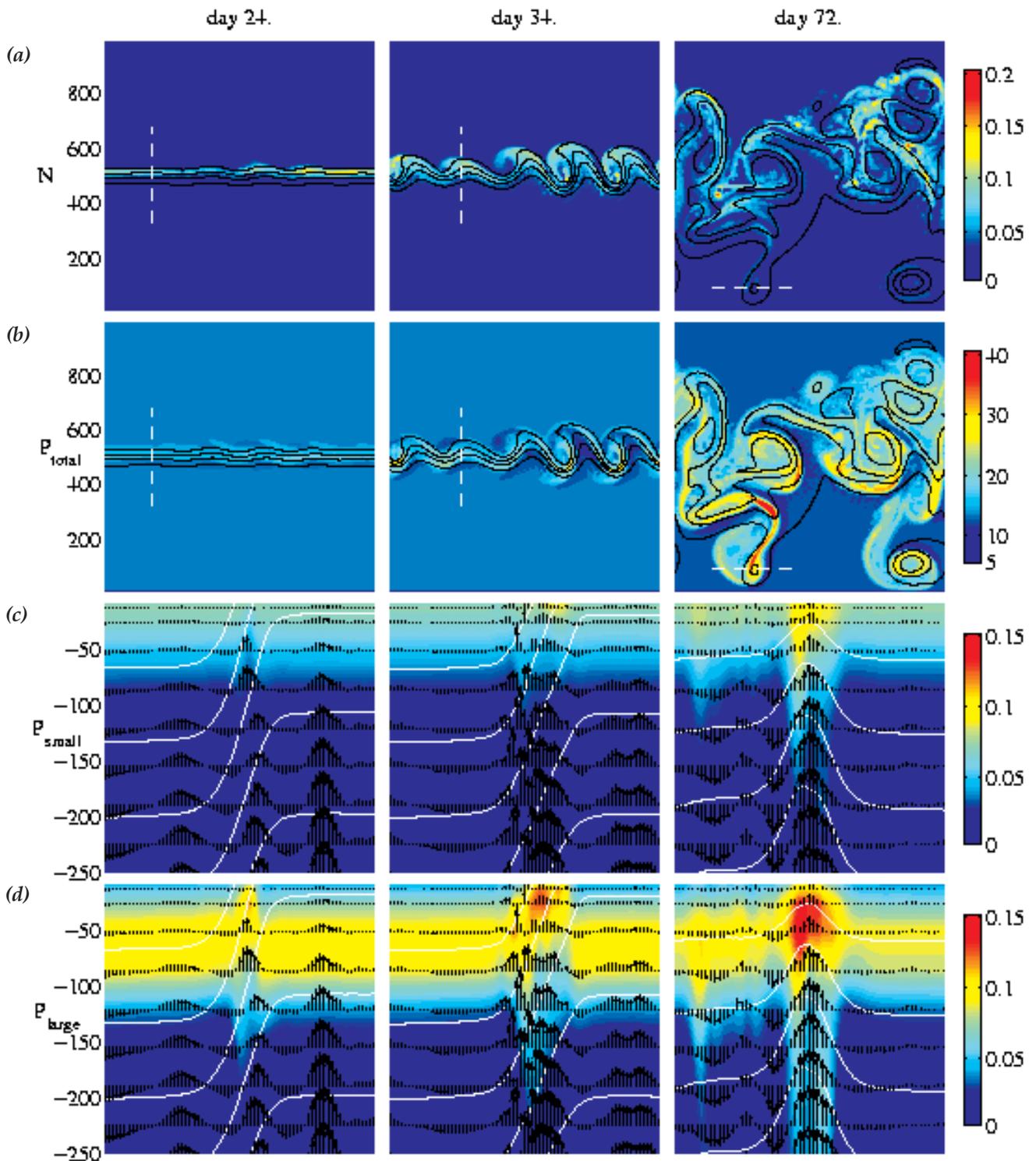


Figure 5. Results from a coupled biological-physical model of an unstable, subtropical front (Lima et al., 2001). (a) Horizontal distribution of dissolved nitrogen at 50 m (mmol N m^{-3}). (b) Horizontal distribution of vertically integrated total phytoplankton inventories (mmol N m^{-2}). The dashed white lines mark the vertical sections (depth on the y-axis) shown in the bottom two rows for (c) small and (d) large phytoplankton biomass (mmol N m^{-3}). Horizontal distances are in km, and vertical distances are in m. Solid lines represent sea-surface elevation in (a) and (b) and isopycnal surfaces in (c) and (d). Black arrows denote direction and magnitude of vertical velocities (m d^{-1}).

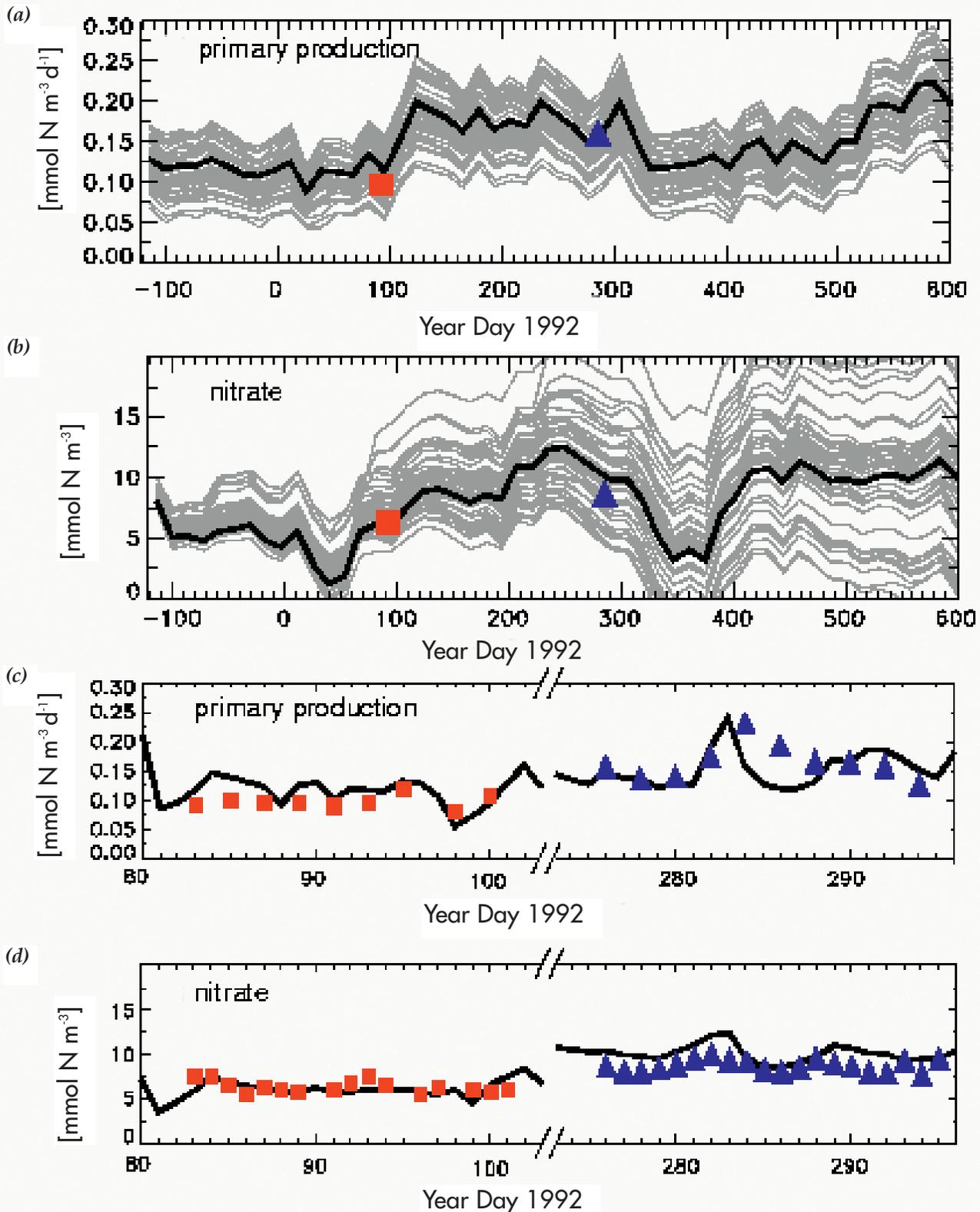


Figure 6. Simulated evolution of primary production and surface nitrate in the equatorial Pacific from a coupled biological-physical mixed-layer model (Friedrichs and Hofmann, 2001). Panels (a) and (b) show an ensemble of 50 cases (each case shown separately as a grey line) with the model parameter values drawn randomly about the base estimates and the parameter set optimized for the first U.S. JGOFS EqPac time-series cruise data (solid line) using an adjoint data assimilation technique (Friedrichs, 2001). The solid symbols represent observations from the two EqPac time-series cruises. Panels (c) and (d) show more detailed model-data comparisons with the EqPac data.

estimated with the optimal parameter set found from the first EqPac time-series cruise using data assimilation techniques.

Using this approach, one can determine objectively whether or not a given model structure is consistent with specific sets of observations. In fact no valid parameter sets are found when either the data from the second EqPac time-series cruise or SeaWiFS ocean color data are assimilated. The model failures are linked to changes in species composition associated with the passage of a TIW in the first case and, in the second, to a brief period of macro-nutrient limitation during the El Niño conditions that prevailed in the region during 1997 and 1998. While not overcoming inappropriate model dynamics, data assimilation can highlight missing processes and guide the reformulation of models.

Many questions remain before data assimilation becomes routine for biological-physical models. For example, the best method for assimilating both physical and biological data simultaneously is still under investigation. Furthermore we currently have far more physical than biological data, though this constraint should lessen in the future as more autonomous *in situ* and remote-sensing methods are developed.

Export flux and subsurface carbon cycling

We do not have as good numbers for the export of organic carbon from surface ocean waters as we do for phytoplankton production. A decade ago, investigators regarded sinking particles as the dominant pathway for the export of organic carbon from the euphotic zone, but a more complex picture has emerged during the JGOFS era (see Ducklow et al., this issue). Vertical migration of zooplankton, mesoscale subduction of living biomass, and the formation and downwelling of dissolved organic matter also contribute to the export flux. Most of the primary production is remineralized (converted back into inorganic form) in the surface layer; the ratio of export to total production varies between 10% and 50%. Export efficiency appears to be highest in cold, productive regions, decreasing in the oligotrophic subtropical gyres both because of the shift in community structure to a system dominated by picoplankton and microzooplankton and increased temperatures, which increase respiration losses at each trophic level (Laws et al., 2000).

Figure 7 presents a map of mean annual organic carbon export in the global ocean from a modeling simulation by Moore et al. (2001). The upper panel shows the sinking particulate component of the export flux, and the lower shows the total flux of biogenic carbon, which also includes export associated with the mixing of semi-labile dissolved organic matter (DOM) to depth (see Hansell and Carlson, this issue). The DOM export in the model accounts for roughly a third of all carbon export from the surface mixed layer. In portions of the mid-ocean gyres and in some high-latitude regions, it exceeds the export via sinking particles.

Almost all of the semi-labile DOM in the model is remineralized in the shallow thermocline.

The partitioning of the export flux of carbon between DOM and sinking particles and the depth and time scales associated with remineralization of sinking material are important determinants of the vertical redistribution of nutrients and carbon in the ocean. Unfortunately, we do not know much about the ecosystem dynamics of the mesopelagic zone, from the base of the euphotic zone to about 1000 m, and most models use empirical particle flux relationships derived from subsurface sediment traps (see Berelson, this issue). Some improvements are underway, such as efforts to link fluxes of organic matter and ballast materials, but more field data at the mechanistic level are clearly needed.

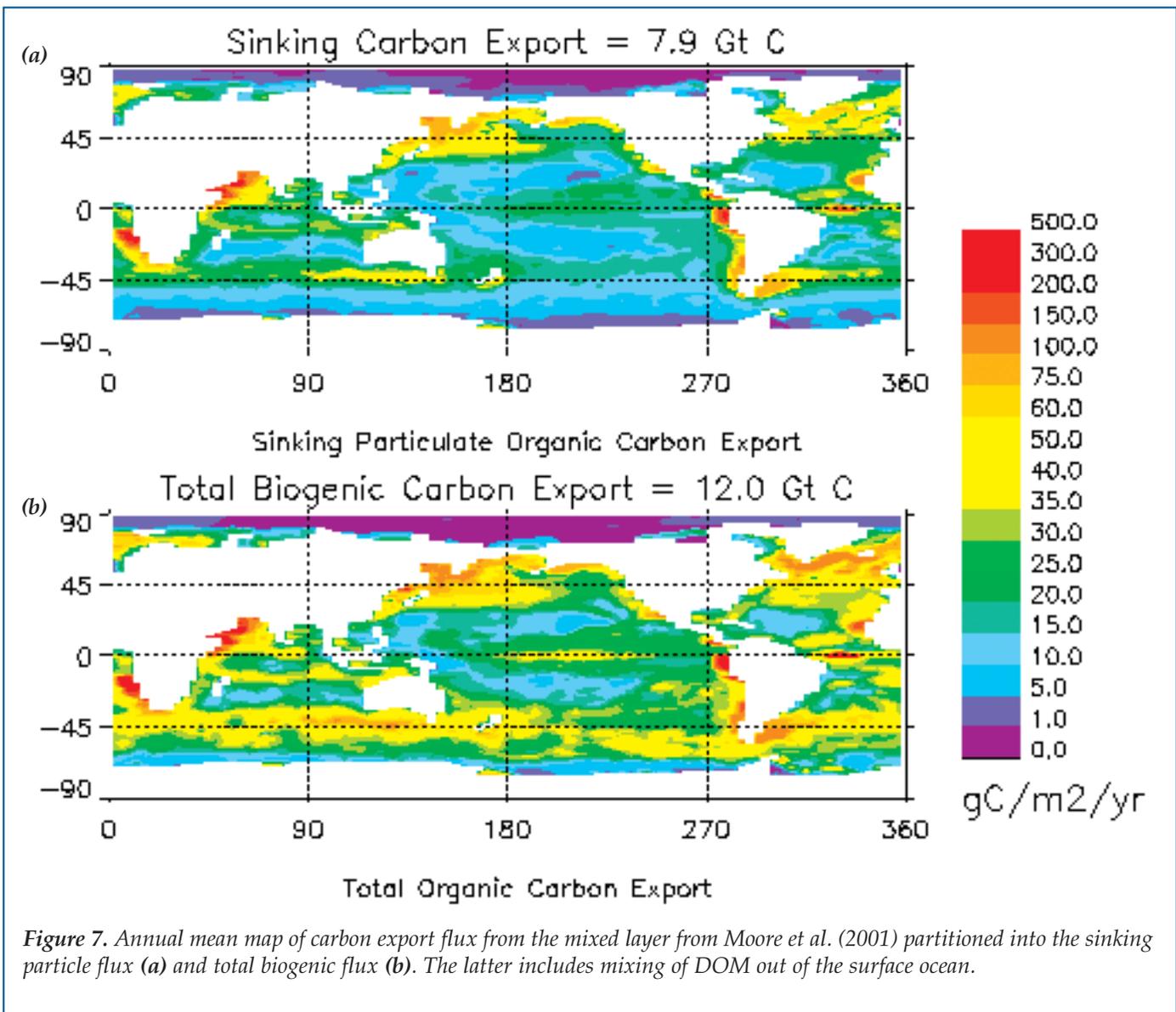
Global three-dimensional ocean carbon-cycle models are now widely used to estimate basin- and global-scale patterns and rates of biogeochemical processes such as export production, remineralization, and the uptake of anthropogenic CO₂ (Maier-Reimer, 1993; see also <http://www.ipsl.jussieu.fr/OCMIP/>). Large-scale data sets of inorganic nutrients, oxygen, dissolved inorganic carbon and hydrography are essential for evaluating such simulations. The JGOFS global survey of CO₂ in the ocean, carried out in cooperation with the World Ocean Circulation Experiment, provides an excellent, high-quality example of what is needed.

Figure 8 shows the distribution of pre-industrial dissolved inorganic carbon (DIC) along north-south transects in the Pacific and Atlantic, determined from both National Center for Atmospheric Research (NCAR) global ocean modeling simulations and observational estimates (Doney et al., 2001). Despite the simplicity of its biogeochemical module, the NCAR model captures many of the large-scale ocean biogeochemical patterns found in the observations. The “biological pump” provides about two-thirds of the increase in DIC from surface to deep waters, and the physical processes of the “solubility pump” provide about one-third. Horizontal gradients in the deep water are determined by a mixture of the thermohaline circulation and the subsurface particle remineralization rate; several of the significant model-data differences can be ascribed, in part, to problems with physical processes in the model.

Future Climate Change

Investigators have incorporated simple marine biogeochemical models in ocean-atmosphere climate simulations to project the effect of the anthropogenic increase in greenhouse gases on the air-sea exchange of CO₂ over the next century (Matear and Hirst, 1999). These simulations suggest that oceanic CO₂ uptake will be reduced as a result of greenhouse warming (Figure 9), providing a positive feedback response to rising atmospheric CO₂.

Three potentially competing mechanisms are involved. First, elevated sea-surface temperatures



decrease the solubility of CO₂ in seawater and thus the capacity to sequester anthropogenic carbon. Second, increased density stratification in the upper ocean and lower meridional overturning rates reduce the transport of anthropogenic CO₂ into the ocean interior. Third, slower ocean circulation rates decrease the supply of nutrient- and carbon-rich water to the upper ocean.

In the Matear and Hirst (1999) simulations, the third factor led to an increase in the overall efficiency of the biological pump, enabling the ocean to absorb more CO₂. The likely responses of biological processes in the ocean to changing climate conditions are not well understood at present, however, and their treatment in current ocean carbon models is quite simplistic. The ongoing JGOFS synthesis will provide the information needed to formulate more realistic models of the present and future behavior of the ocean biological pump and to improve projections of oceanic CO₂ uptake.

In order to develop greater confidence in future projections, it is essential to identify specific climate-change signals from the models that are observable in the real ocean. Model simulations suggest that dissolved oxygen concentrations may be a sensitive early indicator of changes in ocean circulation and warming. Climate change is expected to stimulate a net outgassing of oxygen from the ocean and a corresponding decrease in dissolved oxygen in the ocean. Multi-decadal changes in dissolved oxygen in the Southern Ocean modeled by Matear and colleagues (Matear et al., 2000) agree remarkably well with changes in measured oxygen levels between 1968 and 1996 (Figure 10). Further investigations of dissolved oxygen changes in the ocean are needed to determine their spatial extent and to provide valuable observations to assess climate model projections.

Future Directions

With a few exceptions, the treatment of biological

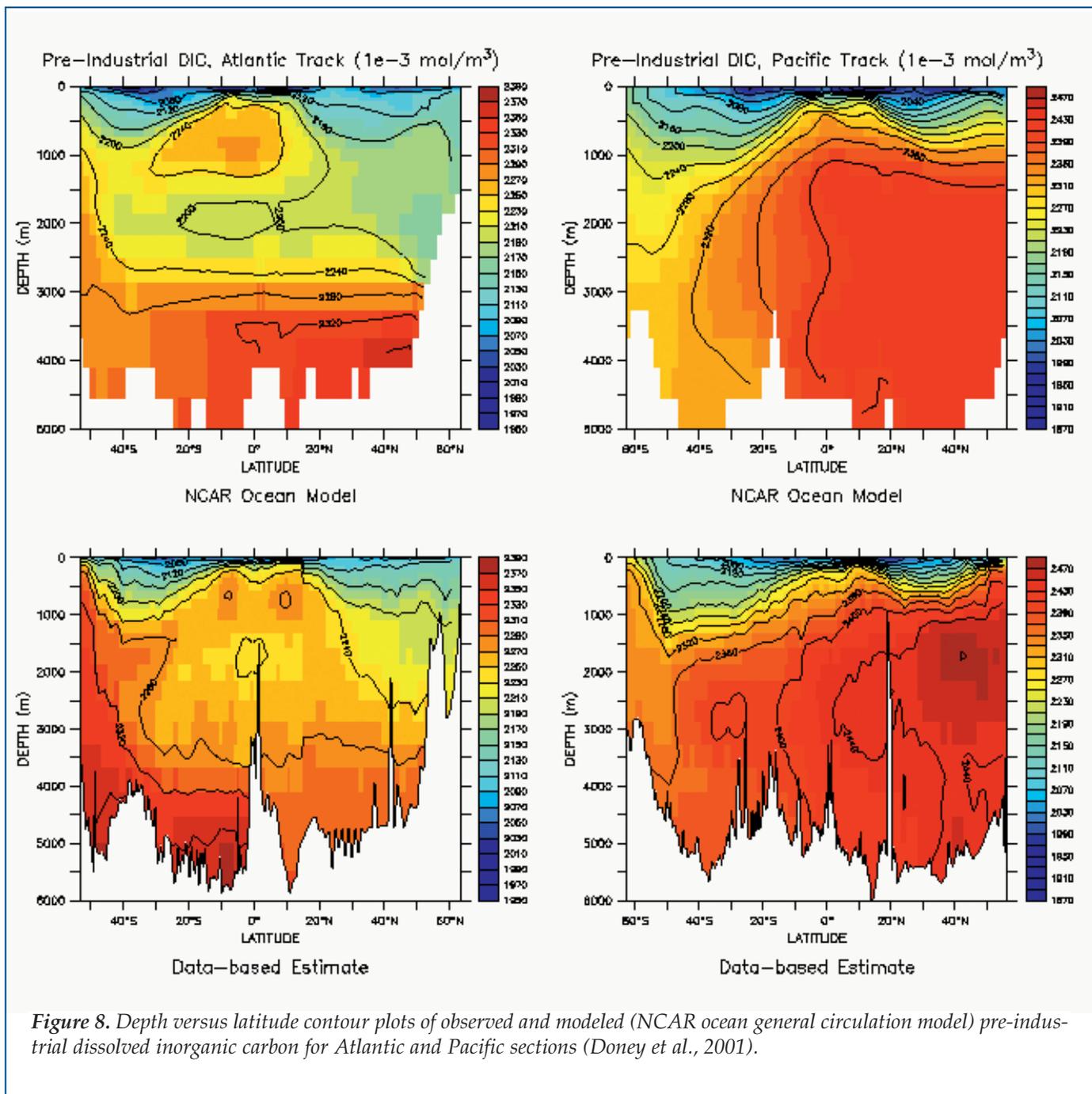


Figure 8. Depth versus latitude contour plots of observed and modeled (NCAR ocean general circulation model) pre-industrial dissolved inorganic carbon for Atlantic and Pacific sections (Doney et al., 2001).

processes and ecosystem structure in global biogeochemical models to date has been rather rudimentary. The next step, already underway, is to combine reasonably sophisticated components for both ecosystem and biogeochemical dynamics in a global modeling framework. The exact form of such models is yet to be determined. But based on the new insights emerging from JGOFS and other recent field studies, one can envision a minimal version covering the basic processes that govern surface production, export flux, subsurface remineralization and the coupling and decoupling of carbon from macronutrients. Important topics that need to be addressed include multi-nutrient

limitation, size structure and trophic dynamics, plankton geochemical functional groups, microbial loop and DOM cycling, subsurface particle transport and remineralization.

Clearly some areas of the field are advancing more rapidly than others, driven primarily by the availability of field observations, particularly by data that elucidate the fundamental mechanisms of the system. Considerable effort has been devoted to understanding processes governing phytoplankton primary production at a number of levels, including molecular biophysics, cellular physiology, community dynamics, seawater bio-optics and remote sensing. Other areas,

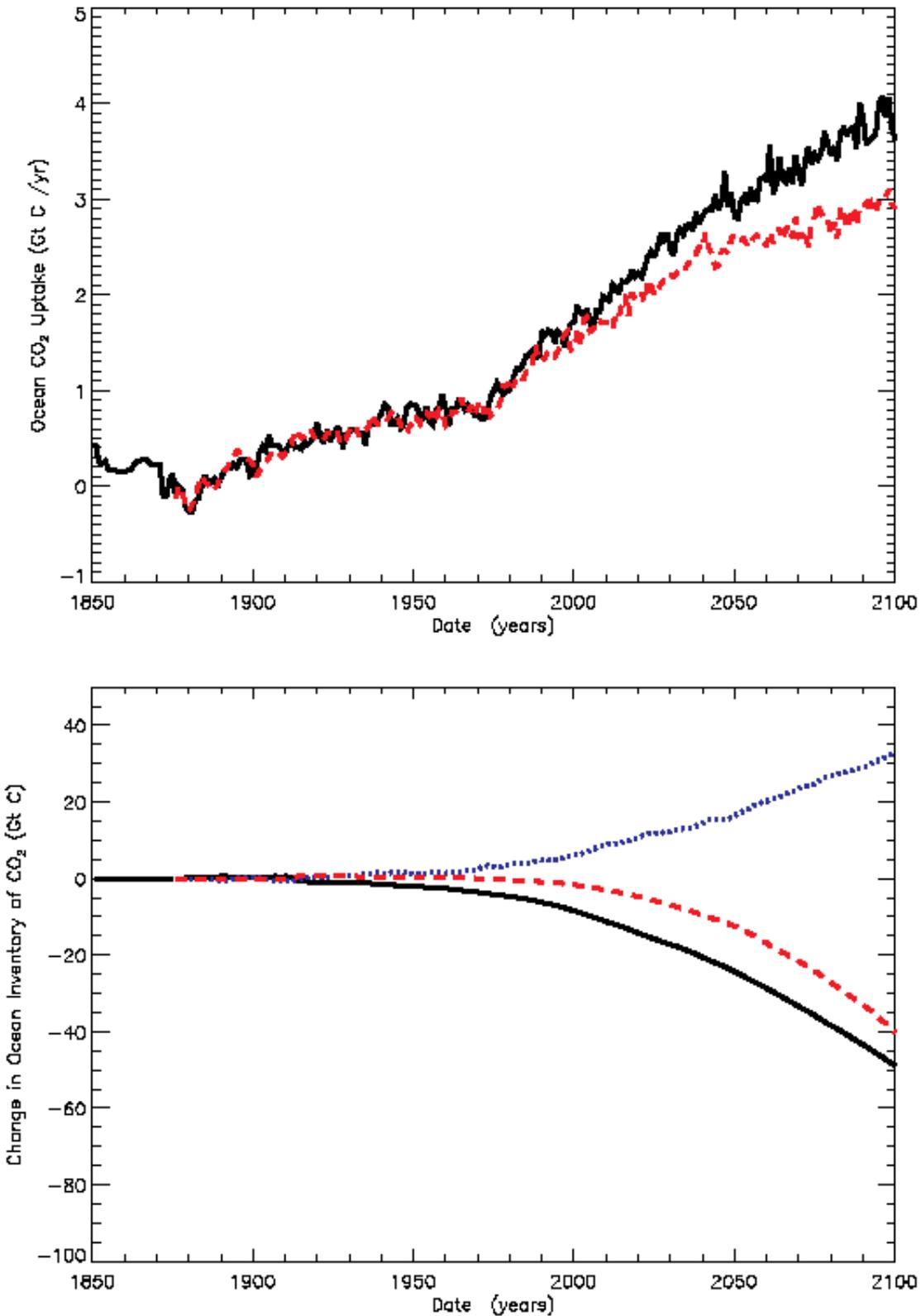


Figure 9. The effect of greenhouse warming on projected oceanic CO₂ uptake over the next century from the CSIRO climate model (Matear and Hirst, 1999). Upper panel shows global oceanic CO₂ uptake for a control case (solid black line) and for a climate-change scenario (red dashed line). Lower panel shows mechanisms driving the cumulative oceanic CO₂ response: elevated sea-surface temperature (black solid line), reduced transport of anthropogenic CO₂ into the interior (red dashed line), and increased efficiency of the ocean biological pump (blue dotted line). The sum of these three curves equals the total change in oceanic CO₂ uptake under the climate-change scenario.

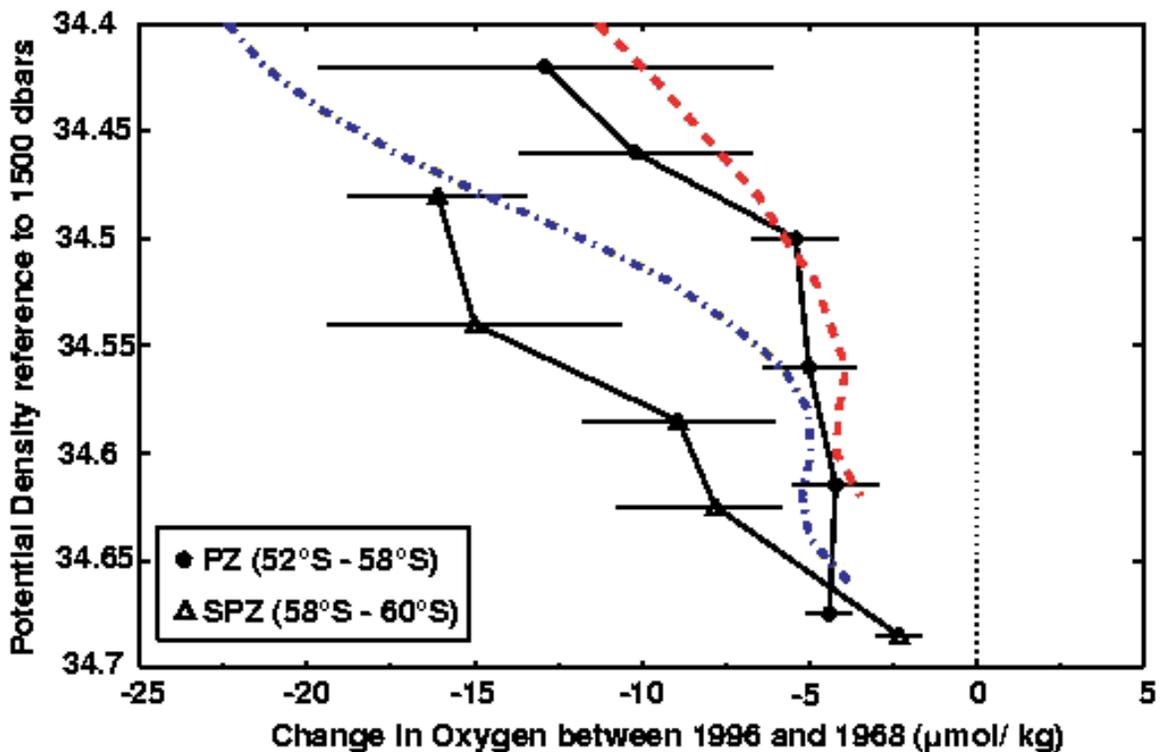
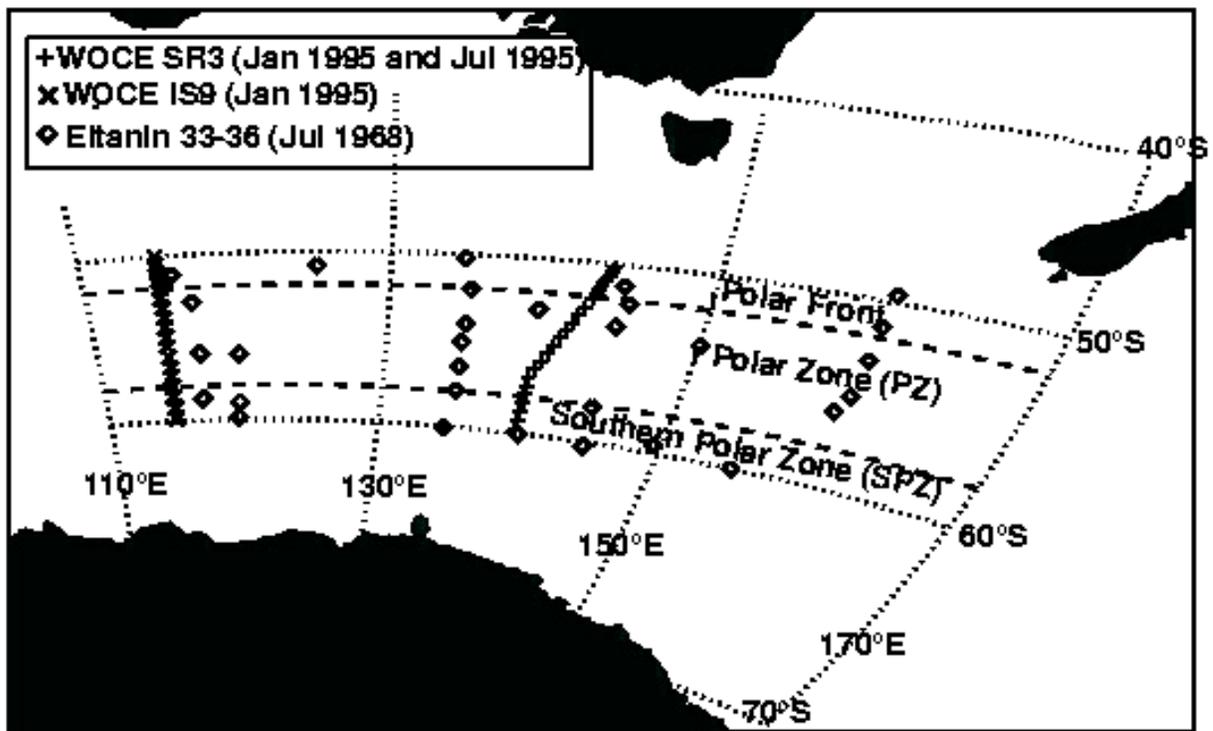


Figure 10. Observed changes in dissolved oxygen between 1996 and 1968 (solid black lines) on potential density surfaces referenced to 1500 dbars in the Polar Zone (PZ), 52°S–58°S, and Southern Polar Zone (SPZ), 58–60°S. The horizontal lines denote the two standard deviation errors. Also shown are the simulated changes in the zonal averaged oxygen (110°E to 170°E) associated with climate change for the 1990–99 period at 59°S (blue dashed line) and 55°S (red dashed line). In the model simulation, the interannual variability in dissolved oxygen was greatest at 33.4, where it was ± 4 mmol kg^{-1} for 55°S and ± 3 mmol kg^{-1} for 59°S. At densities greater than 33.5, the modeled interannual variability in dissolved oxygen was less than ± 1 mmol kg^{-1} . Upper panel shows locations of stations sampled on cruises in 1968 and 1995.

Taking Care Of The Legacy: Data Management In U.S. JGOFS

David M. Glover

*Woods Hole Oceanographic Institution
Woods Hole, Massachusetts USA*

Data management, essential to the success of a large, multidisciplinary scientific enterprise, has been a basic component of the U.S. Joint Global Ocean Flux Study (JGOFS) from the beginning. The planners of the program envisioned two main tasks associated with managing U.S. JGOFS data: the assemblage of a complete, high-quality archive of biogeochemical observations and modeling results that would be readily accessible to the scientific community, and the creation of a flexible and expandable system for providing access to data.

In 1988, with support from the National Science Foundation (NSF), Glenn Flierl, James Bishop, David Glover and Satish Paranjpe began work on an approach to the retrieval, merging and analysis of data stored in a variety of formats and located in different places. This team developed an object-oriented, distributed database system with an interface that permits users to get access to data and "metadata" (information about a data set, such as methodology used) via the world wide web, regardless of the web browser used.

Data management began with the completion of the first JGOFS process study in 1989. George Heimerdinger, northeast liaison officer for the National Oceanic and Atmospheric Administration's National Oceanographic Data Center (NODC), undertook the task of collecting, organizing and distributing U.S. JGOFS data and carrying out quality-control procedures from his office at Woods Hole Oceanographic Institution (WHOI) during the first few years.

An independent U.S. JGOFS Data Management Office (DMO) was set up at WHOI in 1994, and information systems expert Christine Hammond was hired as a full-time data manager. She worked with Flierl and Glover to improve and maintain the database system in addition to overseeing the ongoing work of collecting and organizing the U.S. JGOFS data sets, most of which are stored and served by the DMO. Among her many accomplishments was the creation of a U.S. JGOFS home page, through which users could get access to the steadily growing volume of data from the field programs.

In 2000, Hammond moved to a new position at WHOI, and Glover took over the scientific leadership of the DMO. Cynthia Chandler now acts as system manager and handles the day-to-day operation of the web-based system. David Schneider and Heimerdinger, who returned from retirement to work for the DMO, are responsible for data verification and quality assurance.

Several factors in the development of the DMO have been critical to its mission. The U.S. JGOFS DMO would not have enjoyed the success it has without the unflinching support of the NSF, the U.S. JGOFS Scientific Steering Committee and the investigators. All parties have benefited from the decision of the DMO not to require a fixed format for data submission. Scientists are encouraged to get their data into the DMO, and staff members work with them to get the data into the system.

Equally important was the decision, made at the beginning of the program, to encourage data sharing among investigators as part of U.S. JGOFS data policy. Data were to be made available to all U.S. JGOFS investigators after six months and to the entire scientific community after two years, as required by NSF. The open data policy has enhanced the value of U.S. JGOFS results by encouraging exchanges and intercomparisons.

When the initial development team was working out how users would actually interact with the data system, the powerful idea of using the world wide web as the primary interface was adopted. Any computer user with a web browser of some sort is able to get access to the database without platform-specific code.

Another important early choice was to have a data manager look at the data as they came in. Detection of anomalous data generates a quick call to the originator, and problems are generally solved before anyone makes use of contaminated data.

Finally, the decision to supply data documentation directly linked, as only hypertext can be, with the data themselves has been of great value. Users get not only the data, but also the supporting documentation files when they download data. The information in these documentation files describes methods used, calculations performed on the original data and other forms of metadata information that improve the usability of the data.

One of the new tasks undertaken by the DMO is the support of results from the U.S. JGOFS Synthesis and Modeling Project (SMP). Because of the gridded nature and larger volume of output from modeling studies, the DMO is working with the Live Access Server (LAS) group at the University of Washington and NOAA Pacific Marine Environmental Laboratory (PMEL) in Seattle to provide a visualization and access interface to these results.

Meanwhile, DMO staff members are working on the development of merged data products that combine data objects with com-

mon variables and collection methods to form larger data sets. Currently the DMO is producing both conductivity-temperature-depth (CTD) and bottle merged data objects for the various U.S. JGOFS process studies. These begin at the cruise level and are in due course merged into basin-level objects.

The data scientists leave behind represent their legacy to future generations. Currently all of the originally submitted U.S. JGOFS data are available via the world wide web, as well as some of the newly merged data products that the DMO has been building. With the web interface, users can search, view and download specific "slices" of the data or all of a given data set if they desire.

Some day in the not-so-distant future, U.S. JGOFS and the DMO will come to an end. To ensure the continued availability of U.S. JGOFS data, the DMO plans to produce and distribute a final data product, such as a set of CD-ROMs, that will include all of the field data, the merged data products and a selection of SMP models and results. At the same time, data will be submitted to the NODC, the long-term archive mandated by Congress for oceanographic data. The experts who developed and maintained the U.S. JGOFS data system may go on to support future ocean carbon-cycle research, and the data may be available on the web through these programs. With these various options, U.S. JGOFS data should continue to be available for many years to come.

such as phytoplankton loss rates or the dynamics of mesopelagic ecosystems, have received considerably less attention, not because they are considered less important from a carbon-cycle perspective, but because techniques and measurement systems in these areas are less advanced.

The modeling and synthesis component of U.S. JGOFS came at the end of the program, and with a few notable exceptions, numerical models did not strongly influence the planning for the field studies, either in terms of what was measured or of how the experiments were designed. This lack reflects, in part, the state of modeling a decade ago. But much has changed, and future field programs will require closer, more synergistic relationships between models and observations.

Finally, the recent progress in ocean biogeochemical modeling would not be possible without the treasure trove of JGOFS field data and satellite remote-sensing products now easily accessible in electronic form. U.S. JGOFS has achieved timely, public release of data, including models and modeling simulations, and has provided explicit support for data management, which is neither glamorous nor cheap (see Glover sidebar, this issue). These are major cultural and programmatic advances for our field; they provide lessons that should not be neglected in the future. This is U.S. JGOFS Contribution Number #682. 

References

- Doney, S.C., 1999: Major challenges confronting marine biogeochemical modeling. *Global Biogeochem. Cycles*, 13, 705–714.
- Doney, S.C., K. Lindsay and J.K. Moore, 2001: Global ocean carbon cycle modeling. In: *Ocean Biochemistry: #9 JGOFS synthesis*. M. Fasham, ed., in press.
- Dutkiewicz, S., M. Follows, J. Marshall and W.W. Gregg, 2001: Interannual variability of phytoplankton abundances in the North Atlantic. *Deep-Sea Res. I*, 48, 2323–2344.
- Fasham, M.J.R., H.W. Ducklow and S.M. McKelvie, 1990: A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *J. Mar. Res.*, 48, 591–639.
- Follows, M. and S. Dutkiewicz, 2001: Meteorological modulation of the North Atlantic spring bloom. *Deep-Sea Res. II*, in press.
- Friedrichs, M.A.M., 2001: The assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central Equatorial Pacific Ocean. *Deep-Sea Res. II*, in press.
- Friedrichs, M.A.M. and E.E. Hofmann, 2001: Physical control of biological processes in the central equatorial Pacific Ocean. *Deep-Sea Res. I*, 48, 1023–1069.
- Laws, E.A., P.G. Falkowski, W.O. Smith, Jr., H.W. Ducklow, and J.J. McCarthy, 2000: Temperature effects on export production in the open ocean. *Global Biogeochem. Cycles*, 14, 1231–1246.
- Lima, I.D., D.B. Olson and S.C. Doney, 2001: Biological response to frontal dynamics and mesoscale variability in oligotrophic environments: a numerical modeling study. *J. Geophys. Res.*, submitted.
- Maier-Reimer, E., 1993: Geochemical cycles in an ocean general circulation model: Pre-industrial tracer distributions. *Global Biogeochem. Cycles*, 7, 645–677.
- Matear, R.J. and A.C. Hirst, 1999: Climate change feedback on the future oceanic CO₂ uptake. *Tellus*, 51B, 722–733.
- Matear, R.J., A.C. Hirst and B.I. McNeil, 2000: Changes in dissolved oxygen in the Southern Ocean with climate change. *Geochem. Geophys. Geosystems*, 1, November 21, <http://g-cubed.org/>.
- Moore, J.K., S.C. Doney, J.A. Kleypas, D.M. Glover and I.Y. Fung, 2001: Iron cycling and nutrient limitation patterns in surface waters of the world ocean. *Deep-Sea Res. II*, accepted.