CITATION

COPYRIGHT
This article has been published in *Oceanography*, Volume 22, Number 3, a quarterly journal of The Oceanography Society. Copyright 2009 by The Oceanography Society. All rights reserved.

USAGE
Permission is granted to copy this article for use in teaching and research. Reproduction, systematic reproduction, or collective redistribution of any portion of this article by photocopy machine, reposting, or other means is permitted only with the approval of The Oceanography Society. Send all correspondence to: info@tos.org or The Oceanography Society, PO Box 1931, Rockville, MD 20849-1931, USA.
Integrating Biogeochemistry and Ecology Into Ocean Data Assimilation Systems

BY PIERRE BRASSEUR, NICOLAS GRUBER, ROSA BARCIELA, KEITH BRANDER, MAEVA DORON, ABDELALI EL MOUSSAQUI, ALISTAIR J. HOBDAY, MARTIN HURET, ANNE-SOPHIE KREMEUR, PATRICK LEHODEY, RICHARD MATEAR, CYRIL MOULIN, RAGHU MURTUGUDDE, INNA SENINA, AND EINAR SVENDSEN

ABSTRACT. Monitoring and predicting the biogeochemical state of the ocean and marine ecosystems is an important application of operational oceanography that needs to be expanded. The accurate depiction of the ocean’s physical environment enabled by Global Ocean Data Assimilation Experiment (GODAE) systems, in both real-time and reanalysis modes, is already valuable for various applications, such as the fishing industry and fisheries management. However, most of these applications require accurate estimates of both physical and biogeochemical ocean conditions over a wide range of spatial and temporal scales. In this paper, we discuss recent developments that enable coupling new biogeochemical models and assimilation components with the existing GODAE systems, and we examine the potential of such systems in several areas of interest: phytoplankton biomass monitoring in the open ocean, ocean carbon cycle monitoring and assessment, marine ecosystem management at seasonal and longer time scales, and downscaling in coastal areas. A number of key requirements and research priorities are then identified for the future. GODAE systems will need to improve their representation of physical variables that are not yet considered essential, such as upper-ocean vertical fluxes that are critically important to biological activity. Further, the observing systems will need to be expanded in terms of in situ platforms (with intensified deployments of sensors for O₂ and chlorophyll, and inclusion of new sensors for nutrients, zooplankton, micronekton biomass, and others), satellite missions (e.g., hyperspectral instruments for ocean color, lidar systems for mixed-layer depths, and wide-swath altimeters for coastal sea levels), and improved methods to assimilate these new measurements.
INTRODUCTION

In its original design, the Global Ocean Data Assimilation Experiment (GODAE) was conceived as a practical demonstration of real-time ocean data assimilation in order to provide a regular and complete depiction of global ocean circulation at eddy resolution and better consistency with observations of physical parameters and dynamical constraints (Le Traon et al., 1999). This type of oceanic information was identified in the early stages of GODAE as potentially very valuable to different applications related to the “living ocean,” such as the fishing industry and fisheries management (Griffin et al., 2002). Numerous examples of GODAE products used for environmental applications are found in the literature, ranging from real-time temperature products for monitoring seasonal fish migration (Hobday and Hartmann, 2006) and ecological regime shifts (Brander, in press), to ocean currents and velocities used to understand the transport and spread of fish larvae (Bonhommeau et al., 2009; Johnson et al., 2005), to sources and sinks of atmospheric CO$_2$ (Gruber et al., 2009). Some of these examples will be discussed in the following sections.

Today, advances in modeling biogeochemical processes and increased computer power have made it possible to couple physical and biogeochemical models online to address wider environmental and societal issues by providing hindcasts, nowcasts, and forecasts from short lead times (Siddorn et al., 2007) to climate time scales (Johns et al., 2006). These advances, together with increased pressure for an integrated ecological approach—the so-called ecosystem-based approach—to managing the marine environment, have resulted in active development and application of ocean-ecosystem models. Assimilation of biogeochemical data has a potentially critical role to play in improving coupled simulations and assessing the uncertainty of biogeochemical products. Potential applications of these models are wide ranging, from short-term, limited-area information for fisheries management to climate time-scale estimations of the global carbon cycle. In all cases, they require accurate prediction of both physical and biogeochemical ocean conditions. In this context, a new challenge for operational oceanography will be to lead a practical demonstration of near-real-time biogeochemical products that provide a comprehensive description of the ocean ecosystem variables in support of seasonal-to-decadal climate forecast and analyses, and oceanographic research.

The expansion of operational systems to biogeochemistry can well be expected in an era of climate change, requiring well-planned strategies supported by data at appropriate spatio-temporal resolution to predict and assess trends of marine biogeochemical cycles (especially for carbon) and to safeguard marine ecosystems. This paper examines four main categories of applications. The next section focuses on recent implementations that enable monitoring phytoplankton biomass in the global ocean, based on physical hindcasts of ocean circulation coupled to biogeochemical models. We then show how this type of information can be used to improve monitoring global or basin-scale carbon fluxes across the air-sea interface. We then present several developments regarding the exploitation of living resources using ecosystem-based management approaches, followed by a discussion of issues regarding regional downscaling and coastal applications. Finally, we identify a number of key requirements and research priorities needed to consolidate the approach.

PHYTOPLANKTON BIOMASS MONITORING IN THE OPEN OCEAN

Phytoplankton is a key marine component in a variety of applications, with relevant temporal scales from weeks to decades and spatial scales from global to local. Despite recent improvements in biogeochemical and ecological modeling, the coupled physical-biogeochemical models available today remain rudimentary in comparison with the actual behavior of living marine organisms. In this context, data assimilation is a relevant approach to achieve: (1) better control of the physical circulation that enhances the quality of biogeochemical dynamics, (2) initialization of the biological variables for prediction, (3) estimation of physical and biogeochemical model parameters, and (4) data-based assessments of modeling hypotheses.

The importance of mesoscale activity in primary production is the subject of much discussion in the literature (e.g., Oschlies and Garçon, 1998), and the example below confirms that the resolution of GODAE products is key to realistically monitoring phytoplankton biomass in ocean basins. During the EU-funded MERSEA (Marine Environment and Security for the European Area) project, a prototype of a coupled physical/biological assimilation system based on the PISCES model (Aumont et al., 2003) has been
developed with the objective to routinely estimate and forecast biogeochemical variables over the globe using the Mercator operational setup. The spatio-temporal evolution of biological tracers can be driven using physical hindcasts produced by the Nucleus for European Modelling of the Ocean (NEMO) global 1/4° circulation model that assimilates sea surface temperature (SST), altimetric data, and in situ temperature and salinity profiles. A real-time demonstration was performed during 2008 with weekly delivery of biogeochemical products. Despite technical simplifications due to real-time constraints (the main one is that the PISCES model is run offline), resolving mesoscale processes has an observable positive impact on the biological fields. Figure 1 illustrates the annual mean chlorophyll distribution in the North Atlantic, showing that the prototype provides realistic simulations in terms of location of productive regions at high latitudes (e.g., in the Labrador Sea). Compared to satellite data, the mean chlorophyll concentration is, however, slightly underestimated. A set of secondary products that include nutrients, primary production, euphotic layer depth, dissolved organic matter, dissolved oxygen, and surface ocean pCO₂ were also delivered using the Mercator setup.

In the future, real-time estimates of phytoplankton biomass are expected to better sustain applications such as monitoring of the carbon cycle, living resources, and water quality. Another key challenge will be to better understand and quantify the impact of climate changes on phytoplankton. There is currently no clear consensus on the response of primary producers to the climate variability observed during the last two decades. Recent analyses based on ocean color records from the Coastal Zone Color Scanner (CZCS) and the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) indicate an overall increase in the world ocean average chlorophyll concentration of about 22% (Antoine et al., 2005), while a reduction of biological activity may have taken place in oligotrophic regions (Gregg et al., 2005; Behrenfeld et al., 2006; Polovina et al., 2008). Accurate reanalyses of the physical ocean during the past 60 years, such as those expected from the Mercator reanalysis systems, will offer new insight into how phytoplankton biomass and production has varied over this period.

**OCEAN CARBON CYCLE MONITORING AND ASSESSMENT**

The operational systems available as GODAE ended are not yet able to accurately provide real-time monitoring of global or basin-scale air-sea carbon fluxes as required (e.g., for attempts to obtain reliable regional carbon budgets). Developing this capacity represents a research activity of strategic importance in the context of international conventions. Earlier attempts to obtain maps of pCO₂ and associated air-sea fluxes of CO₂ included the interpolation of in situ measurements with the help of satellite data for ocean color and surface temperature. Because coverage of in situ data is sparse and synoptic satellite images capture only the near-surface variability, and are often dominated by clouds, numerical models can help to combine and interpolate the observational information contained in various data sets. The explicit assimilation of ocean color data into coupled models is promising but also challenging, mainly because of (1) the fairly poor signal-to-noise ratio of the observed signal...
(estimated errors of ± 50% in the open ocean and as high as ± 200% in shallow waters; Ballabrera-Poy et al., 2003), (2) the complex relationship between the remotely sensed property and the model equivalent, (3) the nonlinearities of coupled physical-biological processes, and (4) the still inadequate performance of the modeling tools available today.

The UK Met Office is developing ocean color data assimilation in a pre-operational coupled physical/biogeochemical model. The focus of that work is to estimate the air-sea flux of CO$_2$ driven by the pCO$_2$ differences across the air-sea interface based on the Forecasting Ocean Assimilation Model (FOAM) coupled to the Hadley Centre Ocean Carbon Cycle Model (HadOCC). The primary aim of assimilating ocean color is therefore to improve seawater pCO$_2$ estimates. Biological controls on pCO$_2$ in the coupled model result from biotic modification of total dissolved inorganic carbon (DIC) and alkalinity by a combination of photosynthesis, respiration, and calcification. In this context, an ocean color assimilation scheme must make the best use of the assimilated information for correcting not only the phytoplankton components but also all the components of the system affecting DIC and alkalinity. The scheme does exactly that by balancing daily surface phytoplankton increments, derived from univariate analysis of surface chlorophyll, in a nitrogen-phytoplankton-zooplankton-detritus (NPZD) nitrogen-cycle model (Hemmings et al., 2008). The scheme performs well as, on average, it improves estimates for most of the biogeochemical variables in a way that cannot be achieved by any simple method without detailed nitrogen balancing.

Gloor et al. (2003) developed a complementary approach to determining the exchange of CO$_2$ across the air-sea interface, and it was further refined by Mikaloff Fletcher et al. (2006, 2007). In this inverse method, air-sea fluxes of natural and anthropogenic CO$_2$ are adjusted until the model-derived ocean interior distribution of these two tracers agrees optimally with their observationally derived distribution. Comparison of the inversely derived fluxes of contemporary CO$_2$—the sum of natural and anthropogenic CO$_2$, with those derived from pCO$_2$ data—shows excellent agreement, except for the Southern Ocean (Figure 2). These differences are attributed to a combination of data limitations and poor model physics in this part of the ocean (Gruber et al., 2009). It is interesting to note that one of the 10 models used in the inversion to determine the mean and standard distribution of the estimated fluxes was based on Estimating the Circulation and Climate of the Ocean (ECCO; a GODAE product). In fact, this GODAE-derived transport model was found to be one of the best performing, based on its superiority in modeling the distribution of chlorofluorocarbons and natural radiocarbon (Mikaloff Fletcher et al., 2006, 2007). As a consequence, the results from this particular model were weighted more than those from other
models. So far, these inversions were performed with climatological ocean transport fields, resulting in long-term annual mean fluxes of CO₂ and of its two components. The next steps include the development of methods that permit inversion on seasonal and interannual time scales as well as increasing the horizontal resolution.

**MARINE ECOSYSTEM MANAGEMENT**

Marine exploitation issues need to be addressed from an ecosystem perspective, that is, using an ecosystem-based management approach (e.g., Svendsen et al., 2007). The long-term vision is to develop the systems required for the study, management, and monitoring of exploited and protected marine species. They need to cover pelagic (e.g., tuna, mackerel, billfishes, sharks, marine mammals, turtles), demersal (e.g., cod, pollack), as well as continental margin and coastal domains (e.g., herring, sardine, anchovy, sea birds). One key explanatory variable that is usually missing when attempting to understand individual behavior of these large oceanic predators is the spatial distribution of their prey species. The Spatial Ecosystem And Population Dynamic Model (SEAPODYM) is one proposed approach to modeling the spatial dynamics of large predator (e.g., tuna) populations (Lehodey et al., 2008) that are interacting with their modeled prey. These prey are described by several mid-trophic functional groups of the pelagic system in three vertical (epi-, meso-, and bathy-pelagic) layers. In this model, the dynamics of forage and predators are driven by environmental forcing (temperature, currents, oxygen, and primary production), which can be predicted from coupled physical-biogeochemical model outputs. Another key issue for the coming years will be to gather the information necessary for calibrating this type of mid-trophic model based on functional groups or
other approaches, for example, a continuous size spectrum (Maury et al., 2007). Ultimately, collected data should be used for parameter optimization and data assimilation (Senina et al., 2008).

From the user’s perspective, oceanographic information is used in fisheries management in two basic ways. First, traditional fisheries oceanography has involved a search for explanation of historical catch-related patterns. Environmental variables such as SST, mixed-layer depth, and chlorophyll are used as covariates along with fishing-related variables to explain observed variation in catch rate (e.g., Herron et al., 1989; Cole, 1999; Zagaglia and Stech, 2004; Bigelow and Maunder, 2007). Stock assessment scientists then use this information to correct observed catch rates, and derive historical population estimates (e.g., Bigelow et al., 2002).

These estimates lead to an understanding of how fishing mortality changes population size, and the mortality is the control variable that management seeks to adjust via gear restrictions, quotas, and spatial management.

The relationships with the environment are often quite weak, because although the selected environmental variables may be correlated with fish distributions and abundances, they are not strong drivers (Myers, 1998). Derived variables such as mixed layer depth, productivity, and prey biomass may be more relevant to the distribution and abundance of fish. Thus, improved “physical” and “biological” variables from coupled models may improve understanding of historical patterns, and offer improved management for ocean resources (e.g., Senina et al., 2008).

The second, and growing, way in which fisheries management can use oceanographic information is in either real-time or forecast modes. These approaches offer the potential for strategic rather than reactive marine resource management. For example, oceanographic variables with a known relationship to the species of interest may be used to forecast distribution of suitable habitat. Access to this habitat can then be regulated by the fisheries agency. Such an approach is used in a longline fishery in eastern Australia, where southern bluefin tuna habitat is predicted every two weeks by combining physical fields derived from the BLUEnet ocean model with tag-based environmental preferences (Hobday and Hartmann, 2006). Figure 3 illustrates how real-time oceanographic information is converted into fishing management maps. Thus, fishery managers can restrict access to locations with predicted high concentrations of tuna to authorized fishers only.

Seasonal forecasts incorporating interannual El Niño Southern Oscillation or North Atlantic Oscillation variability would also find a lot of applications in fisheries and ecosystem management and monitoring, especially since many relationships have been demonstrated between climate indices and fish stock fluctuations (e.g., see a review in Lehodey et al., 2006).

In addition to these well-studied seasonal and interannual variabilities, changes in marine ecosystems occur at time scales of one to several decades in all parts of the world ocean. In the North Atlantic, a warm period from the 1920s to the 1960s caused extensive changes in the distribution and abundance of fish, plankton, benthos, and other biota (Brander, in press). Decadal changes are observed in the North Pacific as well, with significant impacts on fisheries (Chavez et al., 2003; Brander, in press). Because decadal changes affect species composition, distribution, and production, GODAE products can provide three kinds of information of great significance to fisheries and marine ecosystem management: (1) reanalyses that provide integrated reconstructions of past physical, chemical, and biological changes and that help in interpreting the processes responsible for decadal changes, (2) real-time indications of system changes, particularly if these show an increased likelihood of a regime shift in progress, and (3) predictions of extreme events that may cause increased mortality in certain species or areas or give rise to prolonged system changes.

**REGIONAL DOWNSCALING AND COASTAL APPLICATIONS**

Broadening our interest from physics to biology necessarily means increasing the focus of modeling systems to the shelf or even coastal zone. Indeed, continental margins play a key role in the ocean carbon cycle, with about 10–15% of global ocean productivity and more than 40% of the global burial of organic carbon occurring in these regions (Muller-Karger et al., 2005). The major part of fish and shellfish productivity, and thus fishing activity, occurs on and along continental shelves. Hazardous environmental events such as harmful algal blooms, hypoxic events, or eutrophication also take place in the coastal zone, so this oceanic area is highly sensitive to global change.

Meanwhile, regional seas and coastal zones bring new challenges in terms of observing, modeling, and assimilation.
systems. Due to the need to consider additional processes, such as river fluxes, water-column-benthic interactions, atmospheric deposition, and higher biodiversity, coastal ecosystem and biogeochemical models require entirely new submodules that are not necessary for open-ocean applications. Complex coastline and topography, the high frequency of physical processes (external and internal tides) and their coupling to biology, high amplitude, and rapidity of events (successive coastal blooms as compared to temporally well-defined oceanic spring blooms) are good examples of the challenges facing modeling and assimilation methodologies. The case of eel larvae that cross the whole North Atlantic Ocean to eventually enter European estuaries is a good example of the different scales to be considered when focusing on particular species (Bonhommeau et al., 2009).

The spatial scale of coastal processes often requires an increase in model resolution (~ kilometers) to levels at which global models become limited by computing resources. Regional nesting or unstructured grids with higher resolution along the coast and over strong topography gradients may help solve this issue so that biological processes are well constrained by the physics; for example, high-resolution circulation patterns allow for biological material to be transported with higher accuracy (Huret et al., 2007; Gruber et al., 2006). Biological tracers or particles need to be transferred in a conservative way at the chosen boundary between the global and regional models in a two-way mode. For that to occur, the slope region, with its complex exchange processes of biological material, should not represent a boundary for regional applications, but rather it needs to be part of the regional high-resolution model. These coupling issues have important implications for the design of GODAE-based systems and their connection with operational coastal models.
THE FUTURE: REQUIREMENTS AND RESEARCH PRIORITIES

In light of what has been achieved by, and learned during, GODAE, several key areas of progress can be identified for the future. Regarding the physics of GODAE products available today, improvements will be needed to increase their relevance, accuracy, and availability to a variety of users interested in environmental applications. Well-identified strengths of the nowcast, forecast, and reanalysis products delivered by GODAE systems are their high horizontal and vertical resolution of the upper ocean, and their capacity to describe the variability of observed mesoscale features. Another key asset is the capacity of the systems to produce ocean circulation reanalyses of homogeneous quality over multiyear periods. However, a number of applications have shown that coupling biological models to physical hindcasts could reveal fundamental problems in hydrodynamic solutions, and sometimes lead to degraded simulations of biogeochemical variables (e.g., McKinley et al., 2004; Berline et al., 2007). A possible source of difficulty lies in the intermittent nature of the sequential methods used in GODAE systems to update the physical state, causing temporal discontinuities in the model trajectory. When the increments are physically unbalanced, the model states can experience strong convective adjustments that give rise to unrealistic vertical (advective and diffusive) fluxes of properties. Modifications of the assimilation protocol based on the Incremental Analysis Update method (Ourmières et al., 2006) have been implemented in several operational systems to alleviate this effect, but more rigorous approaches such as smoother algorithms will be needed in the future.

The lack of relevant observations is probably the most severe limitation to expansion of operational systems to biogeochemical and ecosystem applications. It will be necessary to adapt current observing systems, develop new instruments, and design studies for optimal observation strategies. As an example, Argo is a key component of the global ocean observing system that has been designed primarily to serve climate research. As a result, the technical specifications adopted for the profiling floats in terms of spatial and temporal sampling are not optimal for biogeochemical and ecosystem monitoring. In the current mode of operation, the profiling floats are programmed to sample the upper layer of the ocean with a vertical resolution of only about 10 m, which is often insufficient to capture the strong biogeochemical gradients observed in the euphotic layer. A first requirement for the future will be to achieve better sampling of surface layers where most of the biological activity takes place. In addition, new instruments will be required for in situ observations of geochemical parameters and ecosystem variables resolved by the implementation of biogeochemical models. Ourmières et al. (2009) recently emphasized the key role of nutrient data in constraining coupled physical-biogeochemical simulations in the North Atlantic using an assimilative platform. To gather these data requires deploying new sensors on profiling floats and gliders that will automatically measure biogeochemical parameters such as O2 and nitrate (Gruber et al., 2007; Johnson et al., 2009). Oxygen, fluorescence, and turbidity sensors are being tested on Norwegian Argo floats in the Nordic seas, providing previously unattainable information on the seasonal development and vertical structure of primary production in the open ocean (Figure 4). Regarding Argo, a key question for the future is whether physical oceanographers are willing to redesign their systems to cater to the interests of biogeochemical users, or whether a separate sampling network including re-instrumented profiling floats will be preferable.

The development of automatic acoustic sampling tools, such as the Mid-trophic Automatic Acoustic Sampler (MAAS), and their large-scale deployment for sampling prey organisms will be another important issue (see the Climate Impacts on Oceanic Top Predators [CLIOTOP] project: http://www.globec.org). With regard to space observations, ocean color missions on satellites with geostationary orbits will permit better temporal coverage, offering the possibility of following episodic events at the scale of hours (e.g., red tides, sediment dispersal) and improving the match between the temporal scale of satellite observations and those of models. Other potential applications include examining the daily cycle of ocean properties, for example, by eliminating the effect of clouds from ocean color measurements and better observation of planetary waves.

Finally, strategies for quantitative evaluation of hindcast, nowcast, and forecast model skill (versus persistence) will be critical to determining the models’ predictive powers. The protocols developed for GODAE products have thus far not considered biogeochemical fields in their analyses. Novel approaches based on advanced statistical methods
open new perspectives for exploring the nature and spatio-temporal distribution of errors in complex ecosystem models (Allen et al., 2008). Continued progress will require using existing models for generating end-user products and for developing efficient observational strategies.

REFERENCES
Figure 4. Development, April to October 2006, of (A) density, (B) Chl a, and (C) AOU (a measure for oxygen production from two Argo floats in the Nordic Seas). Author Svendsen and Kjell Arne Mork, Institute of Marine Research, pers. comm., January 3, 2008
The new Hadley Centre climate model HadGEM1: Evaluation of coupled simulations. Journal of Climate 19:3, 327–1,353. doi: 10.1175/JCLI3712.1