Emerging ocean observations for interdisciplinary data assimilation systems

Tommy D. Dickey*

Ocean Physics Laboratory, University of California, Santa Barbara, 6487 Calle Real, Suite A, Goleta, CA 93117, USA

Received 18 January 2002; accepted 21 October 2002

Abstract

Identification, understanding, and prediction of many interdisciplinary oceanographic processes remain as elusive goals of ocean science. However, new ocean technologies are being effectively used to increase the variety and numbers of sampled variables and thus to fill in the gaps of the time-space continuum of interdisciplinary ocean observations. The formulation, accuracy, and efficacy of data assimilative models are highly dependent upon the quality and quantity of interdisciplinary observational data. In turn, the design of optimal sampling networks will benefit from data assimilative-based observation system simulation experiments (OSSEs). The present contribution, which is directed toward both modelers and observationalists, reviews emerging interdisciplinary observational capabilities and their optimal utilization in data assimilative models.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Observation system simulation experiments; Data assimilation; Sampling networks

1. Introduction

Oceanography began as a highly interdisciplinary science as observations from expeditionary ships were extraordinarily difficult to obtain and few research expeditions could be put to sea. Early oceanographers often utilized data from outside their specific interest areas or subdisciplines, and oceanographic research was generally conducted as a truly interdisciplinary science through roughly the 1940s. In the next few decades, disciplinarily focused oceanographic research led to great advances within the physical, chemical, biological, and geological subdisciplines. The return to prominence of interdisciplinary oceanography may be traced to the late 1970s and early 1980s when several problems (e.g., coastal upwelling, El Niño-Southern Oscillation (ENSO), and others) that clearly required interdisciplinary data sets started to bring back together the subdisciplines and importantly stimulated interdisciplinary sampling, analyses, and modeling. Today, a growing fraction of oceanographic research is interdisciplinary in character (e.g., see example in Fig. 1). The future success of this research depends on the merging of interdisciplinary data and models. This point has been made from diverse viewpoints over the past decade by many oceanographers (e.g., see Dickey, 1991, 2001; Bennett, 1992; Wunsch, 1996; Robinson and Dickey, 1997; Robinson et al., 1998; Doney, 1999; Doney et al., 2001; Hofmann and Friedrichs, 2001; Robinson and Lermusiaux, 2002). A general purpose of this review is thus to stimulate and encourage...
dialogs and cooperative efforts among observationalists and data assimilative modelers.

There are several facets of the present contribution: (1) motivation, fundamental concepts, and goals of data assimilative models and systems are introduced; (2) some examples of interdisciplinary oceanographic problems that have been, or can be, addressed with data assimilative models and systems are summarized; (3) uses and requirements of observations for data assimilative models and systems are reviewed; (4) state-of-the-art and emerging technologies appropriate for obtaining and utilizing interdisciplinary data sets in data assimilative models and systems are discussed; (5) some ideas concerning sampling networks and schemes are outlined; and finally (6) a brief summary of future challenges and opportunities is presented.

### 2. Motivation, concepts, and goals

#### 2.1. Motivation

Research advances and paradigm shifts in oceanography have often been stimulated by observations of various processes; in particular, theories and models have typically been developed to explain, quantify, incorporate, or parameterize these processes using balance equations. Examples of this chronology include Ekman transport, western intensification of boundary currents, and seasonal phytoplankton blooms. Limitations of oceanographic data persist in terms of raw numbers and diversity of variables. This is not surprising considering the spatial scale of the oceanic setting and the time scales of interest, both of which can span over 10 orders of magnitude (Fig. 2a; e.g., Dickey, 1991, 2001), and the complexity of the biology, chemistry, and physics of the oceans. Progress in addressing ocean sampling deficiencies using interdisciplinary, multiplatform sampling is best seen by reviewing Fig. 2a and b together (this topic is developed in Sections 5 and 6). Ocean models have become increasingly useful as new processes have been incorporated or parameterized in formulations, numerical techniques have been improved, and more powerful computing capabilities have allowed increased spatial and temporal resolution and range as well as greater numbers of variables and balance equations. Interestingly, as more data have been collected, analyzed, and

---

**Fig. 1.** A conceptual model illustrating coupling of physical, chemical, and biological processes through optical processes. PENET. RAD. and K represents the penetrative component of solar radiation and the diffuse attenuation coefficient of solar radiation. The schematic portrayal of key biological and chemical state variables (phytoplankton, zooplankton, nitrate, silicate, phosphate, ammonium, trace elements, and detritus) basically follows Friedrichs and Hofmann (2001).
Fig. 2. (a) Time-horizontal space scale diagram illustrating several physical and biological processes in ovals. (b) Rectangles indicate the approximate horizontal and temporal sampling domain capabilities of several platforms at present.
interpreted, and as computer model simulations have become more realistic, observationalists and modelers have become more cognizant of and dependent upon each other for making scientific advances. The culmination of this cultural change is epitomized by two methodologies: inverse methods and data assimilation.

2.2. Concepts

The “ocean circulation inverse problem” has been defined by Wunsch (1996) as a problem of inferring the ocean circulation, understanding it dynamically, and forecasting it by quantitatively combining theory and observations. He further defines “inverse methods,” more strictly relevant to the topic of his book on ocean circulation, as techniques used for solving problems or systems of equations that are mathematically underdetermined. Most physical, and certainly all interdisciplinary, oceanographic problems fall into this “underdetermined” category. Further, we are often seriously limited in our studies by undersampling and aliasing. Hence, there is growing interest in application of inverse methods to many oceanographic problems. Inverse methods are often used in the context of data assimilative models as indicated below.

Several introductions to and reviews of data assimilative methodologies for geophysical, physical oceanographic, and more recently interdisciplinary oceanographic problems have been presented during the past decade (e.g., Ghil and Malanotte-Rizzoli, 1991; Robinson and Dickey, 1997; Robinson et al., 1998; Kasibhatla et al., 2000; Robinson and Lermusiaux, 2001, 2002; Hofmann and Friedrichs, 2001). The term data assimilation has been described in a variety of ways, perhaps most succinctly by Hofmann and Friedrichs (2001) as “the systematic use of data to constrain a mathematical model.” Important points from their introduction to data assimilation include: (1) it is assumed that dynamics responsible for processes and distributions of properties are inherent in the data; (2) by inputting data into a specific model, the model will produce a more accurate model simulation; and (3) hindcasts, nowcasts, and forecasts are improved through data assimilation. Here, we consider data assimilation to broadly comprise parameter estimation and setting/resetting of state variables. It is worth noting that atmospheric data assimilation, which is relatively well developed, has several similarities to oceanographic data assimilation (e.g., Ghil and Malanotte-Rizzoli, 1991). This is not surprising considering the many analogies of these two geophysical fluid media.

A brief description of the mathematical formulation of the biological (chemical) data assimilation problem follows that of Robinson and Lermusiaux (2002). The dynamical equations (time-averaged) for \( n \) biological (and/or chemical) state variable fields, \( \phi_i(r, t) \) can be represented as

\[
\frac{d\phi_i}{dt} = \nabla \cdot (K_i \nabla \phi_i) + B_i(\phi_i) \quad \text{with } i = 1 \text{ to } n
\]  

where \( \mathbf{r} \) is the three-dimensional field position vector and \( t \) is time. The total time derivative on the left-hand side includes local and advective effects. Mean (time-averaged) equations are typically formed using Eulerian representations, and fluctuating (turbulent) portions are used to compute biological or chemical “Reynolds stress terms.” These are usually represented with turbulent diffusivities, \( K_i \), and the parameterizations are either assigned or determined using turbulence closure model calculations. The first term on the right-hand side of Eq. (1) is the turbulent diffusion term. The second term on the right-hand side is the reaction term \( B_i(\phi_i) \) that represents the biological (chemical) dynamics and thus sources and sinks of \( \phi_i \) that involve processes such as reproduction, mortality, predation, behavior, and various chemical reactions. These are complex nonlinear processes that are generally difficult to formulate or parameterize. In the model examples considered for this review, we refer to state variables for models that are equivalent to biological (chemical) compartments or components. For example, a simple nutrient–phytoplankton–zooplankton (or NPZ) model has three compartments or components with a set of three coupled equations \((n = 3 \text{ in Eq. (1)})\).

An important step in data assimilation involves the estimation of error related to both the model and the data derived from measurements. To include model error, Eq. (1) can be rewritten in the form

\[
\frac{d\phi_i}{dt} = M_i(\phi_i, P_i) + \eta_i \quad \text{with } i = 1 \text{ to } n
\]  

where \( M_j \) represents a model operator dependent on advection, diffusion, biology, and chemistry, \( P_j \) is the ensemble of model parameters, and \( \eta_i \) is a stochastic or error forcing. The state variables, \( \phi_i \), can be related to measurement data, \( y_j \), through an observational functional or measurement model, \( H_j \), as

\[
y_j = H_j(\phi_1, \ldots, \phi_i, \ldots, \phi_n, P_1, \ldots, P_i, \ldots, P_n) + \epsilon_j
\]

where \( \epsilon_j \) is a stochastic or error forcing. In addition, a sensor data model (Lermusiaux, personal communication) can be formulated as

\[
s_j = S_j(y_j, P_j) + \zeta_j
\]

where \( S_j \) is a functional, \( P_j \) is the ensemble of model parameters, and \( \zeta_j \) is a stochastic or error forcing. The assimilation (melding) criterion is formulated as a minimization problem that can be written in terms of a function \( J \) (called cost, penalty, or objective function)

\[
\min_{\phi_i, P_i} J(\eta_i, \epsilon_j, \zeta_j, q_{\eta}, q_{\epsilon}, q_{\zeta})
\]

where \( q_{\eta}, q_{\epsilon}, \) and \( q_{\zeta} \) are weighting factors. Robinson and Lermusiaux (2002) discuss the mathematical methods for solving this problem using Eqs. (2)–(5). Open questions concern representation, attribution, and propagation of errors.

To summarize, errors are attributable to (1) the dynamical model (e.g., poor representation of processes, omissions or incorrect formulations of key processes, and computational or procedural problems) and (2) observations or measurements (e.g., sensor design, performance, noise, sample averaging, aliasing, etc.). The main point is that estimation of observational or measurement error is essential for application of data sets for data assimilative modeling and specifically for quantifying the predictive skill of models. Further, adaptive sampling can be based on error estimation. For example, sampling platforms can be directed to areas of higher error estimates [as described later using results from the Littoral Ocean Observing and Prediction System (LOOPS) program]. It is beyond the scope of this review to ascribe error for the sensors and systems described here; further, different specific sensors and systems have different resolution, accuracy, and noise characteristics. It is also worth noting that marked improvement in sensors and systems are expected to result in reduced error factors that will be input into data assimilative models and will modify cost functions.

The application of data assimilation methods to interdisciplinary oceanographic problems is nontrivial (e.g., Ishizaka, 1993; Lozano et al., 1996; Robinson et al., 1998; Kasibhatla et al., 2000; Hofmann and Friedrichs, 2001; Robinson and Lermusiaux, 2001, 2002). Like the strictly physical case, broad ranges of time and space scales need resolution to depict relevant processes for interdisciplinary problems. Additional considerations are relevant for the interdisciplinary case. These include the following: (1) the number of important interdisciplinary processes and state variables is several times if not orders of magnitudes greater; (2) biology involves complex physiological and behavioral effects that are either unknown or difficult to model or parameterize presently; (3) biological, chemical, and optical processes are often highly related: biological and chemical processes involve complex transformations (e.g., organic to inorganic forms of matter and vice versa), and require coupled sets of nonlinear equation; (4) several biological, optical, and chemical processes have short time scales (minutes to days); and (5) rates are especially difficult to determine. Because of these many issues, challenging as they may be, interdisciplinary data assimilation is especially important for advancing interdisciplinary oceanography.

One depiction of a “data assimilative system” is shown in Fig. 3 (following Robinson and Lermusiaux, 2001). The system is composed of three fundamental elements: (1) an observational network with data telemetry capability; (2) an interdisciplinary model; and (3) a data assimilation scheme. Schematics illustrating open ocean and coastal observing networks are shown in Figs. 4 and 5, respectively. Interdisciplinary data are collected and transmitted in real time or near real time from various platforms of the sampling network or array. Each platform has its own intrinsic capabilities and limitations in terms of types of sensors and systems that can be accommodated and define spatial and temporal ranges of measurements (Figs. 2a and b and 3). Data and error estimates are transferred for incorporation into an interdisciplinary model. The interdisciplinary dynamical model is composed of several modules, e.g., for physics, biology, and chemistry. Each module is composed of a set of equations with...
variables appropriate for describing the relevant processes. Importantly, these modules are coupled. The data-melding step produces state variable and parameter estimates of observed physical variables (OSV$_i$), observed biological variables (OSV$_j$), observed chemical variables (OSV$_k$), and unobserved variables (USVs) along with associated errors. This information feeds back to the data assimilative model for inclusion in the next model iteration and for model improvement and back to the sampling network enabling adaptive sampling (e.g., based on error simulation as discussed earlier). Adaptive sampling has several possible modes: changing sampling rates, changing gains of sensors and systems, and redirection of mobile sampling assets to locations of special informational value (e.g., fronts, upwelling sites, and regions where processes are leading to large model error estimates).

2.3. Goals

Important goals of the data assimilation procedure are: (1) to develop and improve model formulations...
and functional dependencies; (2) to estimate parameters and variables as well as rates that cannot be measured; (3) to estimate or predict state variables and parameters on time and space scales inaccessible to some or all of the observational sampling assets (e.g., a smart interpolating function); (4) to provide model initializations; (5) to identify and elucidate ocean processes; (6) to design experimental or operational sampling networks using observational system simulation experiments (OSSEs); (7) to optimize use of sampling assets through adaptive sampling; and (8) to make predictions for research and operational environmental management including uncertainty and error estimates. A general goal is to accurately predict the distributions of selected state variables given their values at some initial time \( t_0 \). A limiting problem concerns the quality (accuracy and resolution) and, perhaps more importantly, the quantity of the observational data at \( t = t_0 \). Thus, the available initial data fields need to be interpolated and gridded before the governing equations can be solved (integrated). It is also important to emphasize that the data collected from sampling networks can be effectively used to test, verify, and validate different data assimilation methods and schemes. Examples of evaluating different assimilation methods for biogeochemical

---

Fig. 4. Conceptual illustration of several sampling platforms for the open ocean. Most of these platforms have been used off the island of Bermuda.
data using “identical twin experiments” are presented in Hofmann and Friedrichs (2001) and other references in the present review.

To fulfill the goals of the data assimilation methodology, each of the components of the system described above (Fig. 3) must be optimized. Simply put, we need the most appropriate and accurate data sets to input into well-designed models that include formulations for the key processes; data and models must then be linked together through the best possible data assimilation scheme. It is beyond the scope of the present paper to review two of the system elements: interdisciplinary models and data assimilation methodologies. Readers interested in these topics are directed to recent reviews of interdisciplinary models by Hofmann and Lascara (1998), Nihoul and Djenidi (1998), Doney et al. (2001), Ducklow et al. (2001), and references therein and reviews of interdisciplinary data assimilation models and methods by Lozano et al. (1996), Robinson et al. (1998), Kasibhatla et al. (2000), Hofmann and Friedrichs (2001, 2002), Robinson and Lermusiaux (2001, 2002), and references provided in these papers.

Fig. 5. Depiction of sampling platforms used during the HyCODE study off the New Jersey coast at the LEO-15 site in the summers of 2000 and 2001. Figure provided by Scott Glenn, Oscar Schofield, and Michael Crowley.
3. Examples of interdisciplinary oceanographic data assimilation studies

Interdisciplinary oceanographic data assimilative modeling is considerably newer than its atmospheric and physical counterparts. Nonetheless, some important work has already begun, thanks in part to the increasing availability of interdisciplinary data and improved knowledge and models of biological and chemical processes resulting from process and time series studies (e.g., see Robinson and Lermusiaux, 2002; Hoffmann and Friedrichs, 2002; Dickey, 2001, 2002; Dickey and Falkowski, 2002). The following subsections summarize a few examples of interdisciplinary data assimilation studies. The focus here is on ecosystems, phytoplankton, and biogeochemistry (e.g., Fig. 1). A few examples of other interdisciplinary applications are provided as well. For the purposes of this review, specific emphasis is placed on data sets that have been used for these various studies and on those data sets that have been recommended for future collection.

3.1. Ecosystems, phytoplankton, and biogeochemistry

3.1.1. North Atlantic Ocean: open ocean

An important aspect of data assimilative modeling involves model initialization with actual data. A good example of this step is illustrated by McGillicuddy et al. (1995a,b), who performed model simulations of a coupled physical and biological system under spring bloom conditions during the Joint Global Ocean Flux Study (JGOFS) North Atlantic Bloom Experiment (NABE) in 1989. In particular, they used sea surface height (satellite altimetry data; GEOSAT) and expendable bathythermograph (XBT) temperature profile data to initialize their model. Nowcasts and hindcasts were made using these and biogeochemical data sets as input to a quasi-geostrophic physical model coupled to a surface boundary layer, biological model. The ecological model used four compartments or state variables (nitrate, ammonium, phytoplankton, and zooplankton). The nowcasts were utilized to direct ship sampling (Robinson et al., 1993), while hindcasts were valuable for elucidating processes (McGillicuddy et al., 1995a,b). This activity effectively served as a precursor for later focused work suggesting the importance of mesoscale eddies for nitrate fluxes, primary production, and new production (e.g., McGillicuddy et al., 1998a; Oschlies and Garcon, 1998; McNeil et al., 1999; Letelier et al., 2000; Dickey and Falkowski, 2002).

It should also be noted that a nonlinear optimization method was applied to NABE (47°N, 20°W) data sets by Fasham and Evans (1995), who used a seven-component model of the mixed layer ecosystem and found that a best fit of data for primary production resulted in a poor fit of zooplankton data and vice versa. Their findings led them to suggest that some important processes affecting nitrogen flows (e.g., microzooplankton biomass and bacterial production) need to be more reliably estimated.

Data assimilation studies were used by Lawson et al. (1996) to examine the ability of adjoint methods to recover rate parameters including population growth and death rates as well as component initial conditions and responses to episodic events. In addition, examinations of results based on different sampling intervals were done. The latter aspect was motivated by the practical issue of limitations of ship-based sampling at two sites of biogeochemical time series studies [JGOFS Bermuda Atlantic Time-series (BATS): 31°50′N, 64°10′W, and Hawaii Ocean Time-series (HOT) sites; see review by Karl et al., 2001]. The time-dependent ecosystem model (so-called zero-dimensional in space) of Lawson et al. (1996) included five interactive components consisting of nitrate, ammonium, phytoplankton, zooplankton, and detritus (e.g., see Fasham et al., 1990). The currency of the model was nitrogen. Neither spatial variability nor coupled physics was incorporated in this particular model, primarily in the interest of simplicity. The number of model compartments was limited for similar reasons. Nitrate was input over 2-day “event” periods four separate times. The first event began 4 days after the simulation’s initiation; others followed at 18-, 5-, and 10-day intervals. Data were assimilated at monthly, biweekly (every 2 weeks), and weekly intervals for comparative purposes. The results of this work indicated that biweekly assimilation of data was sufficient to recover rate parameters quite well. However, it was found that lack of information for zooplankton and ammonium at monthly sampling intervals hindered recovery of rate parameters. The authors noted that sampling at frequencies incapable of resolving the controlling processes leads to data sets that are inad-
equate for data assimilation. These two points (i.e., sampling of rate parameters and undersampling) argue strongly for measurements of critical rate parameters and data with temporal resolution adequate to capture key processes (e.g., eddies, wind and dust events, etc.). These points are consistent with reports by observational studies examining these issues using biogeochemical time series and spatial data sets (e.g., Dickey, 1991, 2001; Wiggert et al., 1993; Robinson et al., 1993; Robinson and LOOPS Group, 1999). This sampling issue is nicely illustrated by considering time series data collected from the Bermuda Testbed Mooring (BTM; Dickey et al., 2001) as indicated in Fig. 6. The seasonal cycles are evident in the upper and lower panels of Fig. 6. High-frequency variability and episodic events were also observed. Examples are highlighted in the expanded middle panels. Shown from left to right are (1) the passage of an eddy with high nutrient and chlorophyll concentrations (e.g., McNeil et al., 1999), (2) passage of Hurricane Felix over the BTM site accompanied by inertial pumping and rapid mixed layer deepening (Dickey et al., 1998d), and (3) passage of a warm mesoscale feature with high chlorophyll concentrations and sediment flux to depth (Dickey et al., 2001; Conte et al., 2003). These data illustrate the importance of high-frequency sampling and modeling with time steps of hours or less for many problems. Further, the major mesoscale events had little or no discernible surface expression for satellite sensors.

The work of Lawson et al. (1995) was followed by another data assimilation effort by Spitz et al. (1998), who used BATS data for their study. The two primary objectives were: (1) to determine if monthly sampling by the BATS program is sufficient to estimate all ecosystem parameters and (2) to assimilate BATS data over the period 1988–1993. They chose the variational adjoint method, as it was relatively easy to implement. The seven-component ecosystem model of Fasham et al. (1990) was utilized. The mixed layer’s seasonal cycle was prescribed (based on data from the atlas of Levitus, 1982). A simple mixed layer entrainment parameterization was included. BATS data used for assimilation included nitrate, chlorophyll, particulate organic nitrogen concentrations, bacterial cell counts, phytoplankton primary production rates, and bacterial production rates. The model used for the study simulated the ecosystem within the homogeneous, but seasonally varying mixed layer. Model parameters were based on those used by Fasham et al. (1990), who calibrated their model using data from the nearby Hydrostation S site for the period 1958–1960. The seasonal cycle of key variables of the ecosystem appeared to be well represented. However, additional observations were needed for parameter recovery. Further, it was concluded from results of assimilation of actual BATS observations that some of the assumptions of the Fasham et al. (1990) model needed to be reconsidered and the model be modified; this is consistent with the findings of Fasham and Evans (1995).

The problem of insufficient interdisciplinary time series data sets was pointed out by Spitz et al. (1998). They noted that their use of 5 years of observations (essentially using monthly averages) to create the 1-year data set for the assimilation process is problematic because of interannual variability. This is likely important with respect to the timing and intensities of spring blooms and other interannually varying phenomena at mid-latitude sites like the Sargasso Sea (e.g., discussed by Bates, 2001 and Dickey et al., 2001 for the BATS site). Recently, Oschlies (2001a) examined the correlation between the North Atlantic Oscillation (NAO) index nutrient supply near Bermuda using numerical experiments with a nitrogen-based, four-component (nitrate, phytoplankton, zooplankton, and detritus) biological model embedded in a general circulation model (2/5 x 1/3° resolution) that incorporated a turbulence closure scheme for simulating the mixed layer on a seasonal cycle. High- and low-NAO index conditions were used for intercomparison. Oschlies (2001a) found that nutrient supply decreased by about 30% near Bermuda during the high-NAO state (late 1980s to early 1990s). This is significant as Bermuda hydrographic records show a major shift from relatively deep mixing (200–400 m) in the 1960s (low NAO) to shallower mixed layers (reduced seasonal supply of nutrients from depth) in the late 1980s and early 1990s (high NAO; coincident period of BATS observations). Because of limitations of parameterization data for the forcing period of interest, Spitz et al. (1998) necessarily used parameters based on low-NAO state conditions (experimental runs were used with high-NAO forcing conditions). Furthermore, the problems of simulating event-scale processes (e.g., eddies and synoptic scale weather patterns) are also relevant for this study. These
Fig. 6. Time series of temperature for several depths obtained from the Bermuda Testbed Mooring (BTM; see Dickey et al., 2001). Seasonal cycles are evident in the upper and lower panels. High-frequency variability and episodic events were also observed. Examples of eddy, mesoscale feature, and hurricane passages are highlighted in the expanded middle panels and are discussed in the text.
latter remarks support the need for continuing cooperative efforts of physical and biogeochemical modelers and observationalists to develop improved sampling strategies and methodologies and argue for long-term, high-frequency observations.

A coupled, eddy-permitting, biological–physical (four components: nitrogen, phytoplankton, zooplankton, and detritus) model of the North Atlantic has been used in assimilating satellite ocean color data by Gunson et al. (1999). In their scheme, they “seeded” several “model floats” in different areas of the model domain, followed their trajectories, and applied a one-dimensional biological model that incorporated vertical advection and diffusion. They suggested that by concurrently assimilating ocean color data (with initialization of nutrient, phytoplankton, zooplankton, and detritus profiles) in different biological provinces, it might be possible to constrain all ecosystem parameters.

A specialized data assimilation study using the Harvard Ocean Prediction System (HOPS) model (primitive equation physical model coupled with a five-compartment biological model) was described by Anderson and Robinson (2001; particularly, note model and data assimilation schematics and results in their Figs. 12.16–12.18). They focused on the Gulf Stream meander and ring region. Briefly, feature models and data assimilation methods were used to (1) resolve specific submesoscale and mesoscale dynamical processes, (2) evaluate physical and biological processes causing a high concentration of phytoplankton biomass at the Gulf Stream front, and (3) relate wind-forcing and Gulf Stream and ring-related processes to vertical velocity, nitrate transport, phytoplankton and zooplankton patchiness, primary and new production, particle export, and cross-stream exchanges of nitrate and phytoplankton. Data sets collected in 1988 were used for the data assimilation exercise and included temperature profiles from 320 XBTs, 216 CTD casts (half with chlorophyll), 32 nutrient profiles, and some zooplankton data. These data were collected in the region $38.0 \pm 1.0^\circ N$, $70.0 \pm 2.5^\circ W$ and were supplemented with satellite sea surface temperature and altimetry (sea surface height) data sets in a roughly $600 \times 1000$ km region enveloping the ship-based sampling area. One of the major conclusions of this study was that in late summer, submesoscale ring–stream interactions, opposed to mesoscale meandering, were primarily responsible for generating phytoplankton maxima and patches at the Gulf Stream front and for increasing cross-frontal exchanges. The work also showed the need for and importance of dynamically compatible physical and biological data fields.

3.1.2. North Atlantic Ocean: coastal

The interdisciplinary LOOPS field program was conducted in Massachusetts Bay during the period of August 17–October 5, 1998 (Robinson and the LOOPS Group, 1999; Besiktepe et al., 2003; Robinson and Lermusiaux, 2002). LOOPS utilized multiplatform sampling and interdisciplinary data assimilative modeling components to make real-time forecasts of biophysical variables. The LOOPS system concept can be described as generic, versatile, and portable with application to interdisciplinary, multiscale problems in the coastal zone. Scientific foci of LOOPS have concerned phytoplankton and zooplankton distributions, their evolution, and their patchiness, in relation to physical processes. The biogeochemical and ecological model equations were solved for six state variables: nitrate, ammonium, chlorophyll, phytoplankton biomass, zooplankton, and detritus. Nested domains (see Fig. 12.20 of Robinson and Lermusiaux, 2002) included the following features: western Gulf Stream meander and ring region, Gulf of Maine, and Massachusetts Bay. Real-time interdisciplinary forecasts were made using the Harvard Ocean Prediction System (HOPS, Robinson, 1999; Fig. 12.19 of Robinson and Lermusiaux, 2002). The data assimilation methods included optimal interpolation along with an error subspace statistical estimation (ESSE; Lermusiaux, 1999a,b, 2002; Lermusiaux and Robinson, 1999; Robinson and Lermusiaux, 2002). Observational assets used for, and in some cases redirected by (adaptive sampling mode), the forecasts included ships, buoys, satellites, and autonomous underwater vehicles (AUVs; see Fig. 7; Yu et al., 2002). The circulation patterns of Massachusetts Bay are complex as evidenced by a Cape Cod Bay gyre that can be cyclonic, anticyclonic, or nonexistent (see depiction in Fig. 8). They are also highly variable in space and time because of topographic features, coastal geometry, energetic tides (causing internal solitary waves), wind events, changes in buoyancy currents, occurrences of fronts, and submesoscale eddies.
Fig. 7. Top: Odyssey autonomous underwater vehicle (AUV) as it was configured during the 1998 LOOPS experiment in Massachusetts Bay (see Yu et al., 2002). Lower panels show horizontal–vertical sections obtained with the AUV for the following properties: (a) temperature, (b) salinity, (c) chlorophyll fluorescence (subsurface maximum and patchiness are evident), and (d) optical backscatter (880 nm) (subsurface particle maxima and resuspended bottom sediment are seen).
Some of the first interdisciplinary data collected with an AUV in coastal waters are shown in Fig. 7 (see Yu et al., 2002 for details). The Odyssey AUV sections shown in Fig. 7 were obtained during the 1998 LOOPS study in Massachusetts Bay as the AUV moved along a track of about 6 km in length, changing its depth in time, and creating a sawtooth pattern. Contoured along-track distance/depth data are shown for temperature, salinity, chlorophyll fluorescence, and backscattering of light at 880 nm.
(related to particle concentrations). Major features evident in Fig. 7 include (1) an essentially two-layer system based on temperature, salinity, and density, (2) a subsurface chlorophyll and particle layer lying near the pycnocline with considerable horizontal scale patchiness, and (3) resuspended sediment in the bottom boundary layer. For the LOOPS real-time modeling exercise, physical data were assimilated while the coupled biological model was initialized and run forward. Ship sampling patterns are overlain on model forecast surface temperature error standard deviation (shown in gray scale) in Fig. 8. The shorter track pattern to the south was selected on the basis of the relatively large error in the vicinity and represents a good example of directed adaptive sampling. Real-time modeling fields for currents and chlorophyll at 10 m and a vertical section of zooplankton at the entrance of Massachusetts Bay are shown in Fig. 8. Patchiness on several scales is evident both in the horizontal and vertical. The features of higher concentrations of phytoplankton and zooplankton northeast of Cape Ann and near Boston Harbor appear to be caused by advection of nutrients from Stellwagen Bank (suspended through bottom mixing) and from along the coast with sources attributed to upwelling and episodic wind mixing events (Robinson and Lermusiaux, 2002). Observed chlorophyll at depth was in reasonable agreement with forecast values.

Other interdisciplinary field programs in the coastal zone have been conducted recently off the New Jersey coast (LEO-15 site; see Glenn et al., 2000a,b) and on the west Florida shelf (Bissett et al., 2001) as part of the Hyperspectral Coastal Dynamics Experiment (HyCODE) program. These experiments have also included multiplatform sampling (see Fig. 5) and interdisciplinary data assimilative modeling components. For the LEO-15 site, physical data (including sea surface temperature, hydrographic profiles, and high-frequency radar surface currents) have been assimilated in real time with forecasts made every 12, 24, or 48 h during summer experiments in years 2000 and 2001 (Glenn, personal communication). In addition, a complex, multicomponent, bio-optical, ecological model has been run in hindcast mode, but without biological data assimilation at this point in time (see Bissett et al., 1999a,b, 2001).

3.1.3. North Pacific Ocean

Matear (1995) utilized interdisciplinary data from the Ocean Weather Station P (OWS P; 50°N, 145°W) time series measurement program in the subarctic Pacific. This work was driven in part by the problem that ecosystem models need to specify a large number of biological and chemical parameters that are not constant and sometimes interdependent. An inverse method, simulated annealing optimization, was used to optimize parameters for three different ecosystem models. This methodology is intended to objectively generate a set of ecosystem model parameters that are most consistent with available information. An additional attribute of such an optimization technique is that correlations of model parameters near the optimum can be quantified and possibly used to suggest proxies for variables or parameters that are presently more difficult or impossible to observe or determine. OWS P data sets included mixed layer depth and temperature, surface radiation, nitrate, phytoplankton, mesozooplankton, and primary production. Some data sets were obtained from daily and monthly sampling, while others were available only seasonally (see Table 2 in Matear, 1995). Seasonal variations in mixed layer depth were based on average values observed between 1970 and 1980, while monthly mean total solar radiation for the period 1959–1975 was interpolated to daily values with peak noon values based on a triangular function. The mixed layer portion of the water column was considered in the one-dimensional model. Matear (1995) found that a simple nitrate, phosphate, zooplankton (three-component) model (e.g., based on Evans and Parslow, 1985) was sufficient to successfully reproduce the available observations (see Fig. 12.6 and 12.7 in review by Robinson and Lermusiaux, 2002).

Other model simulations by Matear (1995) included microzooplankton and mesozooplankton (four-component model; see Frost, 1987) and a configuration for the microbial loop (seven-component model; see Fasham et al., 1990). Despite evidence suggesting the importance of including microbial loop processes (e.g., Legendre and LeFevre, 1995), Matear (1995) concluded that use of more complicated models (e.g., inclusion of additional components such as the microbial loop) was not justifiable until additional measurements (e.g., ammonium, bacteria) are obtained. This assertion can likely be generalized in view of recent work related to trace nutrients (e.g., iron) and dissolved...
organic and inorganic matter including the elements nitrogen and carbon (e.g., Hansell and Carlson, 2001). Interestingly, it was found that large numbers of parameters were correlated, with less than half being independent. This is an encouraging result in that it may be possible to reduce the number of measured variables and/or models can possibly be simplified via parameter or structure aggregations.

The results of Matear (1995) appear to be quite consistent with those obtained in the data assimilation study by Prunet et al. (1996a,b), who also used OWS P data. Prunet et al. (1996a) first used a one-dimensional 10-compartment trophic system model, which was constrained to reproduce seasonal surface chlorophyll concentrations through parameter adjustment. They also used a new method for transforming satellite ocean color data (CZCS data from the year 1976) to estimate primary production and carbon fluxes (also see Ishizaka, 1990 and Abbott, 1992 concerning similar studies). Prunet et al. (1996a) hypothesized that a simpler biological model should suffice. The next study by Prunet et al. (1996b) utilized a one-dimensional physical–biogeochemical model. In this work, a turbulence closure mixed layer model was coupled with a four-compartment nitrogen, zooplankton, phytoplankton, detritus model. Assimilated OWS P data included chlorophyll, nitrate, temperature, and partial pressure of carbon dioxide. It was concluded that (1) a relatively simple biological model should be sufficient to describe carbon fluxes, (2) assimilation of satellite or climatological data could be used to adjust parameters for a simple model on a regional basis, and (3) regional parameterizations could be utilized in three-dimensional (spatial) global models. The last two points relate to the concept of “biogeographical provinces” (e.g., Longhurst, 1998), which will be discussed later in the context of observational networks.

In other work in the sub-tropical North Pacific, Christian and Karl (1994) utilized a least squares inverse method with a simple linear model to estimate biogeochemical indicator ratios and to elucidate microbial community structure (e.g., Legendre and LeFebvre, 1995) at the JGOFS Hawaii Ocean Time-series (HOT) site (22°45' N, 158°00' W).

3.1.4. Equatorial Pacific Ocean

Friedrichs (2001a) selected the equatorial Pacific for her biogeochemical data assimilation study. The equatorial Pacific is vast in scale; however, it has been sampled relatively intensively using the Tropical Atmosphere Ocean mooring array (e.g., McPhaden et al., 2001), JGOFS ship-based process studies (e.g., Murray, 1995), and remote sensing (e.g., temperature, sea surface height, winds, and color). Importantly, the region is highly dynamic with the presence of tropical instability waves, Kelvin waves, and Rossby waves (collectively called equatorial longwaves) as well as ENSO phase changes (e.g., both El Niño and “normal” conditions occurred during the JGOFS field experiment, 1991–1993; see Foley et al., 1997). A recent discovery is that physical processes (e.g., tropical instability waves) are reflected in the ecosystem and phytoplankton as evidenced by interdisciplinary mooring and satellite ocean color data sets (e.g., Foley et al., 1997; Chavez et al., 1998, 1999). For her studies, Friedrichs (2001a) utilized a five-component (phytoplankton, zooplankton, ammonium, nitrate, and detritus) ecosystem model (see right-hand side of Fig. 1) that had been calibrated for the central equatorial Pacific (Friedrichs and Hofmann, 2001). The assimilation scheme used for the work was the variational adjoint technique. Assimilated data were obtained from JGOFS cruises and the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) satellite-borne color imager (e.g., Yoder et al., 2001). Depth dependence was included in the time-dependent, one-spatial-dimensional model, and depth-averaged values were computed based on integrations from the surface to the euphotic layer depth. Model-data misfits were reduced by optimizing parameters including phytoplankton growth rates, zooplankton grazing rates, phytoplankton and zooplankton mortality rates, and recycling of nutrients from detritus remineralization. Friedrichs (2001a) found that similar parameter sets were retrieved from independent simulations, namely, for the 1991–1992 El Niño and the 1997–1998 El Niño. However, unrealistic simulations were obtained for cases when tropical instability waves passed (increased abundance of diatoms) and brief periods of macro-nutrient limitation. An important conclusion of the study is that appropriate model dynamics and structure are essential and that data assimilation is valuable for redirecting and guiding future model formulations. In this case, new model structure would account for shifts in species composition and in primary productivity rate control by macro- and micronutrients (e.g., iron limi-
3.1.5. Mediterranean Sea

A final example of data assimilation methods for ecosystem, phytoplankton, and biogeochemical studies is taken from an ongoing study of seasonal and interannual variability of the physics and ecosystem of the Mediterranean Sea, the Mediterranean Forecasting System Pilot Project (MFSPP; Pinardi et al., 1997, 1999, 2003). The data assimilation system for the Mediterranean study has included the following components: a hierarchy of ocean general circulation models (e.g., based on the Modular Ocean Model, MOM, and Princeton Ocean Model, POM; horizontal resolutions of $1/4^\circ$ and $1/8^\circ$ with 31 vertical levels), ecosystem models of varying complexity, data assimilation techniques, in situ observational data sets (e.g., expendable bathythermographs [XBTs using Voluntary Observing Ships (VOS)]) along with satellite sea surface temperature and altimetry (sea surface height) data obtained from TOPEX and ERS-1 satellite altimeters (see Demirov et al., 2003). The study has elucidated many complex circulation and flow patterns on seasonal and synoptic scales. Wind and heat flux variabilities were shown to be important as well as mesoscale and eddy-mean flow interactions. Interannually, current reversals were reported. Water formation (simulated for realistic forcing conditions) and dispersal were demonstrated to be related to eddies and mixing processes. An aggregated ecosystem model of the Mediterranean Sea was used to simulate spatial gradients and temporal changes in surface chlorophyll at the basin-scale. The model further allowed exploration of the roles played by physical forcing on distributions of nutrients and demonstrated the importance of food web structure in determining primary production levels in different Mediterranean subregions.

The coupled ecosystem–physical model of the Mediterranean Sea also simulated coastal to open sea primary productivity. Different food web structures were simulated for three portions (northern, middle, and southern) of the Adriatic Sea and gave evidence of variable Redfield ratios and thus differing levels of phosphorus limitation. Recent MFSPP interdisciplinary data assimilation work has capitalized on data sets obtained from an interdisciplinary mooring array (M3A) off Crete in the Cretan Sea (e.g., Nittis et al., 2003). The mooring collected three hourly measurements of temperature, chlorophyll $a$, nitrate, and oxygen. Triantafyllou et al. (2003) utilized these data to assess the performance and predictability of a one-dimensional coupled physical–biogeochemical model (Allen et al., 1998). The model was tuned and validated for the Cretan Sea, and a set of optimized parameters was determined. The model was run in hindcast mode, and results were in reasonably good agreement with the mooring data. Although phosphorus is the primary limiting nutrient for this ecosystem, it was found that nitrate was quite valuable and the authors suggest that high temporal resolution phosphorus data sets would be especially useful. They also indicate that parameterization of physiological adaptations for light would improve simulations.

Allen et al. (2003) used the M3A data sets with the European Regional Seas Ecosystem Model (ERSEM; Baretta et al., 1995) and an Ensemble Kalman Filter (EnKF) data assimilation method in order to explore estimation efficacy. It was found that this approach resulted in considerable improvement in the capability to hindcast chlorophyll in particular. Interestingly, the authors indicate that nitrate may be a better overall constraint on the ecosystem dynamics than chlorophyll (again, note that phosphate opposed to nitrate is the primary limiting nutrient here). It was concluded that the predictability window of the EnKF method is at least 2 days and that selection error variances affected results. In related work, Hoteit et al. (2003) also used the one-dimensional ERSEM along with the Princeton Ocean Model. A singular evolutive extended Kalman (SEEK) filter (error statistics were parameterized using empirical orthogonal functions, EOFs) was used to assimilate in situ M3A data. The results showed that the SEEK filter was effective and that input of M3A oxygen and nitrate data improved model simulations of chlorophyll time series in the form of decreased estimation error with respect to actual data and improvement of model behavior. The collective works of Allen et al. (2003), Triantafyllou et al. (2003), and Hoteit et al. (2003) highlight the value of interdisciplinary mooring data for developing interdisciplinary data assimilation methodologies. Clearly, the lessons learned in devel-
oping the integration of physical, ecological, and biogeochemical models with data assimilation systems on the scale of the Mediterranean will be valuable for analogous work in larger ocean basins.

3.2. Other interdisciplinary problems

A long-standing problem of biological oceanography concerns predator–prey interactions. Lawson et al. (1995) first utilized data assimilation (adjoint method) to study this problem. They demonstrated the advantages of this approach using a simple predator–prey model. In particular, they were able to recover model initial conditions and to optimize parameters that are most difficult to measure: growth and death rates.

Another important biological problem concerns physical–biological interactions and higher trophic level organisms and their roles in ecosystem dynamics as affected by ocean physics. McGillicuddy et al. (1998b) have examined the mechanisms controlling seasonal variations in the abundance of the calanoid copepod *Pseudocalanus* spp. in the region of the Gulf of Maine and Georges Bank off the northeast coast of the United States as part of the U.S. Global Ocean Ecosystems Dynamics (GLOBEC) program. A two-dimensional (vertically integrated) advection–diffusion–reaction equation for the copepod concentrations was solved. The velocity and diffusivity were based on seasonal hydrodynamical simulations of climatological flows. They used the adjoint data assimilation method in a diagnostic mode to invert for observed changes in copepod abundances and specified transports. The physical circulation model scheme utilized a finite element approach, thus allowing for a relatively realistic depiction of complex geometry. Further, the model was nonlinear and used a turbulence closure scheme (Mellor-Yamada level 2.5). Data for the McGillicuddy et al. (1998b) work were derived from 11 years of interdisciplinary sampling during the MARMAP program, and bimonthly data were binned together. Results of the modeling effort suggest that *Pseudocalanus* spp. populations located in the western portion of the Gulf of Maine and on Georges Bank may be self-sustaining (i.e., Gulf of Maine population separated from the Georges Bank population) rather than interdependent (i.e., Gulf of Maine population serving as a source for Georges Bank population) as suggested in previous studies. A technological aspect of special interest for this work was the development of a molecular genetic technique (Bucklin et al., 1998) that was used to document that two different *Pseudocalanus* species inhabit the region (see Lynch et al., 1998). According to Bucklin et al. (1998), these two species may differ significantly (e.g., distributions, reproductive timing, etc.). Future modeling studies may need to account for this factor.

Further Georges Bank work by McGillicuddy et al. (2001) has been devoted to OSSEs and real-time data assimilative modeling (Lynch et al., 2001). The former studies were directed toward the objective assessment of the synopticity of the observational surveys conducted during the U.S. GLOBEC program in the Georges Bank region. Based on OSSE results, it was determined that relocation of station positions (accounting for mean flow) could reduce “space/time smearing” error by about half. For the latter work, real-time forecasts were made for Georges Bank during the spring of 1999. Three-day forecasts were made and used by shipboard oceanographers for both interpreting results and planning operations. Assimilated data included current velocities (acoustic Doppler current profiler, ADCP), drifter trajectories, and taxa-specific plankton data obtained from a Video Plankton Recorder (VPR; Davis et al., 1996; Gallager et al., 1996). Forecasts in the forms of synoptic maps were found to be useful. An important recommendation resulting from the work was that more data communication bandwidth is needed for transmitting data among ships, moorings, drifters, and satellites.

Another study utilized the Georges Bank region to demonstrate real-time forecasting of physics, biology, and fish stock abundance (Rothschild et al., 2000; Robinson et al., 2001). The HOPS model and a nesting configuration similar to that applied during the LOOPS study were utilized for the study that spanned the period April 17 to May 15, 2000. Data used for the work were derived from satellite sea surface temperature, color, and altimetry and from historic ship-based data. Forcing was accomplished with atmospheric fluxes for year 2000. Surface color data were used for chlorophyll, and a feature model was used to provide chlorophyll depth dependence. A unique aspect of this work was the inclusion of a fish dynamics model to account for cod swimming behavior as related to preferred bottom
temperature along with a dispersive tendency. Nowcasts and forecasts of 1 to 2 days were done twice per week. Simulation products included daily maps of sea surface temperature, subtidal currents, chlorophyll at 15 m, bottom temperature and subtidal currents, and cod abundance near the bottom (see Fig. 12.22 in Robinson and Lermusiaux, 2002). This work was demonstrational; however, it illustrates the potential of data assimilation for highly practical ocean management and operational problems.

Harmful algal blooms (HABs) in coastal waters are gaining increased attention worldwide because of their deleterious effects on the ecosystem, water quality, human health, and economies, and because they appear to be becoming more widespread and frequent (e.g., Anderson, 1995a,b; Cullen et al., 1997). In the past, there was little hope for predicting the occurrences of HABs because of the difficult problems that they present for sampling and modeling. However, emerging instrumentation (e.g., Cullen et al., 1997; Davis et al., 2000) and better models (e.g., Franks, 1997; Walsh et al., 2001) provide impetus for new HAB studies, which logically embody data assimilation techniques. Many of the issues involved in the development of HAB observational networks and forecasting models have been reviewed by Schofield et al. (1999). Important distinctions of this problem juxtaposed with the previously discussed open ocean, ecological, and biogeochemical problem include the following: (1) shallow water coastal dynamics and mixing are especially challenging for models; (2) relevant time and space scales are relatively short; (3) more variables [multiple phytoplankton (including HAB) and zooplankton species] and model state variables are needed; (4) high spectral resolution optical data and models are necessary; and (5) food web and other functional relationships will likely be more complex. These are daunting issues. However, prediction of HABs seems feasible considering advances in integrated, nested arrays of in situ instrumentation, remote sensing, highly structured (many components) bio-optical and biological models (e.g., Franks, 1997; Bissett et al., 2001; Walsh et al., 2001), and data assimilation methodologies.


4. Uses and requirements of observations for data assimilation models and systems

Observational systems and associated data sets have many uses in the context of data assimilative modeling as indicated in the preceding examples. These can be generally summarized as follows.

(1) Data sets obtained during past process-oriented and time series programs can be used to:
   (a) estimate errors and uncertainties in data/models for use in assimilation weights, OSSE designs, etc.
   (b) reanalyze existing data sets (e.g., hindcasts)
   (c) identify key missing processes and variables, which should be incorporated in future interdisciplinary models and/or measured in future observational campaigns
   (d) develop new model formulations (and food web structures) and parameterizations
   (e) quantify space and time scales of significant variability that need to be resolved by models and
   (f) calibrate, validate, and verify models.

(2) New observational programs used with data assimilative models can be used for
   (a) joint “testbeds” for development of advanced data assimilation systems
   (b) conducting and directing focused process-oriented studies (e.g., nowcasts and forecasts; optimizing sampling)
   (c) long-term monitoring of changes in the environment
   (d) responding to environmentally harmful events and
   (e) estimating errors and uncertainties in data/models for use in assimilation weights, OSSE designs, etc.

Based on these multiple uses, there are several extremely important requirements of data in general,
regardless of application. Summarizing, the data need to:

(1) encompass key variables whose numbers seem to continually increase because of the discovery of new organisms and relevant chemical compounds and processes

(2) be collected to maximize resolution of processes of varying characteristic time and space scales (e.g., “oversampling concept”)

(3) be obtained from calibrated instruments

(4) include estimates of instrumental error and sampling statistics as well as instrument calibration data

(5) contain information concerning sampling frequency and averaging (enabling evaluation in terms of aliasing in space and time and synopticity).

It is important to note that presently interdisciplinary observational campaigns rarely satisfy all of these requirements. Much progress is possible and necessary. If data are used for nowcasting and forecasting applications, additional requirements include:

(1) transferring of data as quickly as possible, minimizing latency periods of data sets

(2) duplex or two-way communication so that adaptive sampling can be accomplished by redirecting mobile assets and modifying sampling frequencies in order to change temporal or spatial resolution and

(3) real-time data management and calibration.

The ability to satisfy these requirements will naturally vary in practical terms. For example, many satellite-based data sets will likely need longer latency periods than mooring and drifter data sets. It is worth noting that some of the previous data assimilative modeling efforts have indicated that some variables have longer acceptable latency periods by virtue of longer decorrelation time or space scales.

5. State-of-the-art and emerging interdisciplinary technologies

In this section, we review some of the present interdisciplinary measurement capabilities and suggest some possibilities for future advances. Here, we focus on the application of data assimilative systems. Several other reviews provide additional details about measurement systems and platforms (e.g., see Dickey, 1988, 1990, 1991, 2001, 2002; Robinson and Dickey, 1997; Glenn et al., 2000a,b; Dickey and Chang, 2001; Griffiths et al., 2001; several references in Koblinsky and Smith, 2001). Readers interested in briefer overviews are directed to recent papers by Dickey (2001, 2002). The first subsection is devoted to platforms. The following subsection describes specific disciplinary sensors and systems that can be interfaced with these platforms, making them effective interdisciplinary observational assets. Again, the emphasis of this paper is placed on biogeochemical, ecological, and population dynamics problems.

5.1. Platforms

Platforms represent the backbone of the observational component of any data assimilation system (Fig. 3). Nested, multiplatform approaches have been adopted by several major oceanographic process and time series studies in order to take advantage of the specific sampling capabilities of individual platforms (Figs. 2–5), which generally can carry a variety of interdisciplinary sensors and systems and often have telemetry capabilities. The horizontal spatial and temporal ranges of the various platforms are shown schematically in Fig. 2b. Next, we outline some of the specific capabilities and limitations of several types of platforms for data assimilation systems.

5.1.1. Ships

Observational ships (e.g., Dickey, 2002) can be placed into two categories: those that are dedicated to research activities and those that are used primarily for commercial work. Dedicated research ships are most useful for regional process-oriented studies and for long transect sampling programs designed to provide spatial maps (e.g., Feely et al., 2001). They are also required for sampling of large water volumes and deployment of cutting-edge instruments that are generally too large or require human intervention for deployment from other platforms (e.g., flow cytometers and mass spectrometers). Ships typically have a high capacity for deploying sophisticated instrumentation, but lack the ability to produce synoptic data.
sets and, strictly speaking, are neither Eulerian nor Lagrangian platforms. Ship data are thus most useful when assimilated in dynamical models so that the dynamics extrapolate between ship tracks, hence producing valuable quasi-synoptic fields.

Ships can be used for on-station profiling of instruments and underway sampling with technologically advanced instrumentation (via flowthrough or towed systems). Many research (and some commercial) ships now support nearly complete suites of instruments for meteorological measurements of wind stress and surface heat fluxes (e.g., Taylor et al., 2001) and ADCPs for measuring horizontal currents and acoustic backscatter for inferring zooplankton biomass (described below).

Some recent interdisciplinary data assimilation experiments have used and supported adaptive ship sampling on multiple scales (e.g., Robinson and the LOOPS Group, 1999; Lermusiaux, 2001; Robinson and Lermusiaux, 2002; Besiktepe, 2002; HyCODE, Glenn, personal communication). Commercially operated voluntary observing ship (VOS) and ships-of-opportunity observational programs are especially valuable for obtaining data in remote oceanic regions where few dedicated research ship sampling programs can be routinely executed. There are also interesting examples of uses of VOSs and ferries that make regular transects enabling collection of spatial-time series data (e.g., Rossby, 2001), some with near real-time data telemetry for use in data assimilation models (e.g., to compute spectra, coherence, etc.), to be deployed in harsh environments generally inaccessible by ships (high latitudes, during hurricanes and typhoons), and to telemeter data fairly easily. Biofouling of bio-optical, biological, and chemical sensors and systems remains a problem; however, considerable progress is being made (e.g., Chavez et al., 2000; Dickey et al., 2000) and 4- to 6-month deployments with minimal degradation of data by biofouling are now possible in the open ocean. Disadvantages of fixed-location measurements include the following: (1) they cannot provide horizontal spatial information and are practically limited to selected oceanic regions (ideally placed in representative locations) and (2) mixed temporal–spatial variability is measured and thus partitioning of local versus advective effects requires complementary spatial data sets.

5.1.2. Fixed-location platforms

Fixed-location platforms provide Eulerian data sets. Platforms falling into this class are mooring, bottom tripod, shore-based, and off-shore platforms (e.g., Dickey, 2002), which have been used to collect interdisciplinary time series data to measure changes in the ocean on time scales from minutes to years. An increasing number of bio-optical, chemical, bio-acoustical, and geological parameters are being measured from fixed-location platforms with many of the sensors described below. This work has led to discoveries of new processes including primary production variability associated with ENSO and equatorial longwaves, sediment resuspension caused by internal solitary waves, cloud-induced and diel fluctuations in phytoplankton biomass, and phytoplankton blooms associated with incipient seasonal stratification, eddies, and frontal and eddy-trapped inertial waves (e.g., see reviews by Dickey, 2001; Dickey and Chang, 2001; Dickey and Falkowski, 2002). Moorings with interdisciplinary instrumentation have been used in most of the world oceans and in coastal as well as open-ocean settings. Moored profilers and bottom tripods have also been used in deep-ocean and shallow waters. A draft global map of planned or existing time series observatories is shown in Fig. 9a (based on Send et al., 2001).

Advantages of Eulerian measurements from moorings and tripods include capabilities to capture episodic as well as high-frequency periodic phenomena by sampling time scales from minutes to many years (Figs. 2 and 6), to sample at multiple depths, to minimize undersampling and aliasing, to utilize well-developed data analysis methods (e.g., to compute spectra, coherence, etc.), to be deployed in harsh environments generally inaccessible by ships (high latitudes, during hurricanes and typhoons), and to telemeter data fairly easily. Biofouling of bio-optical, biological, and chemical sensors and systems remains a problem; however, considerable progress is being made (e.g., Chavez et al., 2000; Dickey et al., 2000) and 4- to 6-month deployments with minimal degradation of data by biofouling are now possible in the open ocean. Disadvantages of fixed-location measurements include the following: (1) they cannot provide horizontal spatial information and are practically limited to selected oceanic regions (ideally placed in representative locations) and (2) mixed temporal–spatial variability is measured and thus partitioning of local versus advective effects requires complementary spatial data sets.

Offshore platforms, including research-dedicated and oil production platforms, provide opportunities for conducting oceanographic and meteorological research and monitoring. These large, stable platforms often have space and facilities for manned research laboratories. They offer several advantages over shipboard platforms, including absolute stability in high sea states and suitability for time series measurements. In addition, it should be possible to launch and dock autonomous underwater vehicles (AUVs) and other...
Fig. 9. (a) Draft global map of planned or existing time series observatories. Based on figure in Send et al. (2001). (b) Global map depicting planned global distribution of Argo profiling floats. Based on figure in Argo Science Team (2001).
Fig. 9 (continued).
mobile sampling platforms from offshore. Active platforms are not useful for all types of measurements because of possible chemical contamination and nonrepresentative biology that may result from drilling operations and inherent perturbation of the natural habitat. Shadowing and wake effects are also problematic for optical and physical measurements.

It is expected that future time series observations in the coastal ocean and at selected sites of anticipated high environmental consequence (e.g., high latitude sites of deep water formation and/or \( \text{CO}_2 \) exchange) or special long-term monitoring value (e.g., Southern Ocean) will require mooring platforms. The need to consider costs and return on investment will make it essential to optimally select location sites (e.g., Send et al., 2001). It will be imperative to take advantage of multi-use platforms for interdisciplinary sensors and systems to maximize scientific and monitoring value. As one example, a mooring designed for a tsunami warning system has been used in the Pacific for measuring upper ocean biogeochemical and bio-optical variables relevant to global climate change (Dickey et al., 2001). There is also a need for small and essentially expendable moorings (1) that can be deployed in areas, which require large amounts of shiptime for deployment and recovery, (2) for special observational programs (e.g., in paths and wakes of hurricanes and typhoons and in harmful algal blooms), and (3) in areas where fishing activities, shipping lane traffic, and/or vandalism are problematic (e.g., coastal areas and western equatorial Pacific). Only a few studies have utilized mooring arrays (e.g., McPhaden et al., 2001; Dickey et al., 1998a; Abbott et al., 2001); however, this approach has been effective for studying evolving spatial features (e.g., fronts and eddies) and equatorial longwaves. Increased bandwidth for telemetry of data (e.g., Iridium communication satellite system) should enable transmission of multifrequency acoustical and multiwavelength optical as well as video data. Utilization of the data sets will also be enhanced through synthesis with data collected from other in situ and remote sensing assets and of course data assimilation.

Looking toward the future, use of offshore platforms will require formation of partnerships among government agencies, private industry, and academia (a few test programs are underway). Shore-based instrumentation is already being used to obtain surface current and wave data in several coastal regions (e.g., Glenn et al., 2000a,b). New high-frequency radar systems are extending the working range for current observations to about 200 km offshore. It is likely that high-frequency radar current coverage will span entire coastlines of some nations within the next decade. Coastal sites and piers will also be useful for other ocean observations (e.g., sea level, acoustics, horizontally oriented ADCPs, biological and chemical parameters, etc.). New acoustic instruments are also becoming available for measurements of directional wave spectra, bottom pressure, and currents in the surf zone. Eulerian platforms typically support local measurements; exceptions include high-frequency radar and horizontally oriented acoustic measurements. Thus, complementary measurements (providing spatial context information) from the various types of platforms described below are critical.

5.1.3. Drifters and floats

Drifters and floats (see Griffiths et al., 2001; Dickey, 2002) are Lagrangian platforms that effectively follow water parcels and provide spatial information. They fill in data in the horizontal spatial domain from meters to the basin-scale as indicated in Figs. 2b and 9b. Drifters and floats can be used to collect data in regions of the world oceans that are rarely visited by oceanographic or even commercial vessels. Global Positioning Systems (GPS) are often used with drifters and floats, providing very accurate position data (down to a few meters). Surface drifters can provide interdisciplinary data using many of the same sensors and systems that can be interfaced to moored and tripod sampling packages. In the past, floats typically moved at predetermined depths to give subsurface currents; however, profiling floats (using buoyancy adjustments to move vertically) have been used more recently to provide additional data (e.g., Argo floats for temperature, salinity, and reference velocity at depth; Argo Science Team, 2001; Davis et al., 2001) during rise and descent through the water column as part of their function. They use telemetry systems to send data to land-based stations upon surfacing.

Present projections indicate that roughly 1000 surface drifters (Swenson, 1999) and 3000 profiling floats (Argo floats profiling from near the surface to \( \sim 2000 \) m; Argo Science Team, 2001; Fig. 9b) will be in operation annually within the next few years. The costs
of drifters and floats are expected to decrease as more are produced and used. Emerging systems will likely be simplified and ruggedized to enable easier deployment from ships-of-opportunity and aircraft. New GPS and telemetry (including two-way duplex information exchange) capabilities will continue to improve position accuracy and increase the daily number of reported positions resulting in greater computed current accuracy along with much greater volumes of data throughput. A few oceanographers have deployed optical and/or chemical sensors from drifters and floats (e.g., Abbott et al., 1990; Bishop et al., 2002). Increasing numbers of interdisciplinary variables are expected to be sampled from these platforms in the future as sensor and system size, weight, power, etc. become less limiting. Biofouling, pressure, profiling frequency, data storage volume for telemetry, and costs (most drifters and floats are not recovered) are other issues requiring further consideration for use of interdisciplinary sensors with drifters and profiling floats.

5.1.4. Autonomous underwater vehicles (AUVs)

Autonomous underwater vehicles (AUVs; e.g., Curtin et al., 1993; Griffiths et al., 2001; Dickey, 2002) take several different forms. These include autonomous surface craft, gliders, and propelled AUVs. AUVs have become the focus of many efforts to increase accessibility to the ocean for many purposes and have several attributes for use in data assimilation systems. These include (1) potentially reduced cost per deployment, (2) capability to sample in environments generally inaccessible to ships (e.g., under ice), (3) ability to sample over repeated sections with good spatial coverage, (4) capacity to do feature-based or adaptive sampling, and (5) potential deployment of several AUVs from moorings, mother ships, offshore platforms, and coastal stations, some with data transmission and repowering facilities. For data assimilation applications, sampling flexibility afforded by AUVs is especially important. Most of the activities have been confined to engineering design and some practical field activities thus far. However, these vehicles are beginning to be exploited for a variety of scientific studies, thanks to the development of new, relatively small sensors and systems that consume moderate power and that can be interfaced to the vehicles (e.g., Griffiths et al., 2000; Yu et al., 2002; see Fig. 7). At present, several specialized groups are developing and using AUVs, and numbers of AUVs are expected to grow as mission length capabilities increase, costs decline, reliability improves, operation becomes more routine, and more sensors become available for various sampling needs. Creative uses of the vehicles will involve networking and information feedback loops (e.g., Curtin et al., 1993) to guide sampling programs (e.g., data assimilation systems) and responses to extreme natural and anthropogenically induced events.

5.1.5. Remote sensing

Aircraft-based systems can, in principle, provide very high spatial resolution data for altimetry, color (e.g., HyCODE program), temperature, salinity, and other variables. Aircraft-borne lidar (light detection and ranging) has been used to measure optical properties from backscatter profiles (e.g., Yoder et al., 2001). Work is underway to implement similar methodology using satellite sensors. Autonomous aircraft technologies are progressing, and operating ranges of 2500 km with diverse sensor packages (e.g., for meteorology, sea surface temperature, hyperspectral color, etc.) are projected for unmanned aerial vehicles (UAVs). UAVs could, in principle, fly over a site for a day or longer to enable collection of high spatial resolution time series and provide data in real or near real time. Aircraft- and satellite-based sensors are now capable of providing nearly synoptic regional and global oceanographic data (e.g., references in Koblinsky and Smith, 2001; Yoder et al., 2001; Dickey, 2002; Dickey et al., 2003). The data are typically empirical inferences of surface signals (e.g., either passive or active electromagnetic radiation) and are based on calibration and seafloor data sets obtained from ocean-based platforms. Electromagnetic radiation only penetrates to shallow ocean depths (e.g., infrared to millimeters and visible to meters), thus aircraft- and satellite-derived information must be complemented with in situ observations. Considerable research effort is needed to extract subsurface data using remote sensing and in situ data sets and models. Data assimilation methods can be used for both this purpose and for other objectives such as parameter estimation and prediction. Oceanographically important remotely sensed variables include temperature, salinity, solar radiation, wind stress and direction, rainfall, surface heat fluxes, ocean color (e.g., pigment concentra-
tions), and sea surface height (altimetry). Several recent interdisciplinary studies have used combinations of remotely sensed sea surface height, ocean color, sea surface temperature, and wind data along with in situ data (e.g., see Dickey, 2001).

Temporal and spatial resolutions of satellite remote sensing systems are likely to improve for many parameters. Clouds will continue to be an obstacle for temperature and color measurements. However, by using multiple conventional orbiting or geostationary satellites, the time gaps will be decreased and data sets more nearly completed. Altimetry measurements are essentially unaffected by clouds, but present altimeters have good spatial resolution only in a narrow beam directly under the satellite’s flight path. However, concurrent and coordinated sampling with multiple satellite missions and new wide swath systems will likely make it possible to collect essentially two-dimensional altimetry data (e.g., Mitchum et al., 2001). In addition, new satellite missions will enable improved accuracy of measurements of the global marine geoid. This will allow considerable improvement in the altimetric determination of absolute dynamic topography and ocean circulation.

Work is also progressing in remote sensing of other important oceanic variables. In particular, studies are underway to measure salinity from satellites (e.g., Lagerloef and Delcroix, 2001). Other variables are likely to include colored dissolved organic materials and signatures of different phytoplankton groups, including those associated with harmful algal blooms (e.g., Cullen et al., 1997; Schofield et al., 1999; Kirkpatrick et al., 2000). Planned missions for color and temperature can be expected to achieve spatial resolutions of tens of meters or less (probably over limited selected areas) as well as increased optical spectral resolution down to a few nanometers (Davis et al., 1999). Remote sensing of many important chemical and biological variables (e.g., zooplankton and fish) remains as a major research challenge. Regular observations of higher trophic level organisms from satellite platforms are not presently feasible. However, ocean color, temperature, and current data can be valuable for identifying features (e.g., fronts, eddies, upwelling areas, HABs, etc.) where high biological activity may be located. Further, extremely high-resolution imagery may eventually be available for sensing surfacing mammals and large schools of fish. Data assimilation methods will be valuable for synthesizing such disparate data types and ultimately for attacking difficult problems involving ecosystem and population dynamics.

Possibilities for event-triggered sampling using sensors placed on special “intelligent” satellite platforms are being pursued. Future satellites may utilize steerable (pointable) instruments in geostationary orbit, while others may employ geographic information system (GIS) software and intersatellite communication capabilities. These approaches are most attractive for responses to disasters and for directing field and other remote sensing assets toward key locations to provide data that would otherwise be unattainable. Data assimilation should play an important role in these approaches. Data assimilation modeling activities will also be required to optimize the use of present and future remote sensing data sets, especially in conjunction with complementary in situ data sets. Accessibility of remotely sensed data in real time or near real time remains problematic, especially when commercial licensing is involved. Thus, some data assimilation applications will likely be restricted to hindcasting for some, but not all, remote sensors until this issue is resolved satisfactorily.

5.2. Sensors and systems

More capable oceanographic interdisciplinary sensors and systems are emerging at a fast pace for several reasons. These include technology transfer from other fields (e.g., medicine, engineering, space, microelectronics, military defense, etc.; see Kaku, 1997), increased interest in and support of environmental observations, and formation of functional partnerships among academia, government laboratories, and private industry. The effect of this accelerated technological progress has been to enable the collection of data for more variables at higher sampling rates, thus improving our ability to study many of the processes illustrated in Figs. 1 and 2a with less undersampling and aliasing and expanded temporal and spatial coverage, often in real or near real time. These advances are clearly advantageous for data assimilation activities.

Design factors for future sensors and systems to be deployed from the platforms described above include response time, stability, drift, size, power require-
ments, durability, reliability, susceptibility to biofouling, data storage and telemetry, and cost. Another important problem facing sensor and system developers and users is the proper interpretation of the instruments’ signals. For this reason, testbed sampling programs that use multiple platforms for intercomparisons are absolutely imperative for capitalizing on emerging ocean technologies (Dickey et al., 1998b, 2000, 2001).

Optical and bio-optical sensors have been especially pivotal in moving interdisciplinary oceanography forward. Ocean optics refers to studies of light and its propagation within the ocean; bio-optics concerns biological effects on optical properties and light propagation and vice versa (see Dickey et al., 2003). These closely related subdisciplines are important for many interdisciplinary ocean problems, often entailing biological–optical–physical interactions (Fig. 1). Examples include ocean primary productivity, upper ocean ecology, biogeochemical cycling and the biological pump, sediment resuspension and transport, ocean pollution, and bio-optically modulated variability in upper ocean heating rates (e.g., Dickey and Falkowski, 2002). An important point is that optical properties are primarily biologically modified in the open ocean environment, whereas terrigenous input and resuspended sediment also play major roles in the coastal ocean.

Relationships between physical and biological processes can often be inferred by comparing satellite maps of sea surface temperature and ocean color (e.g., Dickey et al., 2003). Subsurface light fields are affected by both incident solar radiation and properties of ocean waters, which are highly variable in space and time because of a host of complex physical, chemical, and biological processes. Particulates, both organic and inorganic, and dissolved materials play key roles in the variability of ocean optical properties. Also, the spectral quality of light is often fundamental to understanding several processes (Fig. 1; see Dickey et al., 2003).

Descriptions and discussions of in situ sensors and systems are subdivided below in terms of their primary use: physics, chemistry, optics and bio-optics, and biology and bio-acoustics. Again, interdisciplinary measurement suites are essential to maximize cost effectiveness in shared platforms and telemetry systems and for simultaneous, complementary observations enabling interdisciplinary data assimilation modeling.

5.2.1. Physics

Measurements of atmospheric variables above and at the sea surface, currents, and physical water properties are important for interdisciplinary oceanography. Examples of uses of these observations include quantification of air–sea gas exchange and other interactions and advection of chemical and biological species. New meteorological measurement systems developed for ships and buoys now allow net surface heat fluxes to be made with accuracies sufficient to provide global mean surface heat balances of better than 10 W/m² (e.g., Taylor et al., 2001). Satellite capabilities of measuring wind stress, surface heat fluxes, and sea surface temperature are all increasing as well (references in Koblinsky and Smith, 2001). The combination of in situ and remotely sensed surface data is enabling improved data sets for forcing, parameterizing, and evaluating data assimilation models.

Ocean temperature, salinity, and current measurements have advanced, both in terms of quality and quantity (e.g., Dickey et al., 1998c). Shore-based high-frequency radar systems are being used at some coastal sites to great advantage. Systems are being used for shorter ranges offshore (~ 40 km with ~ 1.5-km resolution) and longer ranges (~ 200 km, with ~ 6-km resolution). For example, these systems are being used at the LEO-15 site for data assimilative models as described earlier (Fig. 5). Most systems are shore-based at present. However, offshore platforms and buoys can be used to expand the spatial coverage (e.g., Glenn, personal communication).

Current and turbulence measurements from bottom tripods can be made in conjunction with optical and acoustical measurements to study sediment resuspension and transport. Expendable probes have been used for some physical data assimilation work, primarily for temperature profile data as discussed earlier. Optical, acoustical, chemical, and other probes can be added to expendable packages. An important new method involves tracers (e.g., Fine et al., 2001), which have been exploited for upper ocean, bottom boundary layer, and deep circulation studies. Several tracers are being exploited. These have been introduced either through anthropogenic activity (radio-
active tritium and carbon, chlorofluorocarbons, etc.) or by experimenters (e.g., fluorescein dye, sulfurhexafluoride, etc.). Different time scales of tracers are used to advantage. Tracer data are mostly amenable to inverse methods (e.g., Wunsch, 1996; Fine et al., 2001) and hindcasting rather than nowcasting and forecasting at present.

Looking toward the future, remote sensing will be expanded to increase numbers and types of meteorological and physical measurements over the world oceans. In situ measurements from buoys and drifters will continue to be needed to provide seathruing data and to provide local high-frequency time series data, which cannot be obtained from satellite sensors. For example, localized weather, storm, and hurricane and typhoon systems cannot be adequately sampled from satellites alone.

The cost of physical measurement systems is a primary concern, as large numbers of measurements are needed. Multiple users and uses of data can reduce effective expenses via sharing. In particular, ADCPs can continuously sample vertical profiles of horizontal current and acoustic backscatter (related to zooplankton biomass; Smith et al., 1992; Roe and Griffiths, 1993; see below) while shipboard profiling or tow-yo operations are being conducted. Further, many commercial ships now routinely measure currents using ADCPs. Some AUVs are already using ADCPs as well (Fig. 7). Novel uses are also being explored for underwater sound measurements using ADCPs and other instruments to estimate rainfall rate and wind speed.

Adjustments of specified vertical velocities have often been made to improve ecological model simulation/data agreement. Unfortunately, vertical velocity remains a difficult but important measurement for biological and chemical as well as physical oceanographers. Very few examples of accurate measurements of vertical velocity have been reported (see Dickey et al., 1998c). However, specially designed mechanical instruments and acoustic systems, the latter using both the Doppler effect and backscattering, have been used to investigate vertical motions (e.g., in strong upwelling and convective regimes and Langmuir cells). Near surface-scanning Doppler sonar also has the potential to provide simultaneous measurements of wind direction, wave directional spectra, Langmuir circulation, and mean near-surface flow.

Turbulence measurements are also important for interdisciplinary studies. Only a few specialists have been able to measure ocean turbulence successfully. More routine and cost-effective turbulence measurements, which can be made from a host of platforms (e.g., AUVs—Levine, personal communication), are needed for many types of applications. Some efforts are also being made to extract turbulence data from ADCP measurements (Simpson, personal communication).

Acoustic tomography shows great promise for two important applications: for long-term ocean basin-scale integrated temperature change (acoustic thermography) and for measuring ocean circulation patterns, especially at the mesoscale (Dushaw et al., 2001). This methodology involves the use of acoustic transmissions along multiple paths and requires arrays of sound sources and receivers and inversion algorithms. Acoustic tomography is especially valuable for demanding areas such as boundary currents, straits and throughflows, under ice, and in regions of active convection and bottom water formation. It has been done mostly in deeper waters, but would be valuable in coastal waters as well if technical and analytical issues can be resolved. Some existing hydrophone arrays may be used, and researchers are developing software to make acoustic tomography more accessible and widely used.

Real-time telemetry of data during acoustic tomography operations is feasible, however, considerable data bandwidth is required. One of the important goals of physical oceanography is to optimally use remotely sensed data (e.g., altimetry, wind, sea surface temperature, etc.) and in situ data (from all available in situ platforms, especially moorings, drifters, profiling floats, AUVs, etc.) to provide four-dimensional fields of physical properties and currents. Inverse and data assimilation methods are being used increasingly for these purposes (Le Traon et al., 2001; Stammer et al., 2001).

5.2.2. Chemistry

Examples of scientific areas of special relevance to ocean chemistry include global warming due to the greenhouse effect, nutrients and their roles in primary productivity and the biological pump (for transporting carbon to the deep sea), as well as coastal eutrophication, ocean pollution, and hydrothermal vents. Uses of tracers were mentioned in Section 5.2.1. Some of
the important advances in chemical sensors and potential future applications are described below (also see reviews in Varney, 2000; Blain et al., 2000; Hydes et al., 2000; Clayson, 2000; DeGrandpre et al., 2000; Tokar and Dickey, 2000; Dickey, 2001; Dickey et al., 2001; Fine et al., 2001).

Ocean chemistry was generally done using water samples collected with bottles at-sea in the not too distant past. Analyses of ship-derived water bottle samples are still important. Also, ship-based pumping systems are being used to bring subsurface samples onboard ships for similar analyses as well as water collection. However, several other new methods that will be critical for data assimilation are being developed and put into use. A few examples follow.

5.3. Nonautonomous sampling

Moored serial water samplers are obtaining discrete preserved samples periodically over the course of several months at intervals of days to create chemical time series (e.g., work by Boyle et al.; see Dickey et al., 1998b). Similar systems have also been developed for drifters (Abbott et al., 1990). Trace metals, as well as macronutrients, can be analyzed using the water samples. The importance of atmospheric input of dust and aerosols into the ocean has been recognized (e.g., role of iron fertilization in carbon cycling). Samplers for dust, aerosols, and particular chemical species are being developed for deployment from surface buoys to avoid land contamination (Sholkovitz, personal communication).

5.4. Autonomous sampling

Autonomous underway surface water chemical measurements are now carried out using shipboard water intake systems and laboratory chemical analyzers [e.g., variables include dissolved oxygen, partial pressure of carbon dioxide ($pCO_2$), and nutrients such as nitrate, phosphate, silicate, and ammonium; see Varney, 2000; Dickey et al., 2000; Feely et al., 2001].

In situ autonomous measurements from fixed-location and moving autonomous sampling platforms have major advantages in sampling ocean chemistry as samples do not suffer from degradation and are representative of local environmental conditions at depth. Some sensors and systems are capable of making in situ time series measurements with sampling intervals of a few minutes for durations of months. An example of a new system designed for autonomous measurements of a broad suite of chemical measurements, many of which were indicated above to be important for interdisciplinary data assimilative models, is the spectrophotometric elemental analysis system (SEAS; Byrne et al., 2002). SEAS is capable of spectral analysis in the optical waveband from 400 to 750 nm in both absorbance and fluorescence modes. The system’s sample cell is configured to use long path-length liquid core waveguides (10–500 cm); analyte detection limits are on the order of 1 nM or better. Internal memory of SEAS is sufficient to store 1000 complete measurements. Field deployments of SEAS have been used to measure in situ nitrite profiles to 200 m and surface seawater pH on a long-term mooring. Demonstrated measurement capabilities using liquid core waveguide systems in absorbance and fluorescence modes include ammonium, nitrate, nitrite, iron, copper, chromate, molybdate, hydrogen sulfide, pH, $pCO_2$, total inorganic carbon, total alkalinity, and fluorescence of natural organics. Also, SEAS is currently being configured for measurements of ferric and ferrous iron in rainwater. SEAS and several other new devices are being developed for interfacing to a variety of autonomous platforms. It is worth noting that real-time and near real-time telemetry of chemical variables has also been accomplished, and that there are a few examples of modifying sampling (gain changes, etc.) using two-way or duplex data telemetry systems (Dickey et al., 1998b).

In the future, several new chemical methodologies will likely become useful for oceanographic measurements. For example, autonomous water tracer measurements from moorings and other platforms could serve as valuable complements for ship-based survey tracer sampling (e.g., Fine et al., 2001). These data could be telemetered and used for inverse and data assimilative models. There is clearly a need to increase capabilities for sampling broader suites of chemicals autonomously for many problems such as pollution, with diverse chemical species of concern (e.g., PCBs, DDT, toxic metals, etc.).

It is expected that chemical sensors rather than analyzers will be preferable, if not required, for some platforms such as ship-towed bodies, profiling floats, drifters, and AUVs. It is also anticipated that artificial
neural networks will be useful for some chemical and perhaps biological measurement systems. Some new ocean measurements are utilizing fiber optic sensors (see Varney, 2000; Tokar and Dickey, 2000). Examples of successfully measured chemical species using fiber optic sensors include ammonium, methane, and dissolved carbon dioxide. These are variables of high interest to data assimilation modelers working on ecological, phytoplankton, and biogeochemical problems.

Another promising type of chemical sensor is the microelectromechanical system (MEMS; see Tokar and Dickey, 2000). MEMS is based on a relatively new technology, which is used for making and combining miniaturized mechanical and electronic components out of silicon wafers using micromachining. MEMS have shown encouraging results for sensing physical parameters, but work is still needed to realize their full potential for chemical sensing. Potential advantages of MEMS include autocalibration, self-testing, digital compensation, small size, and economical production, potentially enabling widespread use on most in situ platforms. Verification of data using discrete water samples will be necessary during the developmental phases for many of the emerging chemical systems.

5.4.1. Optics and bio-optics

Phytoplankton are the organisms most easily studied with the techniques described in this subsection. Other optical and video methods used to study zooplankton are discussed in Section 5.4.2.

Ocean optical properties are conventionally classified as inherent optical properties (IOPs) and apparent optical properties (AOPs). IOPs depend only on the medium and are independent of the ambient light field. On the other hand, AOPs depend on both the IOPs and the geometric structure of the subsurface ambient light field. Several instruments are designed for measuring IOPs (e.g., spectral beam attenuation, absorption, and scattering coefficients) and AOPs (e.g., spectral diffuse light attenuation coefficients). Reviews of recently developed IOP and AOP measurement techniques and data sets are presented in Dickey (2001), Dickey and Chang (2001), Dickey et al. (2003), and Dickey and Falkowski, (2002).

Direct in situ measurements of IOPs were generally limited to single wavelength [usually 660 nm because of the availability of photodiodes at that wavelength and because of desired low light attenuation due to colored dissolved organic matter (gelbstoff) in seawater at 660 nm] beam attenuation coefficients until a few years ago. These measurements have been valuable for estimating sediment and particle concentrations and particle production (Siegel et al., 1989), and even proxy measurements of particulate organic carbon (Bishop, 1999; Bishop et al., 2002). However, interest in optically characterizing particulate and dissolved materials stimulated development of instruments that are now capable of measuring light absorption and beam attenuation coefficients at multiple wavelengths (from 9 to about 100 different wavelengths; e.g., Bruce et al., 1996). Likewise, measurements of radiance and irradiance are now being done for 7 to 100 wavelengths. The power of these types of instruments lies in their ability to distinguish phytoplankton from detritus and colored dissolved materials, and potentially to identify phytoplankton (perhaps including harmful algae; Cullen et al., 1997; Kahru and Mitchell, 1998; Schofield et al., 1999; Kirkpatrick et al., 2000; Dickey and Chang, 2001) at least by community groups based upon their characteristic absorption spectra. AOP measurements have the potential for similar analyses, although their use for these purposes is more difficult because of the need to deconvolve the varying ambient light field. Again, this capability is of great value to modelers interested in exploring multicomponent models (e.g., including different phytoplankton groups or if possible species) with differing physiological and behavioral characteristics.

A recent research and development thrust for in situ optical instrumentation is in the area of scattering of light and its angular dependence. These data types, along with absorption, are important for fully characterizing the underwater light field, which has great implications for remote sensing as well as for fundamental optical radiative transfer problems (Mobley, 1994). Concurrent measurements of key IOPs and AOPs are critical for developing inversion models so that IOPs can be determined from AOPs and vice versa. Another important research goal is to estimate the vertical structure of IOPs given remote sensing measurements of water-leaving radiance. Inverse and data assimilation methods should prove useful for this work.

Spectral fluorescence has been used to measure dissolved organic materials as well as special dis-
solved substances associated with aging sewage discharge waters (Petrenko et al., 1997). For this method, ratios of pairs of excitation and emission signals for specific wavelengths are used as indicators of material properties. Optical instruments using Fraunhofer diffraction have been used to obtain particle size distributions, primarily in bottom boundary layers. A different and extremely powerful optical technology, flow cytometry, has been successfully used onshore and on-board ships for counting and distinguishing particles and phytoplankton as well as for characterizing their optical properties. Work is progressing to miniaturize and ruggedize flow cytometers, and some testing is underway at-sea using buoys.

Rate measurements have been lacking, but are especially valuable for data assimilative models as discussed earlier. An important observational goal is to estimate primary productivity. To date, several different optical measurements have been used for these determinations (e.g., Dickey and Falkowski, 2002; Dickey et al., 2003). Examples include the use of chlorophyll fluorescence (based on fluorometer measurements) and photosynthetically available radiation (PAR; measured with quantum scalar irradiance sensors; typically 400–700 nm) data with empirical models and, more directly, the use of sophisticated measurements using “pump and probe” fluorometers (e.g., Kolber et al., 1998). The latter instruments have the important advantage of providing information about biophysical parameters related to photosynthesis (e.g., quantum yield as affected by nutrient and light conditions). Many optical instruments have been deployed from ships, moorings, and drifters. Thus, sampling with spatial and temporal resolutions comparable to those of physical variables is now being accomplished.

One of the common uses of in situ optical measurements has been to calibrate and validate remotely sensed spectral optical data as well as to develop algorithms (i.e., using ratios of water-leaving radiance at different wavelengths) for retrieving chlorophyll concentrations and other properties (Dickey et al., 2003). In addition, variabilities of phytoplankton biomass, primary productivity, and upper ocean radiant heating rates have been estimated using both in situ and remotely sensed data (see Dickey and Falkowski, 2002).

The time series obtained from many of the aforementioned in situ optical systems shows remarkable variability associated with high frequency and episodic events as well as longer term processes. Interpretation of these time series is often difficult because of the complex nature of the observed medium including its varied organic, inorganic, particulate, and dissolved constituents. Optical models are needed to partition these components (e.g., Chang and Dickey, 1999). Finally, more work is being directed toward autonomous sampling of optical properties from moorings, drifters, profiling floats, gliders, and AUVs. Capabilities for the telemetry of optical data are also increasing, but higher data bandwidth capability is needed (again, the Iridium satellite communication system shows great promise).

5.4.2. Biology and bio-acoustics

Some of the most exciting prospects for using new data sets for data assimilation modeling lie in the area of biological oceanography (e.g., see reviews by Dickey, 1988, 1990, 1993, 2002; Smith et al., 1992; Jaffe et al., 2001; Marine Zooplankton Colloquium 2, 2001). One of the important goals of current biological research is to understand and ultimately predict how populations of marine animal species (from phytoplankton through fish) respond to natural and anthropogenic changes. This research is driven in part by the waning of fisheries and concerns for biodiversity. Zooplankton are key to several current research areas as highlighted recently in a review by Marine Zooplankton Colloquium 2 (2001). These include (1) “hot spots” defined as locations where zooplankton tend to aggregate and operate at higher rates (what are the processes leading to these “hot spots” and how do they persist?), (2) individual zooplankton species or genera (what governs their abundances and variability?), and (3) the role of zooplankton in biogeochemical cycling (what are the fundamental processes, what are the key relevant variables, and how can they be quantified?). At the lower trophic levels, it is becoming evident that microbes also play crucial roles in biogeochemical cycling, specifically in terms of their activities in production and decomposition of particulate organic matter (e.g., Fuchs et al., 2002). Key indicators and ecologically relevant variables must be carefully chosen because of the tremendous variety of marine life. These will often be dictated by special regional oceanographic characteristics and ecologies (e.g., biogeographical provinces; Longhurst, 1998).
Net sampling was used almost exclusively for early studies of phytoplankton, zooplankton, and higher trophic level organisms. An advantage of this method is that an individual organism can be captured, numbered, analyzed, and studied in detail. However, large amounts of shiptime and personnel are generally required for net-based sampling, leading to high costs and thus limiting geographic and temporal coverage. In addition, nets do not necessarily capture representative specimens (e.g., problems of net avoidance, bias toward capturing unhealthy organisms, damage to organisms, and other sampling issues). Nonetheless, important data sets have been collected with various types of net samplers. One important long-term effort is the Continuous Plankton Recorder (CPR) program that has collected plankton samples using ship transsects in the North Atlantic since the 1930s (Reid et al., 1998).

Studies of higher trophic level organisms (e.g., zooplankton and fish) are also using new video and acoustical techniques. One method of sampling zooplankton employs light sheets (Optical Plankton Counters or OPCs). These systems are usually profiled or towed behind ships and provide zooplankton biomass and size distributional data. Calibration and interpretation of the data relies upon occasional in situ net sampling. Video imaging systems (e.g., Video Plankton Recorders, VPRs; Davis et al., 1996) provide detailed organismal information. Image analysis is a demanding aspect of this approach, but considerable progress is being made (Tang et al., 1998). The informational value of these collective systems is high because statistical methods can be applied to examine important questions such as plankton patchiness, “hot spots,” and predator–prey interactions (e.g., Ashjian et al., 2001). OPC and VPR systems have been profiled and towed from ships. Work is underway to deploy these from other platforms for autonomous data collection. Methodologies such as holography appear promising for in situ applications as well (e.g., review by Jaffe et al., 2001). In principle, data resulting from these various imaging systems can be telemetered; however, again, large data bandwidth capability is needed.

Fisheries research has used acoustics since the 1930s. Breakthroughs in the use of acoustics for biological research (e.g., see review by Griffiths et al., 2002) have resulted in part from the development of multifrequency systems (e.g., Holliday and Pieper, 1995) and the utilization of ADCP backscatter data (e.g., Smith et al., 1992; Roe and Griffiths, 1993). Acoustical inversion models have been developed to use multifrequency data to estimate total biomass of zooplankton and to determine size distributions (in some cases, with vertical resolution of tens of centimeters; see Holliday et al., 1998). Zooplankton size (and possibly taxa) distributions in space and time can be obtained by deploying acoustic systems. Again, this is important information for the development and testing of ecological, biogeochemical, and population dynamics model elements of data assimilative systems as indicated earlier. Since the choice of the size ranges of targeted organisms is dictated by the choice of transducers (e.g., lower frequencies for larger organisms), a rather complete size spectrum of organisms can be obtained (Holliday and Pieper, 1995). Choices of targeted size classes dictate the number of different transducers or alternatively the selection of frequency band choices using chirp acoustical scanning over a broad frequency range. Deployments of acoustic systems using a side-looking mode suggest that distributions of migrating fish (e.g., salmon) can be observed. Also, acoustic tomography may be useful for some applications (e.g., mapping fish parameters over large areas by using measures of absorption losses at resonance frequencies of key organisms). The interpretations and analyses of both video and multifrequency acoustical data types remain as major challenges. Data assimilation modeling will likely be a valuable tool for these technologies.

Some investigators have used backscatter acoustical signal data obtained from monofrequency ADCPs to estimate biomass (e.g., Smith et al., 1992; Roe and Griffiths, 1993; Griffiths et al., 2002). The power of this approach is that ADCP data are being collected by commercial as well as research interests; thus, it may be possible to collect very large volumes of data with continuously sampling ADCPs. However, interpretation (e.g., size class information) remains problematic without multiple frequencies. Three-dimensional acoustic, optical, and video systems can collect data concurrently, optimizing the advantages of each method (e.g., Benfield et al., 1998; Jaffe et al., 2001). Such hybrid sampling can be used to provide information concerning predator–prey interactions. These methods still need occasional net sampling to calibrate.
and interpret data. The optimization of combinations and integrations of diverse data sets such as those described here can be accomplished using data assimilation techniques.

The measurement of microbes appeared to be out of range until Fuchs et al. (2002) recently reported a new methodology involving very thin light sheet microscopy. This method appears to provide a promising way to observe microbes in natural settings. Other optical methods are developing for in situ studies of predator–prey interaction observations (e.g., Jaffé et al., 1998). Interesting new tools for biological oceanography involve molecular genetic and species-specific molecular probes. In particular, molecules such as ribosomal RNAs have already begun to be used for detection and determinations of abundance of some bacterioplankton species and for studying evolutionary relationships among several species. Also, species-specific probes (e.g., DNA antibody) for HABs have been developed and offer great promise in helping us to understand the context of these events (Scholin et al., 1997). This technique would fit well with future data assimilation work on HABs. Applications of genetic techniques to higher trophic levels (zooplankton to whales) are also being pursued (e.g., Bucklin et al., 1998). Finally, measurements of bioluminescent plankton and nekton may lead to development of proxy indicators of particular organisms.

6. Sampling networks and schemes

Platforms and measurement systems for data assimilation systems require consideration of the oceanographic settings of interest along with the placements and movements of in situ fixed and mobile platforms, respectively. It is presumed that satellite temporal and spatial coverage parameters are predetermined, although this may not be a restriction for some future satellites (e.g., pointing capability). The following discussion is subdivided into coastal and open ocean subsections.

6.1. Coastal

The coastal ocean domain considered here lies between the shoreline and the continental shelf break. A general introduction to data assimilation methods for coastal observing systems is given by Walstad and McGillicuddy (2000). Sampling considerations must account for (1) physical processes including tides, upper and bottom boundary layers, surface and internal gravity waves, submesoscale, mesoscale, and frontal features, and buoyancy flows, (2) biological processes including primary and secondary productivity, grazing, mortality, larval dispersal, trophic/food web dynamics, HABs, and eutrophication, (3) special coastal bio-optical considerations as described earlier, and (4) chemical processes involving input of riverine and stormwater nutrients and other naturally occurring and anthropogenically produced chemicals, toxic chemicals, and anoxia. Most coastal interdisciplinary problems require a large number of measurements of biological and chemical variables. Also, biofouling of chemical and bio-optical sensors and systems is considerably more problematic in the coastal zone because of high biological productivity.

There are several attractive aspects of designing coastal networks and sampling schemes vis-à-vis their open ocean counterparts. For example, (1) in situ and shore-based (e.g., meteorological towers, piers, and high-frequency radar systems for current measurements) sampling assets can be deployed with relative ease and modest expense, (2) high-density sampling arrays can be established as testbeds for research and development of instrumentation and data assimilative models and systems (e.g., LEO-15; Fig. 5), (3) telemetry of data can be done by using fiber optical or electromagnetic conducting cables, radio frequency and cell phone telecommunications, and communication satellites, (4) it is relatively easy to redirect mobile sampling assets based on data assimilative model predictions, (5) conducting cables running from shore to nodes may be used to supply power to moorings and for recharging AUV batteries, (6) manned or unmanned aircraft can be used for remote sensing, and (7) emerging satellite remote sensing systems designed specifically for the coastal ocean will allow higher spatial resolution for several variables.

Past studies have utilized drifter, ship transect, aircraft, and satellite data for temperature and optical properties (including chlorophyll) to compute horizontal spatial decorrelation scales, spectra, and coherences (e.g., Gower et al., 1980; Lewis and Platt, 1982;
Denman and Freeland, 1985; Yoder et al., 1993; Denman and Abbott, 1994; Abbott and Letelier, 1998) and moorings located at varying distances offshore to determine temporal decorrelation scales, spectra, and coherences (Chang et al., 2002). As an example, Chang et al. (2002) recently found that decorrelation scales generally increase with distance from shore; however, they note that small spatial-scale features associated with sharp fronts or convergences and short time scale events associated with processes like internal solitary waves can occur far offshore as well. A practical guide for platform placement is to increase sampling resolution closer to shore in order to optimize information. This concept is embodied in nesting or telescoping of sampling assets and model computational grids. At the same time, it is highly desirable to use adaptive sampling to better sample small-scale offshore physical and ecological features, particularly when based on data assimilation nowcasting and forecasting (e.g., LOOPS example presented earlier). More specifically, this approach can allow intensive sampling in areas with high gradients in physical, chemical, and biological variables (e.g., upwelling fronts) or where environmentally hazardous conditions are present or forecast to be developing (e.g., oil spills, stormwater or agricultural nutrient runoff, fish kills, HABs, and storm surges). Other recent references on various aspects of coastal ocean observing systems include papers by Glenn et al. (2000a,b) and Malone and Cole (2000).

6.2. Open ocean

Open ocean horizontal scales of variability are typically greater than those of the coastal ocean (e.g., Dickey et al., 2001). Mesoscale processes and their relationships with variability in biogeochemical properties and fluxes have been foci of several papers and debate for the past few years (e.g., recent reviews by Garcon et al., 2001; McGillicuddy, 2001; Oschlies, 2001b, 2002a,b; Lewis, 2002). In particular, McGillicuddy et al. (1998a,b) suggested that mesoscale eddies may play a major role in supplying nutrients to the upper ocean (euphotic layer) and may represent a major source of fueling of new production in the Sargasso Sea. However, Oschlies and Garcon (1998) simultaneously forwarded a counterview as they indicated that there might be a less important role for the eddy process than suggested by McGillicuddy et al. (1998a,b). Interestingly, several of the estimates have been based on similar methodologies: remotely sensed sea surface height and numerical models. The concept of using satellite altimetric data to estimate quantities such as new production (e.g., McGillicuddy et al., 1998a,b; McGillicuddy, 2001; Siegel et al., 1999) has considerable appeal. A premise of this work is that a cyclonic eddy (first baroclinic mode type) has a lower sea level anomaly, isopycnal surfaces and nutrient rich waters upwell into the euphotic layer, and phytoplankton utilize the nutrients resulting in large phytoplankton (chlorophyll) concentrations (McGillicuddy et al., 1998a,b). This rectification process is apparently only one of several scenarios, and we are not sure that this is in fact the most common one. For example, in situ measurements of physical and biogeochemical parameters have revealed that a far more complicated set of processes is also governing nutrient fluxes into the euphotic layer, primary production, and new production (e.g., McNeil et al., 1999; Dickey et al., 2001; Lermusiaux, personal communication). In fact, the North Atlantic, including the Sargasso Sea, is populated with a variety of eddy, mesoscale, and submesoscale features. More specifically, high temporal resolution (minute time scales) mooring-based time series off Bermuda (Fig. 6) have also shown high biomass (and inferred productivity) associated with second baroclinic mode eddies (e.g., mode eddies in the Sargasso Sea; McNeil et al., 1999) and warm mesoscale features (e.g., Dickey et al., 2001; Conte et al., submitted for publication). In addition, we hypothesize that the biogeochemistry of mesoscale and eddy processes is affected by synoptic-scale weather (e.g., including hurricanes and typhoons), seasonal, interannual, and decadal variability. It is also important to further note that each eddy has its own unique shape, current and property structure, genesis, life cycle, and ecological and biogeochemical characteristics.

In designing a sampling system or data assimilative model grid spacing capable of resolving mesoscale features, it is useful to consider statistical decorrelation scales in horizontal space. For eddies, these are expected to generally scale with the Rossby radius of deformation. The need for interdisciplinary data sets with horizontal sample spacing considerably less than the Rossby radius is well accepted (e.g., Lewis, 2002; Oschlies, 2001b, 2002a,b). There is also considerable
vertical variability and structure for biological and chemical parameters that simply cannot be observed from remote sensing platforms (e.g., subsurface chlorophyll maxima, higher modes of baroclinicity, water mass intrusions at various depths, etc.). Horizontal spatial structure functions derived from data collected with Airborne Oceanographic Lidar (AOL) for the NABE area have shown significant correlations in surface fluorescence (chlorophyll) and temperature over the range of 10 to 290 km (Yoder et al., 1993). Also, “Meddies,” which are apparently relatively deep features (>500 m) with diameters of roughly 20 km and thicknesses of ~200 m, have been observed in the North Atlantic Ocean and the Sargasso Sea in particular. These are now thought to originate not only from the Mediterranean Sea, but also the northwest corner of the North Atlantic Current (Prater and Rossby, 1999). Meddies and similar features are difficult to observe and model; yet, they may well play important roles because of their physical dynamics, propensity to reduce water stability, and anomalous water mass and nutrient characteristics. Data assimilative models are useful for describing and predicting variability at the mesoscale that is often inaccessible via observations or most numerical models (note that even eddy-resolving models likely do not always capture salient features).

“Nonpurely” eddy processes (e.g., shear-induced mixing, fronts, jets, meanders, and warm outbreaks) are important as well as biological processes (species succession, phytoplankton physiology, photoadaptation, etc.), some with quite short time scales (minutes to hours) as suggested by Granata et al. (1995) and Lewis (2002). Temporal decorrelation times at mid-latitudes have been found to be in the range of ~5–15 days for the key physical and biological variables (Dickey et al., 2001). These time scales are consistent with passages of synoptic-scale weather patterns and mesoscale eddies. In addition, transient forcing conditions leading to stratification or mixing are often reflected in phytoplankton bloom and bust sequences (Dickey et al., 1994, 1998a, 2001). Several examples of physical and biological processes contributing to observed variability are presented in Dickey and Falkowski (2002). Interestingly, a recent modeling study examining biological and physical variability in the central Arabian Sea has indicated that relatively high-frequency processes such as diurnal cycling are important for simulating observed phytoplankton biomass and primary production (McCreaey et al., 2001). This general result is supported by previous observational work as well (e.g., Dickey et al., 1994; Stramaska and Dickey, 1998). There is increasing observational evidence that nonsteady-state, nonequilibrium processes may be disproportionately important for explaining ecological and biogeochemical variability (e.g., Dickey et al., 1998b; Honjo et al., 1999; Conte et al., submitted for publication). Thus, high-frequency, high-spatial resolution, long-term, interdisciplinary observations through the water column are essential. Coupled biological–physical models are capable of simulating some, but not all, of the aforementioned processes.

An important point of agreement seems to be that observational methods and models have not been adequate to allow us to solve the biogeochemical mesoscale/submesoscale problem as evidenced by continuing controversies (e.g., Dickey et al., 2001; Garcon et al., 2001; McGillicuddy, 2001; Oschlies, 2001b, 2002a,b; Lewis, 2002; Conte et al., submitted for publication). Comprehensive, coordinated studies utilizing a full complement of in situ and remote sensing sampling assets along with data assimilation modeling are desperately needed (e.g., Robinson and Dickey, 1997; Doney, 1999; Doney et al., 2001; Lewis, 2002). These can also be valuable for developing model parameterizations of unresolved physics, ecology, and biogeochemistry, and for constraining regional-, basin-, and global-scale estimates of gross primary and new primary production and nutrient and carbon fluxes.

7. General challenges and opportunities

In this final section, we summarize some general methodological and technical challenges and follow with a brief summary of some key opportunities for the future.

7.1. Methodological challenges

First, it is worth noting the challenges of data assimilation and, specifically, inherent limitations in predictability in the form of nonlinear error propagation and error-scale transfers (Robinson and Lermu-
siaux, 2001). Other, perhaps less daunting, observational and modeling problems remain for future work. For example, remote sensing of many physical and biogeochemical variables is most attractive for reasons described earlier. Unfortunately, many of the more important biological and chemical variables cannot be measured remotely, even at the surface at present, and cloudy conditions negate ocean color (and some temperature) data retrievals sporadically and regionally. In addition, measurements and models of the ocean at depth are vital because of the considerable vertical structure and variability, which cannot be determined from remote sensing.

Another methodological problem concerns sampling of episodic events, some of which may contribute substantially if not dominantly to variance in physical and biogeochemical properties. Fortunately, there will likely be more platforms available in the future for providing high temporal and spatial resolution data at representative ocean sites; data obtained from these locations will be valuable for data assimilation systems (see Figs. 2–6 and 9). While the number of variables needed for open ocean studies is typically much smaller than for the coastal ocean, many important biological and chemical variables and rates remain difficult to measure from autonomous in situ and remote sensing platforms. Data assimilative modeling using data collected from all of the available platforms (Figs. 4, 5, and 9) will be needed to characterize, quantify, and model the processes depicted in Figs. 1 and 2a. Another nagging issue concerns the constraint of using combined multiyear data sets for assimilation of an annual cycle. This has been necessitated by lack of nearly continuous (high frequency), long-term measurements of relevant variables. We are fast approaching a time when this should not be the case as past technical limitations (e.g., time-space sampling continuum is filled in for more variables) are removed as obstacles to progress (e.g., Dickey, 2001, 2002; Dickey and Falkowski, 2002).

7.2. Technological challenges

There remain several key areas for future technological work. For example, most of the in situ and remote sensing systems described here were not designed for sampling within 30-m perpendicular distance of the shoreline, one of the most important ocean zones for direct human interaction. Propelled AUVs are still power-limited (and thus duration- and distance-limited); hence, fuel cell advances are needed. Biofouling of many of the in situ sensors is still a problem for longer duration sampling, especially in coastal waters. Satellite observation of higher trophic level organisms also remains a major research goal.

Interpretation of many of the signals is problematic, and intensive cross-sampling and intercalibrations are critically needed. Data assimilation models also need to have error estimates of measurements (e.g., Bennett, 1992; Wunsch, 1996; Evans, 2003). The problem of estimating errors from data sets and models and ultimately assessment of data assimilation uncertainties remains a major challenge that requires interactions among observationalists and modelers. Technology is moving so rapidly that many of the measurement systems will become obsolete quite quickly. However, continuity of standardized and well-calibrated measurements must be hallmarks of ocean observing systems intended to quantify abrupt and long-term change. Many of the new sensors, systems, and platforms described above are presently in developmental phases and in the hands of only a few researchers. Mass production of instruments and platforms must be done by the commercial sector to substantially increase ocean observations on a global scale.

There are presently relatively few opportunities for telemetering data from the open ocean. Some communication cables may be used in specific locations and under special circumstances; however, satellite communication systems must be relied upon in general. As indicated earlier, data bandwidth remains limiting (e.g., see Dickey et al., 1993, 1998b for more details) and new satellite communications systems (like Iridium) are badly needed for virtually all sampling assets at sea. However, new data compression methods may alleviate some of the constraints (e.g., Davis et al., 1999). In all cases, cost of transmission of data must be considered and may be prohibitive for some applications. Nonetheless, telemetry of data will be critical for predictive data assimilation applications. Rapid communication among those parties responsible for remote sensing and in situ measurements and data assimilative models will likely require use of advanced satellite data communication networks.
Finally, the utility of interdisciplinary oceanographic data and model products will depend on their presentation to scientists as well as the general public. Major advances are being made in the visualization of observational data and model results (e.g., see examples in Djurcilov et al., 2001; Dickey, 2002). Meteorologists have led the way in creating both still and motion picture graphics (some with three-dimensional visualizations) that enable scientists and the public to optimally utilize complex information and to better understand weather and atmospheric dynamics and thermodynamics. Ocean color images from space are now presented via the World Wide Web, and in the future, four-dimensional motion pictures or loops of physical, optical, bio-optical, biogeochemical, biological, and geological fields and their predictions can be added.

7.3. Opportunities

Large numbers of sensors are being developed for many purposes; some may be directly or indirectly useful for ocean measurements. One possible benefit of advanced sensor technologies may well be more capable, smaller, simpler and less expensive sampling systems (e.g., “chip-based” chemical and biological sensors with automated data processing; Kaku, 1997), which will benefit oceanographic programs regardless of present technical capabilities and skill levels. There is increasing interest in making platforms available (e.g., platforms-of-opportunity) for a broader suite of interdisciplinary measurements. Also, there is growing support for systems approaches that incorporate data acquisition, data telemetry, data assimilative modeling, prediction, and capability for adaptive sampling through redirection of assets.

There is now a strong international focus on global ocean observing systems for both the coastal and open oceans. These developments will likely reflect in improved sampling systems that will provide more variables and better spatial and temporal coverage (e.g., international partnering for Argos and time series observatories programs). Importantly, data assimilation modeling is considered a key element of these systems (as described by Le Traon et al., 2001; GODAE program), and thus observational and modeling components are being developed synergistically. Joint instrumentation/data assimilation system test-beds should be initiated as soon as possible to accelerate progress in this direction (e.g., Robinson and Dickey, 1997). There remain a myriad of challenges to bring interdisciplinary data assimilation systems to maturity; however, the oceanographic community is in a very strong position to make major advances because of burgeoning technologies and increasing public concern for the ocean environment and climate.

Acknowledgements

I would like to thank Allan Robinson for introducing me to data assimilation and working with me for several years in promoting sampling and observing systems in the context of interdisciplinary data assimilation systems. This paper was stimulated by an invitation from Pierre Lermusiaux and Marielle Grégoire to review observational capabilities at the Liege Colloquium on Interdisciplinary Data Assimilation. Numerous collaborators have contributed ideas that have influenced this paper. Special thanks for information and discussions are extended to Scott Doney, Marjy Friedrichs, Scott Glenn, Eileen Hoffman, Marlon Lewis, Dennis McGillicuddy, and Oscar Schofield. Pierre Lermusiaux deserves special credit for providing extensive input and for carefully reviewing this paper. Nadia Pinardi graciously provided several recent references. Geoff Evans is also thanked for his thoughtful review. This work was supported by the following agencies and programs: U.S. National Ocean Partnership Program [Littoral Ocean Observation and Prediction System (LOOPS) and Ocean-Systems for Chemical, Optical, and Physical Experiments (O-SCOPE)], National Science Foundation (Ocean Technology and Interdisciplinary Coordination Program, Chemical Oceanography, Oceanographic Instrumentation and Technical Services), Office of Naval Research (Ocean Optics), and the National Atmospheric and Space Administration. This is U.S. JGOFS contribution number 805.

References


Baretta, J.W., Ebenhof, W., Ruardij, P., 1995. The European Re...


