See also

Acoustics, Deep Ocean. Acoustic Scattering by Marine Organisms. Baleen Whales. Marine Mammal Social Organization and Communication. Seals. Sea Otters Sirenians. Sonar Systems. Sperm Whales and Beaked Whales.

Further Reading

Au WWL (1993) *The Sonar of Dolphins*. New York: Springer Verlag.

Au WWL, Popper AS and Fay R (eds) (2000) *Hearing by Whales and Dolphins*. Springer Handbook of Auditory Research Series. New York: Springer Verlag.

Burgess WC, Tyack PL, LeBoeuf BJ and Costa DP (1998) A programmable acoustic recording tag and first results from free-ranging northern elephant seals. *Deep-Sea Research* 45: 1327–1351.

Kastak D and Schusterman RJ (1998) Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology. *Journal of the Acoustical Society of America* 103: 2216–2228.

Medwin H and Clay CS (1998) *Fundamentals of Acoustical Oceanography*. New York: Academic Press.

Miller P and Tyack PL (1998) A small towed beamforming array to identify vocalizing resident killer whales (*Orcinus orca*) concurrent with focal behavioral observations. *Deep-Sea Research* 45: 1389–1405.

Rayleigh, Lord (1945) *The Theory of Sound*. New York: Dover.

Tyack P (1998) Acoustic communication under the sea. In: Hopp SL, Owren MJ and Evans CS (eds) *Animal Acoustic Communication: Recent Technical Advances*, pp. 163–220. Heidelberg: Springer Verlag.

Tyack PL (2000) Functional aspects of cetacean communication. In: Mann J, Connor R, Tyack PL and Whitehead H (eds) *Cetacean Societies: Field Studies of Dolphins and Whales*, pp. 270–307. Chicago: University of Chicago Press.

Wartzok D and Ketten DR (1999) Marine mammal sensory systems. In: Reynolds JE III and Rommel SA (eds) *Biology of Marine Mammals*, vol. 1, pp. 117–175. Washington, DC: Smithsonian Press.

BIOGEOCHEMICAL DATA ASSIMILATION

E. E. Hofmann and M. A. M. Friedrichs,
Old Dominion University, Norfolk, VA, USA

doi:10.1006/rwos.2001.0410

Introduction

Data assimilation is the systematic use of data to constrain a mathematical model. It is assumed that the dynamics that are responsible for a particular process or distribution are inherent in the data. By inputting data of various types into a mathematical model, the model, which is a truncated version of the real world, will more accurately stimulate a particular environment or situation. Through data assimilation, the hindcast, nowcast, and/or forecast of the model will be improved.

Data assimilation was first used in the 1960s in numerical weather forecasting models, with the goal of providing short-term predictions of meteorological conditions. The use of data assimilation techniques was made feasible by the development of a worldwide atmospheric data network that could provide the measurements needed. Data assimilation provided a methodology for using these observations to improve the forecasting skills of the operational models. Although weather forecasts are now taken for granted, to a large extent the accuracy of these forecasts results from assimilation of meteorological observations.

In the 1970s, numerical ocean general circulation models (OGCMs) became an important tool for understanding ocean circulation processes. Initial applications of these models focused on simulation of the large-scale structure of ocean currents. From these simulations, the limitations of the OGCMs were clear. Data assimilation was looked to as an approach for constraining these dynamical models with available data. For example, data assimilation could be used to quantitatively and systematically test and improve poorly known sub-grid-scale parametrizations and boundary conditions that are so abundant in OGCMs. With recent advances in data availability, it is also now feasible to use data-assimilative OGCMs for making forecasts of the ocean state, such as the El-Niño–La Niña cycle in the equatorial Pacific Ocean.

Implementing data-assimilative biogeochemical models has been problematic because of the paucity of adequate data. Historically, biological and chemical data were obtained almost exclusively by ship surveys, and thus were extremely limited in both space and time. However, advances in satellite and mooring instrumentation, as well as in the understanding of the structure and function of marine ecosystems, now makes it feasible to begin the development of data-assimilative biogeochemical models. As a result, since the mid 1990s there has been a dramatic increase in the types of data that are input into marine ecosystem models, and the development of robust and varied approaches for assimilating these data. This research provides a framework for future studies of biogeochemical data assimilation and predictive biogeochemical modeling that will inevitably play a major role in the next generation of large interdisciplinary oceanographic observational programs.

The following section provides a brief history of how the field of marine biogeochemical modeling has matured as more and more data have become available. This is followed by a description of some data assimilation methods and specific examples of how two of these methods can be used in conjunction with a simple marine ecosystem model. The final section provides a summary.

Biogeochemical Models and Data Availability

Mathematical models provide a quantitative framework for investigating processes that are responsible for the biological and chemical distributions that underlie the structure and function of marine ecosystems. Mathematical models were first used to study marine ecosystems in the late 1940s and these models had their basis in the predator–prey models developed in the early 1900s. These early modeling attempts were focused on understanding the processes that allow large blooms of phytoplankton and zooplankton to occur. The models were simple in nature, including only average population characteristics, basic biological processes resulting in plant and animal growth, and interactions at the lowest trophic levels, e.g., primary and secondary producers. Effects of environmental factors such as temperature and circulation, which are important in marine systems, were not explicitly included in these models.

The following generation of models included more complex biogeochemical processes, differentiation of species, and coupling of the marine biogeochemical system to circulation models. These more realistic models were made possible by the development of large multidisciplinary oceanographic programs in the 1970s, which, for the first time, provided concurrent physical, biological, and chemical measurements that could be input to coupled circulation–biogeochemical models. The resulting models clearly demonstrated the utility of modeling for integrating and synthesizing large multidisciplinary oceanographic datasets. However, more importantly, the realism of the simulated distributions obtained from the coupled modeling efforts helped establish this approach as an important research tool for understanding marine biogeochemical systems. Present-day multidisciplinary oceanographic programs now routinely include a mathematical modeling component.

At the time coupled circulation–biogeochemical models were being developed, significant advances were being made in the measurements of biological and chemical distributions in the ocean. In the 1980s, the Coastal Zone Color Scanner satellite was launched, which provided large-scale ocean color distributions, from which phytoplankton chlorophyll distributions, and their evolution over space and time, could be derived. The ocean color data also facilitated making inferences about the relative roles of circulation versus biogeochemical processes in controlling phytoplankton distributions. The availability of large-scale observations of chlorophyll has been enhanced with the subsequent launch of the Sea-viewing Wide Field-of-view Sensor in 1997. Instrumentation capable of providing biological measurements at fine space and time scales, e.g., moored optical and acoustic measurements, now provide *in situ* observations that can be combined with ocean color to reveal a more complete view of chlorophyll distributions. Also, moored buoy arrays provide concurrent physical data. Thus, high-resolution datasets that can be used to study marine systems are becoming increasingly available.

These new, high-quality datasets can now be used to validate coupled circulation–biogeochemical models. In many cases simulated distributions from the models reproduce many of the features seen in the ocean color observations. However, the simulated fields are often unable to reproduce the variability observed on short space (tens of kilometers) and time (days) scales. There are many potential explanations for these discrepancies, including mismatches in the space and time scales that the model resolves versus those resolved by the measurements. There may also be inconsistencies between the model structure and the observations. For instance, specific parameter values or choices for empirical formulations, forcing functions, or initial conditions may be in error.

Along with these new types of data comes the possibility of developing a new generation of data-assimilative biogeochemical models that will not only be better able to reproduce the observed variability in biogeochemical fields but may also have the potential for significantly improving the accuracy of model predictions. In the 1990, researchers began to investigate the use of data assimilation as an approach for improving these coupled models, and as an approach for making better use of the many types of environmental and biogeochemical data that are becoming available.

Data Assimilation Methods

Many techniques exist for systematically combining data with mathematical models. The development and refinement of many of these techniques have been through the use of meteorological and oceanic general circulation models. These data assimilation methods are just starting to be tested in marine biogeochemical models; however, it is not clear whether the same methods can be used with these multidisciplinary models, since biogeochemical ocean models differ substantially from their physical counterparts. For instance, biological systems have no analogue to the Navier–Stokes equations that form the basis for fluid dynamics. Thus, biogeochemical models are by necessity largely empirical and nonlinear, and abound with poorly known formulations. For example, such models typically include large numbers of parameters that are difficult (*in situ* growth rates), or even impossible (mortality rates) to measure with current oceanographic instrumentation. The timescales of these models are also typically short, since the model must resolve the rate at which populations double in number, which for most of the abundant phytoplankton species is one day or less.

Because of these innate differences between physical models and biogeochemical models, the application of data assimilation techniques to biogeochemical ocean models, and specifically to marine ecosystem models, presents many exciting new challenges. Although considerable effort will undoubtedly be put into developing new assimilation schemes specific to these types of models, the data-assimilative marine ecosystem models that have been developed using existing assimilation methods already show much promise.

One of the most straightforward methods that has been used to combine model dynamics with data entails simply replacing the model solution with data whenever such information is available. This technique, referred to as data insertion, integrates the model forward in time until additional observations become available, at which point the model is reinitialized and the process is repeated. A basic assumption underlying this method is that there is adequate knowledge of the governing model dynamics and parameter values.

This technique has been used to estimate velocity fields by inserting temperature and salinity data into relatively complex physical oceanographic models. In these analyses, however, model–data inconsistencies caused the resulting simulations to compare poorly with observations. This led to the development of a technique in which the model solution is 'nudged' toward observations whenever they become available, instead of being directly replaced by the observations. Although this more gentle method of nudging may provide a significant improvement in simulation skill, like data insertion, it still lacks a means by which information on data uncertainty can be incorporated, and does not provide an estimate of the errors in the resulting solution.

More advanced assimilation schemes, such as optimal interpolation and Kalman filtering, have been successfully applied by meteorologists, yet hold little hope for marine ecosystem models because of the inherent nonlinearities of biological systems. Instead, variational schemes, which have recently been applied to nonlinear physical oceanographic systems, may be more applicable to multidisciplinary problems containing biological components. These variational methods of data assimilation, such as the adjoint method and simulated annealing, have their basis in optimization theory and rely on minimizing the differences between observed and simulated quantities, pursuant to predetermined minimization criteria. At the most basic level, these methods can be thought of as nonlinear least-squares analyses, which determine the optimal solution (including parameter values and initial and/or boundary conditions) that maximizes agreement between the model simulation and observations.

The adjoint method is a variational scheme that has found considerable success in the field of physical oceanography. Although this method is now also being used in marine ecosystem modeling, the nonlinear nature of these types of models may result in the recovery of suboptimal parameter sets. Simulated annealing is another assimilation scheme that has been used with data-assimilative ecosystem models. Although this method is typically capable of recovering a single optimal parameter set, the stochastic, 'random-walk' nature of simulated annealing causes this technique to be considerably less efficient than the adjoint method. As a result, simulated annealing may be computationally too intensive to be of use in large-scale marine biogeochemical assimilation analyses.

Numerical Twin Experiments

Before the application of data assimilation techniques to marine biogeochemical models becomes routine, a number of methodological questions need to be addressed. For instance, how many data are needed for these studies? What types of sampling strategies are optimal? What level of uncertainty in the data can be tolerated? Are these models too complex, or too simple? What types of data assimilation schemes will work best for these highly nonlinear models?

One method for addressing such methodological questions is through the use of identical twin experiments. In an identical twin experiment, the model is initially run using best estimates for the model parameters in order to provide a 'true' simulated time-series. This time-series is subsampled to generate a synthetic data set (**Figure 1A**). The model is then run a second time, using an imperfect parameter set in order to generate a 'reference' (no assimilation) time-series. This same imperfect parameter set is used in the third and final model run, but this time the synthetic data are assimilated into the model (**Figure 1B**). The success of the assimilation process is judged by the difference between these results and the true simulation, and is typically normalized by the difference between the true and reference simulations.

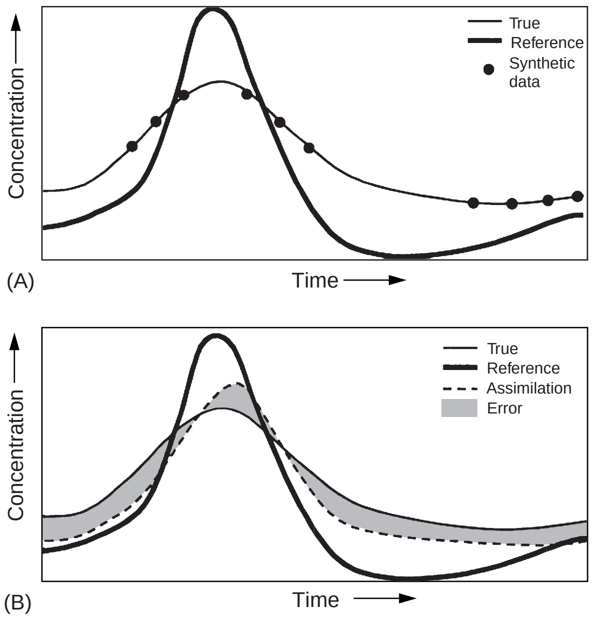

Figure 1 Schematic illustrating the implementation of an identical twin experiment. The true simulation (thin solid line) represents the solution for one component of the model (e.g., phytoplankton, zooplankton, or nutrient concentration) obtained using the best estimates of the model parameters and initial conditions. A second simulation, using a different parameter set or initial conditions, provides a reference simulation (thick solid line). (A) The true simulation is subsampled to create a synthetic data set. (B) The assimilation of the synthetic data into the reference simulation results in a third model solution (dashed line). The difference between this solution and the true solution (shaded region) is a measure of the error in the data-assimilative solution.

Identical twin experiments are a necessary precursor to true data-assimilative model runs, and have the potential to provide considerable insight into a number of important issues regarding the assimilation process. For instance, they can be used to rigorously compare different assimilation schemes, to determine optimal sampling strategies, and to assess the effects of assimilating observations that are associated with known levels of noise. Furthermore, identical twin experiments can be invoked to determine whether a certain set of model parameters can be estimated independently, and thus whether or not a given model may need to be simplified. Although the utility of identical twin experiments is well accepted within the fields of meteorology and physical oceanography, this approach has only recently been applied to ecosystem modeling analyses. The two examples described below illustrate some of the strengths and weaknesses of this approach for understanding data-assimilative marine biogeochemical models.

Example 1: Data Insertion and Nudging

The pros and cons of using data insertion or 'nudging' to assimilate biogeochemical data can be illustrated using a three-component marine ecosystem box model (nitrogen, phytoplankton, zooplankton). Simulations with a non-data-assimilative version of this model (**Figure 2**) show the behavior that is expected in this type of marine system. Nitrogen concentrations decrease over time, as nitrogen is used to support a bloom of phytoplankton. Zooplankton, the primary grazer of the phytoplankton, blooms subsequently and results in a decrease of the phytoplankton.

The accuracy of this basic simulation can be improved by inserting phytoplankton concentrations, such as those derived from ocean color measurements. Results of an identical twin experiment

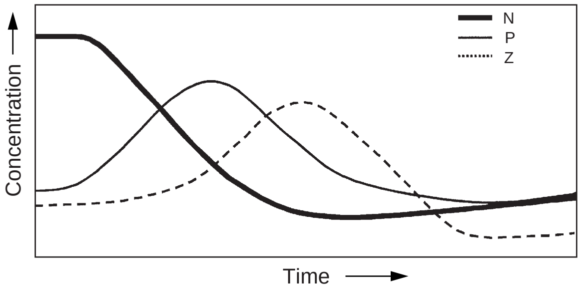

Figure 2 Schematic of the time evolution of nutrient (N), phytoplankton (P), and zooplankton (Z) concentrations, obtained from a marine ecosystem model simulation.

demonstrate that at times when average chlorophyll concentrations are inserted into the box model (e.g., every other day in **Figure 3A**), the error in the simulation of phytoplankton decreases to zero. Nudging yields similar results, except that the error in the phytoplankton field would be reduced to a fixed nonzero value, dependent upon the strength of the nudging.

One requirement for data insertion or nudging is that data must be available on timescales coincident with those of the dominant biological processes. Because biological processes, such as phytoplankton growth, have timescales of 1–2 days in many regions of the ocean, data with this level of time resolution are required for data insertion methods to adequately represent the biological dynamics. As illustrated in **Figure 3A**, the improvement in simulation skill lasts only 1 to 2 days beyond the point at which phytoplankton data are no longer available for insertion. Hence, when using these methods, fully data-assimilative marine biogeochemical models can potentially create a huge demand on data resources.

Another primary factor limiting the use of data insertion methods for marine biogeochemical models is that these models usually consist of many components, all of which must be updated to be in balance with the assimilated data. For instance, although estimates of phytoplankton biomass are improved as a result of the assimilation of phytoplankton data, the accuracy of the other model components is reduced (**Figure 3A**). The error produced in the other model state variables is variable, and depends upon the level of coupling between the various components of the marine biogeochemical system. This problem may be alleviated if data exist for the other model components, since the assimilation of these additional data (**Figure 3B**) can substantially reduce errors in the other model components. Since the adjustment timescales for the model components differ (**Figure 3B**), the insertion of data may be required at varying time intervals for the different model components. Unfortunately, data sufficient to update the other ecosystem model components often do not exist. In this case *ad hoc* approaches, perhaps based on maintaining ratios between different ecosystem components, can be invoked; however, such approaches also have the potential to introduce errors that may negate the gains made through data assimilation.

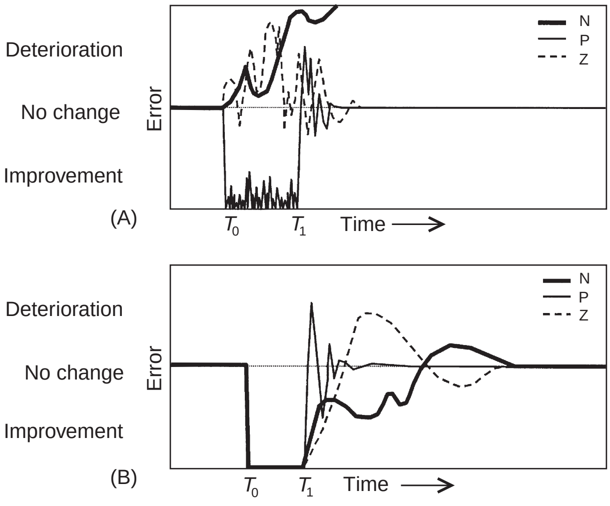

Figure 3 Schematics of the change in model solution error obtained in an identical twin experiment, when data are inserted into the nutrient–phytoplankton–zooplankton model (**Figure 2**) between $t = T_0$ and $t = T_1$. (A) Only phytoplankton (P) observations are inserted every other day, and (B) nutrient (N), phytoplankton, and zooplankton (Z) observations are inserted daily. Deterioration represents movement of the data-assimilative solution farther from the true solution; improvement represents convergence of the data-assimilative and true solutions. No change occurs when the data-assimilative model solution remains the same as in the non-data-assimilative (reference) solution.

Thus, data insertion and nudging are easy to implement and in certain instances may improve the accuracy of biogeochemical model predictions. However, many issues remain to be addressed before this method can be used successfully for data-assimilative biogeochemical models. For instance, because data insertion assumes that the input data sets are perfect representations of the real world, model–data inconsistencies can be magnified and can cause model solutions to become dynamically unbalanced. This is especially a problem for simulations of systems in which the circulation is the dominant control on the biogeochemical distributions. Perhaps most importantly, however, neither data insertion nor nudging readily lends itself to improving model parametrizations or model structure.

Example 2: Adjoint Method

Even a relatively simple marine ecosystem model, such as the model shown in **Figure 2**, typically contains 10–20 model parameters that must be specified for a given simulation. This is a crucial aspect of ecosystem modeling, since even small changes in some of these parameters may result in large differences in simulation results. Unfortunately, values for these parameters are often poorly constrained in space and time, and some, such as *in situ* zooplankton mortality rates, are virtually unknown. Thus, the specification of an optimal parameter set for a given biogeochemical model is a challenge at best, and in most cases, nearly impossible. For these reasons, the adjoint method, which searches parameter space to find a parameter set that minimizes model–data misfits, holds considerable potential for use in data-assimilative marine biogeochemical models.

The utility of the adjoint method can be demonstrated by applying identical twin experiments to the three-component ecosystem model (**Figure 2**). Because marine biogeochemical models typically contain many parameters that are very highly correlated, it is often not possible to recover a parameter set in its entirety. Therefore, sensitivity or correlation analyses can be performed in order to choose a subset of relatively uncorrelated parameters that will be recovered.

If an identical twin experiment is carried out in which phytoplankton data are assimilated every other day using the adjoint method, phytoplankton and zooplankton parameters may be recovered precisely (**Figure 4A**). As a result, the errors in the phytoplankton and zooplankton simulations will be significantly reduced. (In **Figure 4B** the error in the phytoplankton is shown to be zero for the entire model run). However, if no nutrient data are assimilated, the parameter(s) on which nutrient concentration is most highly dependent, e.g., N_1 (**Figure 4A**), will not be recovered, and therefore no significant improvement in the simulation skill of the nutrient component will result (**Figure 4B**). If synthetic phytoplankton, zooplankton, and nutrient data are all available for assimilation, all parameters may be recovered precisely, and the errors in all model components may be reduced to zero for the entire model run.

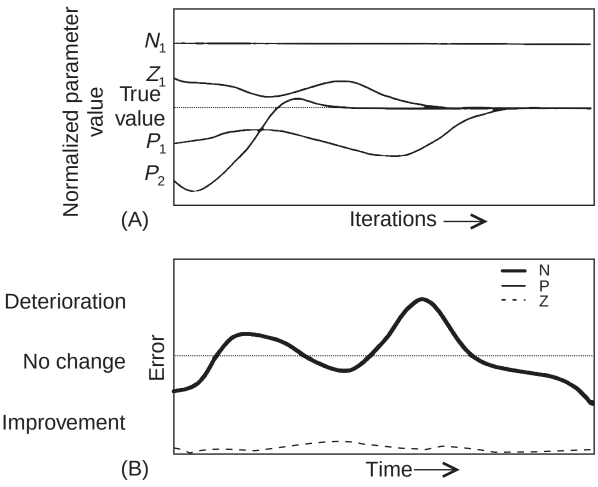

Figure 4 Schematics illustrating the results of using the adjoint method to assimilate phytoplankton data into the nutrient–phytoplankton–zooplankton model (**Figure 2**) in an identical twin experiment. (A) The phytoplankton (P_1, P_2) and zooplankton (Z_1) parameters converge to their true values, but the nutrient parameter (N_1) cannot be recovered without the assimilation of nutrient data. (B) Model solution error is greatly reduced for both phytoplankton (P) and zooplankton (Z), but not nutrient (N) concentrations. Deterioration represents movement of the data assimilative solution farther away from the true solution; improvement represents convergence of the data assimilative and true solutions. No change occurs when the data-assimilative model solution remains the same as the non-data-assimilative (reference) solution.

Research on data-assimilative marine ecosystem models has shown that under certain conditions the results described above are characteristic of those obtained when real data are assimilated using the adjoint method. However, in other instances it is possible that the assimilative model may fail to recover an optimal parameter set. This can occur even if the model has been well tested and calibrated, and implies that the model is in some way inconsistent with the assimilated data set. For instance, changes in plankton community dominance might result in inconsistencies in the model and data that cannot be resolved simply through data assimilation. If this is the case, it may be possible to isolate the specific model assumption(s) that have been violated, e.g., the assumption of a constant species composition, to reformulate the model in a more realistic fashion and to repeat the assimilation analysis in order to test this hypothesis.

Sometimes the adjoint method may recover multiple parameter sets, each dependent on the initial choices made for the model parameters. In these situations, rigorous approaches for choosing between the possible parameter sets are required. One approach is to establish a specific uncertainty range, either from experimental or from theoretical considerations, for each parameter that is allowed to vary in the adjoint analysis. Alternatively, the optimal parameter set could be selected on the basis of the ability of each parameter set in simulating an independent data set.

Summary

Data assimilation techniques for marine biogeochemical models are just beginning to be explored. Initial results are encouraging and data assimilation approaches, such as adjoint methods, hold great promise for improving the capability of these models. For instance, recent analyses of data-assimilative biogeochemical models demonstrate that the assimilation of biogeochemical data can reduce model–data misfit by recovering optimal parameter sets using multiple types of data. Perhaps even more importantly, these data assimilation analyses can demonstrate whether or not a given model structure is consistent with a specific set of observations.

When model and data are shown to be consistent, the specific mechanisms underlying observed patterns in simulated distributions can be identified. If a model is determined to be inconsistent with observations, it may be possible to isolate the specific model assumption that has been violated, and to reformulate the model in a more realistic fashion. Thus, although the assimilation of data into a marine biogeochemical model cannot necessarily overcome inappropriate model dynamics and structure, it can serve to guide model reformulation.

During the 1990s, large interdisciplinary oceanographic programs included model prediction and forecasting as specific research objectives. However, new studies are revealing that much more work needs to be performed before this becomes a realistic and achievable goal. Until high-resolution biological and chemical data are available over large regions of the ocean, and until a much clearer understanding of the intricacies of marine ecosystems is attained, data assimilation in biogeochemical models will be more useful for model improvement and parameter estimation than for model prediction and forecasting. By providing a means for recovering the best-fit set of parameters for a given model, certain assimilation techniques may prove to be a crucial tool for marine biogeochemical modelers.

The importance of inclusion of data in all steps of model development and implementation cannot be emphasized enough. It is through model and data comparisons that models are advanced and better observation systems are developed. Therefore, an important aspect of furthering the development of predictive marine biogeochemical models is recognizing the need for interdisciplinary multiscale observational and experimental networks. The availability of such data will necessitate the development of techniques for input of these data into models, and facilitate the development of data-assimilative marine biogeochemical models.

See also

Data Assimilation in Models. El Niño Southern Oscillation (ENSO) Models. Forward Problem in Numerical Models. Inverse Models. Inherent Optical Properties and Irradiance. Moorings. Ocean Color from Satellites. Population Dynamics Models. Primary Production Processes. Regional and Shelf Sea Models.

Further Reading

Fasham MJR and Evans GT (1995) The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47°N 20°W. *Philosophical Transactions of the Royal Society of London*, *Series B* 348: 203–209.

Ishizaka J (1993) Data assimilation for biogeochemical models. In: Evans GT and Fasham MJR (eds) *Towards a Model of Ocean Biogeochemical Processes*, pp. 295–316. New York: Springer-Verlag.

Lawson LM, Spitz YH, Hofmann EE and Long RB (1995) Data assimilation applied to a simple predator–prey model. *Bulletin of Mathematical Biology* 57: 593–617.

Matear RJ (1995) Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at Station P. *Journal of Marine Research* 53: 571–607.

BIOLUMINESCENCE

P. J. Herring, Southampton Oceanography Centre, Southampton, UK
E. A. Widder, Harbor Branch Oceanographic Institution, Fort Pierce, FL, USA

doi:10.1006/rwos.2001.0200

Introduction

Bioluminescence is the capacity of living organisms to emit visible light. In doing so they utilize a variety of chemiluminescent reaction systems. It has historically been confused with phosphorescence and the latter term is still frequently (and erroneously) used to describe marine bioluminescence. Some terrestrial species (e.g., fireflies) have the same ability, but this adaptation has been most extensively developed in the oceans. Bioluminescent species occur in only five terrestrial phyla, and only in one of these (Arthropoda, which includes the insects) are there many examples. In contrast, bioluminescence occurs in 14 marine phyla, many of which include numerous luminescent species (**Table 1**). All oceanic habitats, shallow and deep, pelagic and benthic, include bioluminescent species, but the phenomenon is commonest in the upper 1000 m of the pelagic environment.