Adaptive Coupled Physical and Biogeochemical Ocean Predictions: A Conceptual Basis

P.F.J. Lermusiaux[1], C. Evangelinos[2], R. Tian[1], P.J. Haley[1], J.J. McCarthy[1], N.M. Patrikalakis[2], A.R. Robinson[1], and H. Schmidt[2]

[1] Harvard University, Cambridge, MA 02138, U.S.A.
[2] Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Abstract. Physical and biogeochemical ocean dynamics can be intermittent and highly variable, and involve interactions on multiple scales. In general, the oceanic fields, processes and interactions that matter thus vary in time and space. For efficient forecasting, the structures and parameters of models must evolve and respond dynamically to new data injected into the executing prediction system. The conceptual basis of this adaptive modeling and corresponding computational scheme is the subject of this presentation. Specifically, we discuss the process of adaptive modeling for coupled physical and biogeochemical ocean models. The adaptivity is introduced within an interdisciplinary prediction system. Model-data misfits and data assimilation schemes are used to provide feedback from measurements to applications and modify the runtime behavior of the prediction system. Illustrative examples in Massachusetts Bay and Monterey Bay are presented to highlight ongoing progress.

1 Introduction

A team of scientists and engineers is collaborating to contribute to the advancement of interdisciplinary ocean science and forecasting through an effective union of ocean sciences and information technologies, focusing on adaptive modeling and adaptive sampling for coastal physical-biogeochemical-acoustical processes. Accurate ocean predictions and optimized rapid responses are essential for many oceanic applications, including fisheries, pollution control, hazard management, and maritime and naval operations. Because of multiple uncertainties in oceanic measurements and dynamical models, these capabilities require assimilation of physical, biogeochemical and acoustical data into dynamical models. Importantly, the coastal environment can be highly variable and intermittent on multiple scales, and oceanic variables, parameters and interactions that matter vary in time and space. Thus, efficient models must evolve during predictions. This evolution occurs as new data is injected and assimilated into the prediction system. The data-model comparisons of the data assimilation process then also involve a direct feedback of the data to the models, forming a dynamic data-driven application system (DDDAS [1]). The focus of this paper is to present and illustrate a conceptual basis for such adaptive ocean modeling and prediction.

The advances in oceanic numerical models and data assimilation (DA, [2]) schemes of the last decade have given rise to interdisciplinary Ocean Observing and Prediction Systems (e.g. [3]) that are used in operational settings. The next generation of such systems will advance the interaction between simulation and measurement to a new level where the forecast application changes its runtime behavior to adapt to new measurements. Importantly, the data assimilation community [4] is starting to recognize the importance of this adaptation, from the correction of model biases to the multi-model data assimilation and automated evolution of model structures as a function of model-data misfits.

Automated objective adaptive modeling allows the optimal use of approximate models for rapidly evolving ocean dynamics. Presently, a model quantity is said to be adaptive if its formulation, classically assumed constant, is made variable as a function of data values. Both structural as well as parametric adaptation are possible. Physical adaptive modeling includes regime transition (e.g., well-mixed to stratified) and evolving turbulent mixing parameterizations. Biogeochemical adaptive modeling includes variations of biological assemblages with time and space (e.g., variable zooplankton dynamics, summer to fall phytoplankton populations, etc) and evolving biogeochemical rates and ratios. This is especially important because biogeochemical modeling is in its infancy and model uncertainties are very large. The adaptive component also greatly facilitates quantitative comparisons of competing physical or biogeochemical models, thus ultimately leading to better scientific understanding.

In what follows, Section 2 outlines several properties of ocean forecasting and DA schemes that are relevant for adaptive modeling. Section 3 describes the conceptual basis of our implementation of adaptivity. Section 4 illustrates some initial progress toward such dynamic data-driven systems. Section 5 concludes.

2   Adaptive Ocean Predictions and Data Assimilation

The presently employed interdisciplinary nowcasting and forecasting system is the Harvard Ocean Prediction System (HOPS, [3]). The DA scheme is Error Subspace Statistical Estimation (ESSE, [5]). The high level architecture involves the component encapsulation of binaries using XML [6]. HOPS is a portable and generic system that simulates the 4D ocean dynamics. It has been applied to many regions [3] and has provided accurate operational forecasts. With adaptive modeling, the data assimilated by HOPS corrects not only field estimates but also the forecast model itself, leading to a dynamic system. ESSE combines forecasts and data based on their errors. It captures the uncertainties that matter by focusing on the largest error modes. These dominant errors are evolved by an ensemble (Monte-Carlo) approach. The ESSE predictions of uncertainties in the forecast are central to adaptive modeling based on Bayesian system identification.

To carry-out adaptive ocean modeling based on dynamic data-driven schemes, one must account for specific properties of ocean predictions and data assimilation. These include: forecasting timelines, data transfers and processing, measurement models to link data to predictive models, dynamical adjustments, and multivariate error covariances. These properties are discussed next.
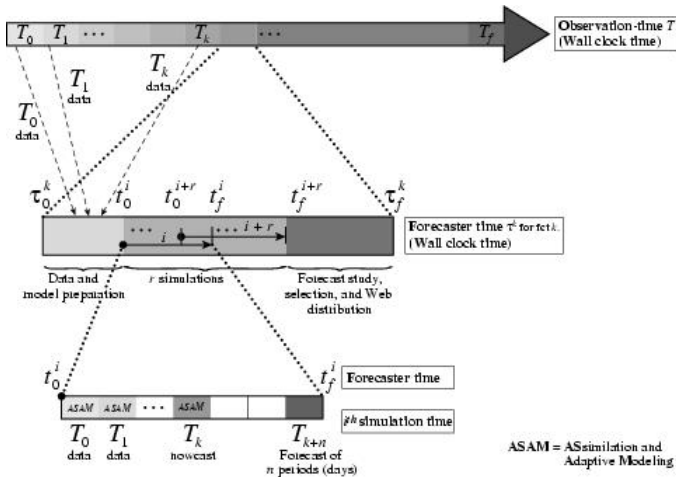
Fig. 1. Forecasting timelines. Top row: "Observation" or "ocean" time T during which measurements are made and the real phenomena occur. Middle row: "Forecaster" time τ^k during which the kth forecasting procedure and tasks are started and finished. Bottom row: "ith simulation" time t^i which covers portions of the real "ocean" time for each simulation. Multiple simulations are usually distributed on several computers, including ensembles of forecasts for uncertainty predictions (ESSE).

An important consideration for real-time adaptive modeling relates to the different times involved in ocean forecasting: the observation time, forecaster time and simulation time (Fig. 1). New observations are made available in batches (Fig. 1, first row) during periods T_k, from the start of the experiment (T_0) up to the final time (T_f). During the experiment, for each prediction k (Fig. 1, zoom in middle row), the forecaster repeats a set of tasks (from τ_0^k to τ_f^k). These tasks include the processing of the currently available data and model (from τ_0^k to t_0^i), the computation of $r+1$ data-driven forecast simulations (from t_0^i to t_f^{i+r}), and the study, selection and web-distribution of the best forecasts (from t_f^{i+r} to τ_f^k). Within these forecast computations, a specific forecast simulation i (Fig. 1, zoom in bottom row) is executed during t_0^i to t_f^i and associated to a "simulation time". For example, the ith simulation starts with the assimilation and adaptive modeling based on observations T_0, then integrates the dynamic model with data assimilation and adaptive modeling based on observations T_1, etc., up to the last observation period T_k which corresponds to the nowcast. After T_k, there are no new data available and the simulation enters the forecasting period proper, up to the last prediction time T_{k+n}.

Data-driven ocean applications involve measurement models which link the measured variables to the model state variables. These measurement models can be complex, e.g. for linking the measured acoustic travel times to the simulated fields of zooplankton concentrations and temperature. In addition, many of the state variables are usually not observed and inferring the non-observed variables by dynamical adjustment (e.g. [7]) prior to data assimilation is often crucial.

This reduces unphysical data shocks/adjustments that would otherwise occur after assimilation. With a fully multivariate DA scheme such as ESSE, forecast error covariances can infer the non-observed variables from the observed ones but it usually remains helpful to impose weak-constraint dynamical adjustments.

3   Implementing Adaptivity

We have been considering the following cases in our design thus far (Fig. 2):

1. running a single, adaptive interdisciplinary model,
2. running a single physical model, coupled to a set of competing biological models whose parameters are adapted,
3. running competing interdisciplinary models and adapting their parameters.

In all cases, the basis of the adaptation are the misfits between model estimates and data. When misfits are large, models are adapted. In the latter two cases, models can be rejected when seen as too inadequate. In the context of ESSE, expected bounds on misfits are given by the forecast uncertainties. Importantly, the computation of uncertainties for adaptive modeling can be based on small-size ESSE ensembles, one for each candidate models. This is feasible because error variances often converge faster than covariances. The adaptation triggers are provided to the software through external files, regularly checked at runtime and updated when data-forecast misfits warrant a specific model structure/parameter modification.

For example, in the first case (Fig. 2, left), the code may modify the carbon-to-chlorophyll ratio, increase the number of state variables (by adding meso-zooplankton) or alter the form of a source term. The latter two structural adaptations are cleanly implemented by using C function pointers to choose between Fortran modules. Physical and biological adaptation must be dynamically compatible but don't need to be computationally concurrent. In the second case (Fig. 2, right), the forecasts of competing biogeochemical models are compared based on their respective data-forecast misfits and the best biological functional forms selected for each time period T_i (Fig. 1). The third case is a combination of the former two.
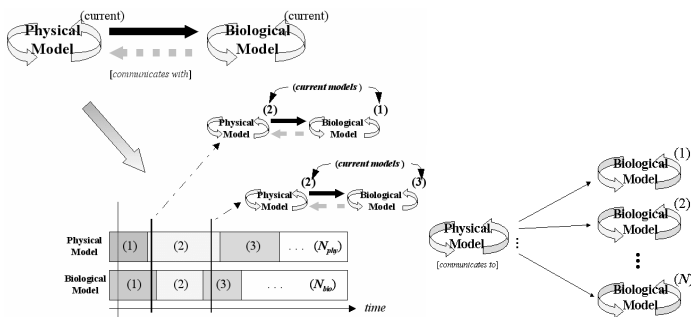

Fig. 2. Approaches for adaptive coupled physical-biogeochemical modeling

We are developing all three cases using a model of concurrent processes directly coupled (using MPI); we are considering the use of a more flexible coupling framework [8] for the future. Several computational challenges involved in the implementation of adaptivity for coupled physical-biological forecasts remain to be researched. They include the optimum operational choice between cases (i) to (iii), the impact of multiple observing and forecasting timelines, the interaction between parameter and structural adaptivity, the details of the quantitative schemes which trigger the adaptive switch/behavior, the relationships between adaptivity and consistent model re-initializations and data assimilation, the different computational meshes, and finally, the issues of communications costs and load imbalance in many-to-one scenarios.

4   Initial Progress and Prospectus

Three initial results are now presented. The first two are experiments which exercise the concept of adaptive modeling but do not yet utilize the whole dynamic data-driven computational system. The last result summarizes the development of a new adaptive biogeochemical computational model.

4.1   Biogeochemical Adaptive Modeling in Massachusetts Bay

The first example corresponds to dynamic data-driven predictions of coupled physical and biological dynamics in Massachusetts Bay during June 2001, as part of the Assessment of Skill for Coastal Ocean Transients (ASCOT-01) experiment. The goals of ASCOT-01 were to enhance the efficiency and extend the scope of nowcasting and forecasting of oceanic fields for Coastal Predictive Skill Experiments and Rapid Environmental Assessments. The limited physical dimensions of the Bay allowed a relatively comprehensive sampling and forecasting of transient conditions. The focus was on coastal ocean responses to wind events, including upwellings and subsequent advections of nutrients and organisms. The predictions illustrated by Fig. 3 were carried-out with the Harvard coupled models, initialization procedure and data assimilation schemes (see [7]).

In June (end of the Spring bloom), the ecosystem is usually in a different state than it is in Aug-Sept (summer/summer-to-fall transition). For June 2001,
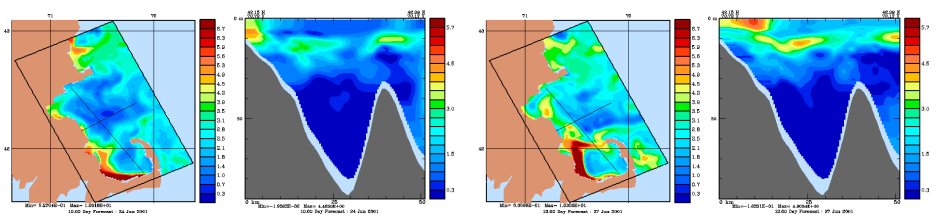

Fig. 3. Adaptive biogeochemical modeling during ASCOT-01 in Mass. Bay, June 2001. Two left panels: June 24 nowcast of surface Chl (mg/m^3) and corresponding west-east cross-section. Two right panels: June 27 forecast. All fields are after adaptation.

important state variables were the 4D fields of phytoplankton, zooplankton, detritus, nitrate, ammonium and chlorophyll-a (e.g. Fig 4), as in Aug-Sept 1998 [7]. However, model parameters needed to be modified. The parameters whose values where adapted to the June 2001 conditions included light attenuation scales, photosynthetic parameters, nutrient uptake constants, phytoplankton and zooplankton mortality rates, zooplankton grazing parameters, settling/sinking velocities for phytoplankton/detritus, and re-mineralization timescales.

Once the experiment started, model parameters were adapted to the new data collected, using the concept of DDDAS. The techniques utilized are illustrated on Fig 3 (right side, cases 2 and 3). Firstly, a set of physical-biological predictions was executed with a single physical model but different biological model parameters. The adaptive biological parameters were the: photosynthetic parameters, half saturation constants for ammonium/nitrate uptakes, zooplankton grazing parameters and zooplankton mortality rates. Each member of the set of predictions was then compared to the new data collected. The biological parameters of the coupled prediction which was the closest to this new data (within error bounds) were ultimately selected as the best parameters. This process was repeated several times during the experiment. Secondly, the representation of the Charles River outflow from Boston Harbor was adapted. This was necessary because the real-time data likely indicated that the outflow was larger in June 2001 than in August 1998. The physical strength of the simulated nitrate outflow was therefore increased. A sample of the nowcast and forecast chlorophyll concentration fields after adaptation are shown on Fig. 3. Note the effects of the southwesterly winds during the June 23-27 evolution (advection of Chl plume and increase in sub-surface maxima). Importantly, the adaptive modeling allowed the 2-3 day forecast of the Chl field to be better than persistence (assume oceanic conditions are constant) by about 5 to 20 percent.

4.2   Physical Adaptive Modeling in Monterey Bay

The second example corresponds to the Autonomous Ocean Sampling Network (AOSN-II) field experiment in the Monterey Bay region. It utilized a plethora of remote and in-situ sensors and platforms including multiple satellite images, drifters, gliders, moorings, AUV and ship-based data (www.mbari.org/aosn). These data were assimilated into numerical models and daily predictions of the ocean fields and uncertainties were issued. The different sensors provided correlated, sometimes redundant, measurements of the environment. Another important component was the multiple predictive models. In addition, multiple ESSE runs of the same model were carried-out, including stochastic forcings to represent model uncertainties. These ensembles of ESSE simulations produced plausible dynamical scenarios for the region.

Prior to the experiment, model parameters were calibrated to historical conditions judged to be similar to the conditions expected in August 2003. Once the experiment started, several parameters of the physical ocean model were adapted to the new 2003 data. This adaptation involved the parameterization of the transfer of atmospheric fluxes to the upper-layers of the sea. As shown on
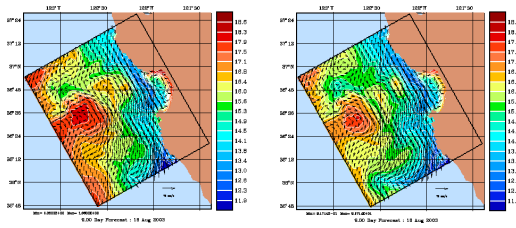
Fig. 4. Adaptive physical modeling during AOSN-II in Monterey Bay, August 2003: surface T on 16 Aug 2003. Left: before adaptation. Right: after adaptation.

Fig. 4, the new values for wind mixing clearly modified surface properties and improved the temperature fields and ocean currents (not shown).

4.3   Generalized Adaptive Biogeochemical Ocean Model

Oceanic states evolve and go through transitions. Efficient predictive models must have the same behavior. This is especially important for marine ecosystems. To automate the dynamic switch of biogeochemical parameters and structures, a new generalized biogeochemical ocean model is being developed [9].

This generalized model is fully modular and flexible (Fig. 5). It has been scientifically constructed based on a serious study of all possible functional groups and parameterizations for coastal ecosystems such as Massachusetts Bay and Monterey Bay (Sects. 4.1,4.2). A general set of state variables and of mathematical structures representing their interactions was selected, based on importance, completeness, efficiency and accuracy. This led to a generalized model (Fig. 5) with the following functional groups of state variables: nutrients (N_i), phytoplankton (P_i), zooplankton (Z_i), detritus (D_i), dissolved organic matter (DOM_i), bacteria (B_i) and auxiliary variables (A_i). Within each functional group, the number of state variables varies from 1 to n. The parameterizations of the interactions among state variables is also variable. Changes in their number at each trophic level can result in automatic changes in these parameterizations.
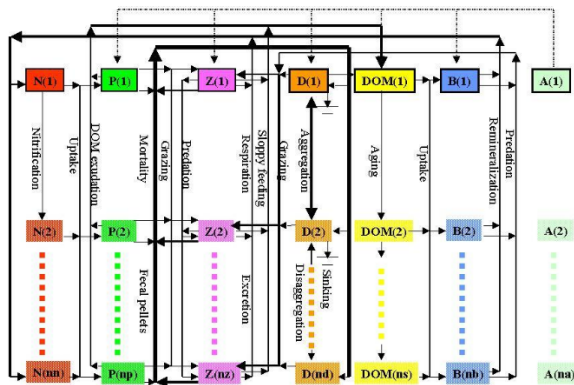

Fig. 5. Generalized biogeochemical model. Thicker arrows imply more processes.

With this flexibility, the generalized biogeochemical model can adapt to different ecosystems, scientific objectives and available measurements. Details of its properties and implementation are given in [9].

5   Conclusion and Future Research

Adaptive modeling based on DDDAS will become essential for interdisciplinary ocean predictions. The present manuscript provides a conceptual basis and illustrative examples for this adaptation. Opportunities and challenges in adaptive ocean modeling abound over a wide spectrum of needs, from coastal ocean science to global climate predictions. Adaptive ocean modeling must be anticipated to accelerate progress in fundamental research and to enable rapid operational predictions otherwise not possible.

References

1. Darema, F., et al.: NSF sponsored workshop on dynamic data driven application systems. Technical report, National Science Foundation (2000) http://www.cise.nsf.gov/cns/darema/dd_das/dd_das_work_shop_rprt.pdf.
2. Robinson, A., Lermusiaux, P.: Data Assimilation in Models. In: Encyclopedia of Ocean Sciences. Academic Press Ltd., London (2001) 623–634
3. Robinson, A., et al.: Harvard Ocean Prediction System (HOPS) (2001-2004) http://oceans.deas.harvard.edu/HOPS/HOPS.html.
4. Stammer, D., Rizzoli, P., Carton, J., Cummings, J., Lermusiaux, P., Moore, A.: A U.S. plan for sustained data assimilation in ocean research (2004) http://www.atmos.umd.edu/~carton/dameeting.
5. Lermusiaux, P., Robinson, A., Haley, P., Leslie, W.: Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation. In: The OCEANS 2002 MTS/IEEE, Holland Publications (2002) 795–802
6. Evangelinos, C., Chang, R., Lermusiaux, P., Patrikalakis, N.: Rapid real-time interdisciplinary ocean forecasting using adaptive sampling and adaptive modeling and legacy codes: Component encapsulation using XML. In Sloot, P., et al., eds.: ICCS 2003. Volume 2660 of LNCS., Springer (2003) 375–384
7. Besiktepe, S., Lermusiaux, P., Robinson, A.: Coupled physical and biochemical data driven simulations of Massachusetts Bay in late summer: real-time and post-cruise data assimilation. J. of Marine Systems **40** (2003) 171–212. M. Gregoire, P. Brasseur and P.F.J. Lermusiaux (Eds.).
8. Larson, J.W., Jacob, R.L., Foster, I., Guo, J.: The Model Coupling Toolkit. Technical Report ANL/CGC-007-0401, Argonne National Laboratory (2001)
9. Tian, R., Lermusiaux, P., McCarthy, J., Robinson, A.: The Harvard generalized biogeochemical ocean model. Technical report, Harvard University DEAS (2004)