Chapter #

# TOWARDS A DYNAMIC DATA DRIVEN SYSTEM FOR RAPID ADAPTIVE INTERDISCIPLINARY OCEAN FORECASTING

N.M. PATRIKALAKIS[a], J.J. MCCARTHY[b], A.R. ROBINSON[b], H. SCHMIDT[a], C. EVANGELINOS[a], P.J. HALEY[b], S. LALIS[a], P.F.J. LERMUSIAUX[b], R. TIAN[b], W.G. LESLIE[b] AND W. CHO[a]
[a]*Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*
[b]*Harvard University, Cambridge, MA 02138, U.S.A.*

**Abstract.** The state of the ocean evolves and its dynamics involves transitions occurring on multiple scales. For efficient and rapid interdisciplinary forecasting, ocean observing and prediction systems must have the same behavior and adapt to the ever-changing dynamics. The present research aims to set the basis of a distributed system for real-time interdisciplinary ocean field and uncertainty forecasting with adaptive modeling and adaptive sampling. The scientific goal is to couple physical and biological oceanography with ocean acoustics. The technical goal is to build a dynamic system based on advanced infrastructures, distributed/Grid computing and efficient information retrieval and visualization interfaces. Importantly, the system combines a suite of modern legacy physical models, acoustic models and data assimilation schemes with new adaptive modeling and adaptive sampling software. The legacy systems are encapsulated at the binary level using software component methodologies. Measurement models are utilized to link the observed data to the dynamical model variables and structures. With adaptive sampling, the data acquisition is dynamic and aims to minimize the predicted uncertainties, maximize the sampling of key dynamics and maintain overall coverage. With adaptive modeling, model improvements are dynamic and aim to select the best model structures and parameters among different physical or biogeochemical parameterizations. This presentation outlines and illustrates the concept, architecture and components of such a Dynamic Data Driven Application System (DDDAS). Current technical and scientific progress is highlighted based on examples in Massachusetts Bay, and Monterey Bay and the California Current System.

**Keywords.** Oceanography, interdisciplinary, adaptive, sampling, modeling, dynamic, data-driven, DDDAS, data assimilation, uncertainty, error estimates, distributed/grid computing.

## 1. INTRODUCTION

Effective ocean forecasting is essential for efficient human operations in the ocean. Application areas include among others fisheries management, pollution control and maritime and naval operations. Scientifically ocean science is important for climate



*F. Darema (ed.), Dynamic Data Driven Applications Systems, pp-pp*

dynamics, biogeochemical budgets and to understand the dynamics and ecosystems of the food web in the sea. Advances in oceanographic numerical models and data assimilation (DA) schemes of the last decade have given rise to complete Ocean Prediction systems [37] that are used in operational settings. Recent developments in the availability of high-performance computing and networking infrastructure now make it possible to construct distributed computing systems that address computationally intensive problems in interdisciplinary oceanographic research, coupling physical and biological oceanography with ocean acoustics [33].

Poseidon [32] is such a distributed computing based project, that brings together advanced modeling, observation tools, and field and parameter estimation methods for oceanographic research. The project has three main goals: 1) to enable efficient interdisciplinary ocean forecasting, by coupling physical and biological oceanography with ocean acoustics in an operational distributed computing framework; 2) to introduce adaptive modeling and adaptive sampling of the ocean in the forecasting system, thereby creating a dynamic data-driven forecast; and, 3) to initiate the concept of seamless access, analysis, and visualization of experimental and simulated forecast data, through a science-friendly Web interface that hides the complexity of the underlying distributed heterogeneous software and hardware resources. The aim is to allow the ocean scientist/forecaster to concentrate on the task at hand as opposed to the micro-management of the underlying forecasting mechanisms.

The Poseidon project employs the Harvard Ocean Prediction System (HOPS) [36] as its underlying advanced interdisciplinary forecast system. HOPS is a portable and generic system for interdisciplinary nowcasting and forecasting through simulations of the ocean. It provides a framework for obtaining, processing, and assimilating data in a dynamic forecast model capable of generating forecasts with 3D fields and error estimates. HOPS has been successfully applied to several diverse coastal and shelf regions [37], and analyses have indicated that accurate real-time operational forecast capabilities were achieved. Error Subspace Statistical Estimation (ESSE) [28], the advanced DA scheme of HOPS that provides an estimate of the dominant uncertainty modes in the forecast, is central to the project's stated goal of adaptive modeling and adaptive sampling. The architecture of Poseidon is being designed based on HOPS, while also keeping in mind possible future HOPS developments so that elements of HOPS could easily be replaced by other components, e.g. employing different physical oceanographic models for adaptive physical modeling. Moreover, the ESSE methodology, that is computing and data intensive, is also an important driving force behind the architectural design decisions.

In the remainder of this paper, Section 2 provides an overview of the dynamic data driven architecture of the Poseidon system, concentrating on the HOPS/ESSE-based forecast workflows, and the concepts of adaptive sampling and adaptive modeling. Section 3 discusses the design of the new computational components of the system and the initial accomplishments in distributed/grid computing and implementing user interfaces. Section 4 illustrates interdisciplinary ocean modeling and forecasting applications, including generalized biological modeling and

objective, non-automated adaptive sampling and adaptive modeling. Section 5 concludes the paper.

## 2. OVERVIEW OF DYNAMIC DATA DRIVEN SYSTEM ARCHITECTURE

The Poseidon system architecture aims to bring together field and remote observations, dynamical, measurement and error models, data assimilation schemes and sampling strategies to produce the best-available estimates of ocean state, parameters and uncertainty. Poseidon's Information Technology approach focuses mainly on key modules or *components* that lead to large gains in efficiency. In general, complex software is thus not re-written, but only modified or updated so as to allow efficient and adaptive distribution. By allowing for *interdisciplinary interactions* (§4), linking in computations physics with biology and acoustics as they are linked in nature, Poseidon aims to capture a more accurate picture of the ocean. At the same time, the system adapts to measurements not only through direct data assimilation but also through data assimilation feedbacks, by the modification of model structure and parameters (adaptive modeling, §4.2-§4.3), and of observational strategies when the most useful data are collected based on ocean field and error forecasts (adaptive sampling, §4.1). This makes Poseidon a dynamic data-driven system [10].

ESSE is a data assimilation scheme that allows for multivariate, inhomogeneous and non-isotropic analyses, with consistent assimilation and adaptive sampling schemes. It is ensemble-based (with a nonlinear and stochastic model) and produces uncertainty forecasts (with a dynamic error subspace and adaptive error learning). It is not tied to any ocean model but its specifics are currently tailored to HOPS. A schematic description of ESSE is shown in Figure 1.

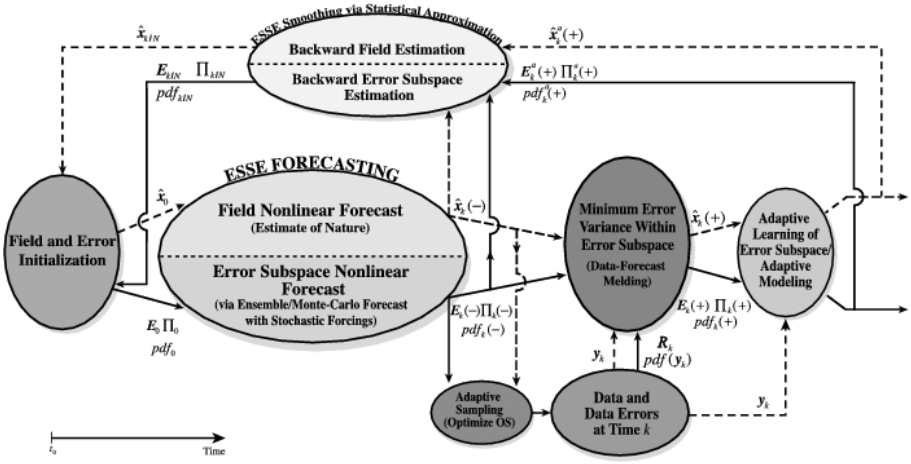


*Figure 1: The ESSE schematic workflow, adapted from [29].*

The Poseidon system builds on ESSE (§3.1) and applies it for the data assimilation of an ocean modeling system including interdisciplinary interactions

between physical oceanography, biological oceanography and ocean acoustics. The target is rapid prediction of interdisciplinary ocean fields and *uncertainties.* Our Dynamic Data Driven Application System [10] (DDDAS) employs autonomous adaptive models for physics and biology, allowing adaptation for parameter values, model structures and state variables and requiring error metrics and criteria to trigger and direct adaptation. It also supports adaptive sampling by concentrating future measurements in regions of high forecast uncertainty or energetic dynamical features (to be identified through *feature extraction* algorithms [18,19] (§3.2.2), which would serve as key components in the fully automated adaptive sampling loop as part of a dynamic data-driven observation system).

The computational framework supporting the Poseidon system is primarily based on *Grid computing technology* [15] (§3.1) and is flexible, allowing for the scalability and incorporation of different models in the future. It also provides the transparent interoperability of the distributed resources. *Metadata* (§3.2.1) are used to describe both datasets (observational and forecast) and software components (code) to allow for advanced automated, distributed and transparent data management as well as the validated composition of several system components into complex information processing workflows that can be executed in a scheduled or on-demand fashion [21]. Finally, it provides a lightweight and user-friendly *Web interfaces* (§3.2.1) for *remote access, control and visualization* (§3.2.2).

## 3. COMPONENTS OF THE POSEIDON SYSTEM: ARCHITECTURE DESIGN AND PROGRESS TO DATE

Here we show initial accomplishments in distributed/grid computing, user-interfaces and overall design of an automated DDDAS. We based our design parameters on the computational and user requirements of Poseidon with respect to the underlying computational framework, its interfaces and the acoustical and biological adaptive modeling and adaptive sampling components. In the subsequent subsections we present some of the research issues, our resulting design and early implementations of the architecture of our evolving system.

### 3.1 Distributed/Grid Computational Strategies

Rapid interdisciplinary ocean forecasting relies heavily on measurements (*in situ* and remote) and models, with associated storage and computation requirements. Data and models are brought together through the process of data assimilation and, in the case of ESSE, the computational work is based on a massive ensemble of forecasts (at least several hundreds). This imposes significant demands on computational power and storage while at the same time being an ideal example for high throughput distributed computing. ESSE ensembles, however, differ from typical parameter scans (one of the most common high throughput applications) in more than one way: (a) there is a hard deadline associated with the execution of the ensemble, as a forecast needs to be timely; (b) the size of the ensemble is dynamically adjusted according to the convergence of the ESSE procedure; (c) individual ensemble members are not significant (and their results can be discarded if unsatisfactory or ignored if unavailable) – what is important is the statistical coverage of the ensemble; (d) the full resulting dataset of the ensemble member

forecast is required, not just a small set of numbers; (e) individual forecasts within an ensemble, especially in the case of interdisciplinary interactions and nested meshes, can be parallel programs themselves.

The significant computational and data requirements of ESSE have driven the adoption of an underlying Grid computing based framework for the Poseidon system allowing for future scalability beyond the confines of a single laboratory and at the same time capitalizing on the significant corpus of work in existence and development in the area of Grid computing technologies (specifically the Globus [46] Toolkit, etc.).

Our low-level computational strategy was shaped by the often-conflicting targets of (i) maximizing computational performance, (ii) maintaining programming investment and (iii) accommodating the needs of code developers. To avoid the resulting major discontinuity in code development [45] we chose to use the constituent domain science codes themselves rather than transforming them into the subroutine form suitable for classical component [6,7] or Java agent based distributed computing [21,23] (which would also require more effort to integrate with a Globus-based Grid computing environment). Component interaction thus generally takes place via file I/O within automated workflows. To address performance issues on the other hand, we plan to parallelize (using MPI and coupling frameworks [24]) tightly coupled interdisciplinary applications (e.g. biology-physics) rather than allow for the far less efficient exchange of data files. Finally, adaptivity that cannot be efficiently expressed at the workflow level is to be implemented within the code in an elegant and efficient manner using function pointers and mixed-language programming.

A high level view of the Grid computing based system architecture of Poseidon is provided in Figure 2. It illustrates the Grid (upper arrow), the computational components currently being developed (boxes on the left and upper portions) and the existing HOPS/ESSE system (schematized on the lower right hand corner), which allows non-automated objective adaptive sampling and adaptive modeling (see feedback arrows from the state/parameter estimates to the data and models respectively).

The main goal of the new computational components (rectangular boxes in Figure 2) is to improve the efficiency of the existing system, especially the automation of multiple tasks carried-out in modern interdisciplinary ocean observing and prediction experiments. The design of these computational components is evolving in a manner compatible with the emerging Earth Science Grid initiative [11] allowing us to capitalize on the new developments in that area. Grid computing will provide the transparent interoperability of the distributed resources: the Globus Toolkit is employed for multi-user distributed authentication/authorization, data and compute access, etc. Remote users will connect to the Poseidon system through a Grid Portal, as well as directly from more powerful clients. Observations will be transferred to Grid/OpenDAP [8] enabled Data Storage Resource Managers while their associated location (and that of any valid cached copies) is recorded in a Replica Location Service (RLS [5]). The same data grid assets will be used to store the results of forecasts. The metadata for any observational and simulation datasets will be stored in an Oceanographic Metadata Catalog Service (MCS [43]), allowing

for searching for datasets based on their content rather than their filename. The combination of MCS and RLS will allow the location of the most appropriate physical copy of the dataset to be used by the system for (a) the computations performed at the Grid-enabled computational resources (clusters, users' workstations and Teragrid resources) and (b) any visualization and data analysis tasks the user requires. A software metadata repository will store the description files (see §3.2.1) used for the remote configuration of the computational tasks.

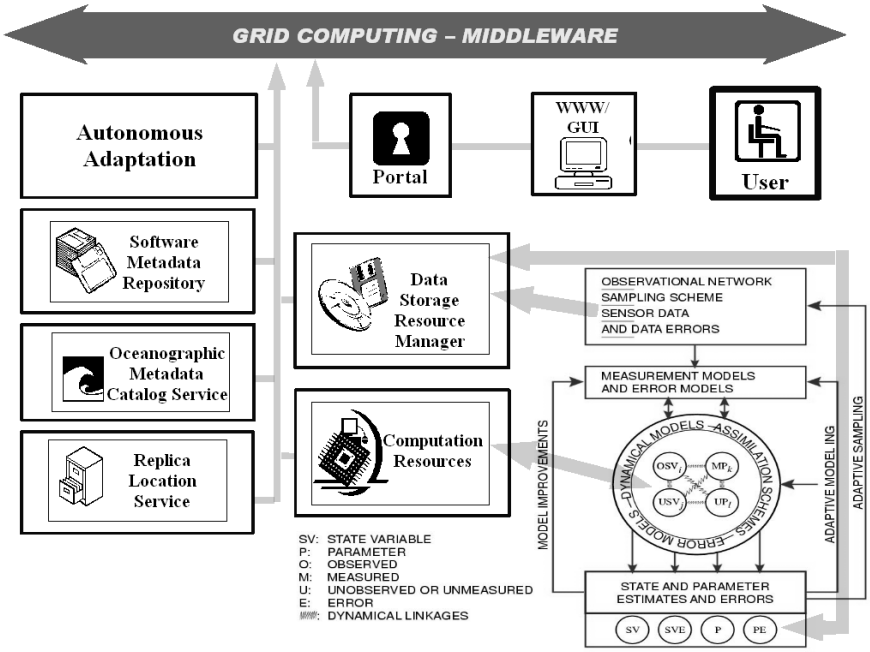


*Figure 2: An overview of the functional components of the Poseidon system and how it improves the efficiency and capability of the HOPS/ESSE system (schematized in the lower right corner).*

## 3.2 User Interfaces

For the eventual adoption into production use of a complex interdisciplinary system such as Poseidon, it is very important for it to be user-friendly, hiding as much of the underlying computational and management complexity from the ocean scientist. While a significant part of the intricacy of the use of Grid computing middleware can be hidden behind Web-enabled Computational Portals and Problem Solving Environments, the complicated (build- and run-time) configuration of the actual interdisciplinary computational components remains a challenge. Ubiquitous access (from remote sites, e.g. on ships) via a lightweight graphical user interface (GUI) was determined as a very important requirement. At the same time visualization, which is an integral part of the ocean forecasting process, both for purposes of

adaptive sampling and for the eventual interpretation of the forecasts, needs to be dealt with within the same framework. The visualization system design needs to balance the needs for interactive exploration of datasets with the restrictions of widespread access: low-bandwidth connections and heterogeneous low-end mobile clients.

### 3.2.1 Generic Web User Interfaces

In the process of designing Poseidon we had to address the fact that HOPS (as other ocean applications, e.g. for physical oceanography ROMS [20] or for ocean acoustics OASES [42]) are, like most scientific applications, legacy[1] programs. Their native binaries expect a standard input (*stdin*) stream, maybe some command line options and a set of input files and generate a set of output files as well as standard output (*stdout*) and error (*stderr*) streams. In such a setup, workflows are either executed interactively in a step-by-step fashion (a very common approach) or (after potential problems are handled) as hard-coded shell-scripts that can be executed in the background. While such an approach, which dates from the days when GUIs were not available, is efficient for a skilled user, it still is cumbersome, error-prone and entails a steep learning curve. Runtime configuration files are complex in general, follow their own formatting rules, may obey complicated dependency and conflict rules and are rarely self-documenting. Add to this the extra complication of configuring the rebuilding of the code (e.g. via a set of preprocessor defines setup in a Makefile, each with their own dependencies on each other and on runtime configuration options) and one arrives at a scenario that is not suited for remote use over the Web.

After examining various ways of dealing with this issue, without costly changes for developers [45] and keeping in mind that the Poseidon system should allow for future handling of non-HOPS components without excessive recoding, we decided to avoid changing the codes or generating specialized GUIs; instead we opted to describe their functionality and requirements - essentially “software metadata” - using the eXtensible Markup Language (XML) [12]. Thus we create what is for practical purposes a computer-readable manual for the codes with information to check for option/parameter correctness (type, range and dependencies) and produce properly formatted input files, scripts, Makefiles and command lines. We have been developing a hierarchy of XML Schemata [41,48] for our software metadata descriptions, attempting to cover as general a case of a legacy application as possible beyond the HOPS and acoustics binaries in Poseidon [4,13,14]. A prototype Java-based tool, called LEGEND (LEGacy Encapsulation for Network Distribution) [16] has been developed: using a repository of software metadata and associated schemata it automatically generates a validating GUI to produce scripts for building and running the binaries and allows for controlling their Grid or local execution.

---

1The term “legacy” should not be misconstrued to imply outdated code in this context: these are all codes with an active development community and recent enhancements. For various reasons they are still b/eing developed mainly in Fortran and in any case are command-line and not GUI-driven.

### 3.2.2 Remote Visualization and Feature Extraction

The graphical output from the individual components of the Poseidon system is based on a variety of software: e.g. NCAR Graphics [30] and MATLAB are used in HOPS and MINDIS or PLOTMTV for OASES. While it is possible to use these tools remotely over the network (via X-Windows or VNC [47]), such a solution is not efficient (with secure access exacerbating the situation), can become unusable over slow connections and imposes extra software and hardware restrictions on the client machines. A major requirement has been the handling of the NetCDF self-describing portable file format that HOPS and most ocean modeling codes use. Based on scientists' usage patterns, we concluded on three different visualization approaches:

1. To cover the standard set of 2D horizontal and vertical slice-based visualizations ocean scientists always look at, we will employ LEGEND-configurable shell scripts that automatically generate Web pages with the required results embedded as images. Such scripts use the existing tools (NCAR Graphics etc.) thus leveraging the robust corpus of model-specific visualization work.
2. For more interactive and capable visualization work, we employ OpenDX [31] and Java Explorer [22] that (using applets for remote control) allow the rendered visualization output to be updated on the user's Web browser. OpenDX offers us the capability to graphically compose complicated interactive visualization (possibly distributed) workflows – these are then exported via Java Explorer for Web usage. While the interactive response of this approach is worse compared to using OpenDX locally, such an approach fits remote lightweight clients.
3. At the same time, the Poseidon system will still allow users to still transfer datasets via the Grid to their local workstations and use their traditional (or future) local tools in a manner very similar to their existing mode of operation.

Beyond user-friendly remote visualization, the relevance and usefulness of the visual picture has also been an important target. Without feature extraction, the human operator needs to visually identify important dynamical events, this being vital for human-directed adaptive sampling. We have been developing a suite of tools to automatically identify oceanic flow features such as eddies/gyres, upwelling etc. and graphically present these results to enhance the effectiveness of the human forecaster and operations planner. For the more involved problem of vortex (eddy etc.) identification we have developed an efficient two-stage algorithm that first identifies vortex cores and then locates the boundaries of the closed streamline region around them [18,19]. The same tools, appropriately modified would serve as key components in the fully automated adaptive sampling loop as part of a dynamic data-driven observation system.

## 4. REAL-TIME INTERDISCIPLINARY MODELING AND FORECASTING AND PROGRESS TO DATE TOWARDS AN OCEAN SCIENCE DDDAS.

Here we illustrate interdisciplinary ocean modeling and forecasting applications during which objective adaptive sampling [39] and non-automated physical adaptive modeling [27] were carried-out. The adaptive sampling was carried-out in Monterey Bay in real-time and enabled by an improved Poseidon-based distribution of the

ensemble of parallel ESSE computations (§3.1). The adaptive modeling was carried out manually in real-time during the same experiment. In order to allow automated adaptive biogeochemical modeling, a new generalized biogeochemical modeling system is being implemented. This new system will be used in future real-time interdisciplinary simulations and progress to date is exemplified for Monterey Bay. Finally, the preliminary development of acoustical-biological measurement models to estimate biological properties from acoustical sensing is outlined.

### 4.1 Objective Adaptive Sampling Using ESSE

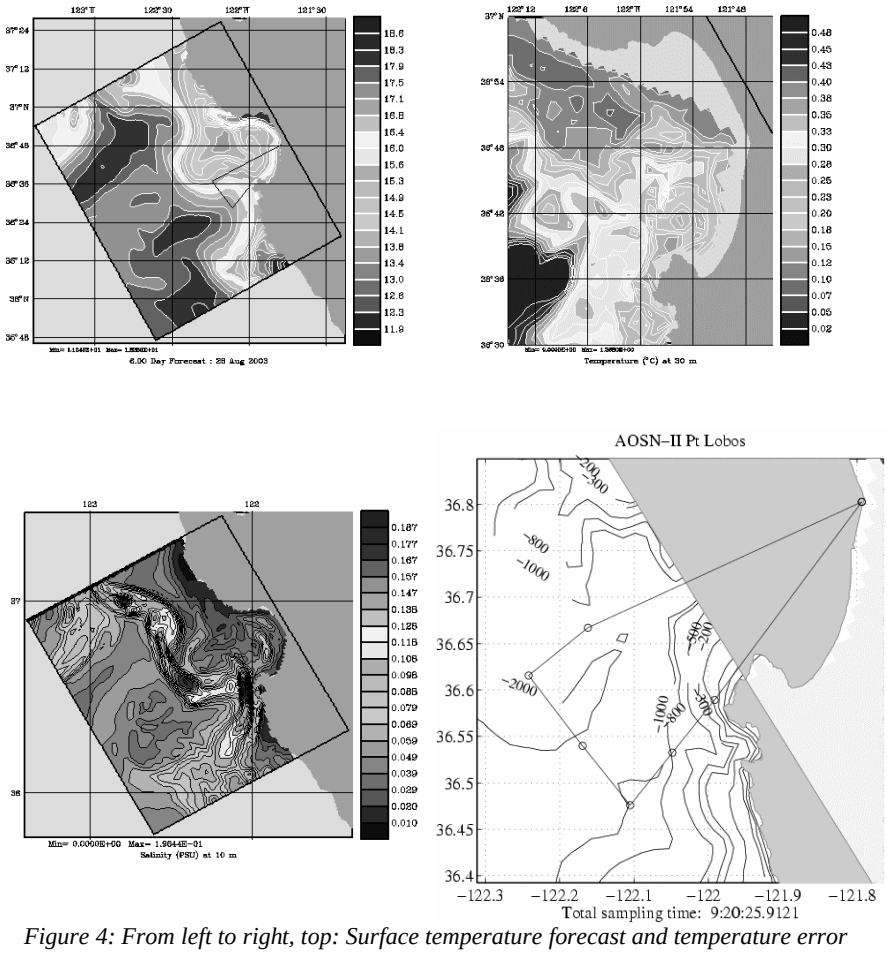


*Figure 4: From left to right, top: Surface temperature forecast and temperature error forecast; from left to right, bottom: salinity error forecast and adaptive sampling pattern for Pt.Lobos during AOSN-II, August 26, 2003.*

With adaptive sampling, the most useful data are collected based on ocean field and error forecasts, either subjectively or objectively through the use of quantitative

criteria or goals. A goal characterizes the ideal future sampling among the possible choices, in adaptive accord with the constraints, available forecasts and past data, e.g. [3,17,25,34,38]. Typically, the areas to be sampled will be chosen based on: a) forecast uncertainty (e.g. error variance, higher moments, probability density functions); b) interesting interdisciplinary phenomena and dynamics (e.g. feature extraction, Multi-Scale Energy and Vorticity Analysis); and, c) maintenance of synoptic forecast accuracy.

In current adaptive sampling [40,39,26], field and error forecasts are combined with *a priori* experience to intuitively choose the future sampling. An example of this comes from the Autonomous Ocean Sampling Network (AOSN-II) [9] field experiment in Monterey Bay, CA during the summer of 2003 [1]. The model forecast for 26 August 2003 predicted a meander of the coastal current that advected warm, fresh water (Figure 4 top left) towards the Monterey Bay Peninsula. The temperature and salinity error fields (Figure 4 top right and bottom left) from a 450-member ensemble (computed using the first version of the distributed ESSE scheme, see §3.1) indicated a high degree of uncertainty in both the position and strength of the meander. In fact, specific ensemble members had either essentially no meander or shifted the meander to the north. Based on this information, and constrained by operational limitations, a sampling pattern (Figure 4 bottom right) was devised for the research vessel Pt. Lobos.

Several different methodologies for obtaining the areas of interest for targeted observations (breeding vectors, singular vectors, ESSE, feature extraction) are being examined in combination with the problem of most intelligently combining areas corresponding to different attribute sets (feature of type n, uncertainty of magnitude E). Optimal methods to schedule such observations given a set of observational assets and corresponding constraints are under investigation.

In the specific case of acoustical adaptive sampling, physical features must be accounted for to compensate for their backscatter, or to sample more effectively the water column - the pycnocline and thermocline typically concentrate plankton layers and can lead to specular or coherent pressure wave reflection. Mixing (of nutrients as well as generation of small-scale sound velocity gradients), presence of sand in upwelling plumes or bubbles in a surface layer, solitons and multireflections between a quiescent sea and a flat sediment bottom are features likely to generate undesired sonar echoes.

### 4.2 Generalized Biological Modeling and Non-Automated Physical Adaptation

Data-driven adaptive modeling and real-time forecast of marine ecosystems [27] is an increasing challenge in marine sciences. In the context of global climate warming and increasing anthropogenic stress, marine ecosystems are becoming more and more vulnerable and uncertain. Eutrophication, harmful algal blooms, red tide, oil spills, toxic element pollution can all deteriorate the health and functioning of marine ecosystems.

Traditionally marine ecosystems are modeled with simulation models of fixed structure and static data inputs. However, forecasting evolving marine ecosystems, in space and time, in response to environmental perturbation, necessitate rapid

response of dynamic data-driven adaptive simulation models. Presently, a model is considered to be adaptive if its formulation, classically assumed constant, is made variable as a function of data flows.

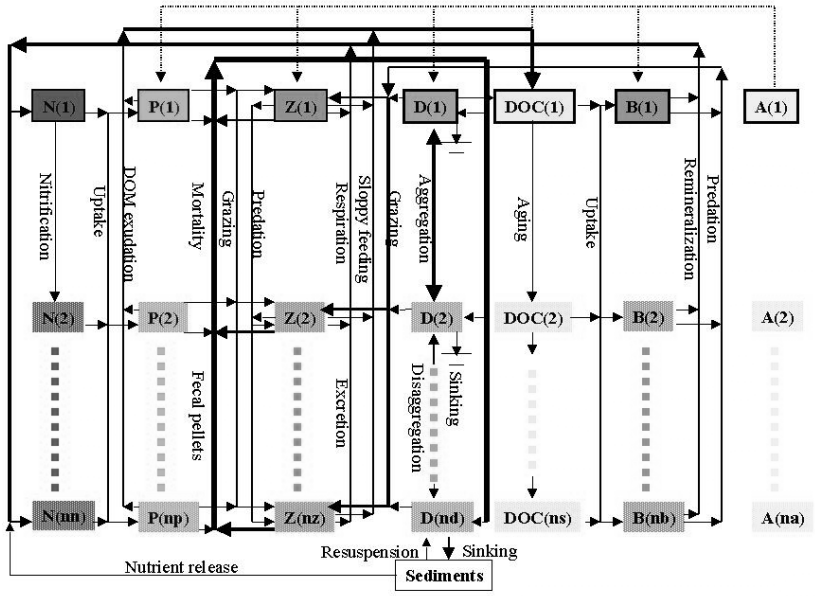


*Figure 5: Generalized biological model. N: Nutrients; P: Phytoplankton; Z: Zooplankton; D: Biogenic detritus; DOM: Dissolved organic matter; B: Bacteria; A: Auxiliary state variables; nn, np, nz, nd, ns and na are the total numbers of state variables of the functional groups.*

We have developed a preliminary version of a generalized, flexile biological model specifically designed for adaptive modeling and real-time ecosystem forecast (Figure 5). Marine ecosystems function through a series of highly integrated interactions between biota and the habitat and dynamic links among food web components. Based on the trophic and biogeochemical dynamics, the generalized model is composed of 7 functional groups: nutrients ($N_i$), phytoplankton ($P_i$), zooplankton ($Z_i$), detritus ($D_i$), dissolved organic matter ($DOM_i$), bacteria ($B_i$) and auxiliary state variables ($A_i$).

Traditionally, the number of compartments in a biological model is fixed with each compartment representing a specific biological community or species. In our generalized biological model however, the number of components of each functional group is a variable (varying from 1 to $n$) and users define the biological correspondents while applying the model to a specific ecosystem. In the code, each trophic level and trophic link is computed by using loops from 1 to $n$. The changes in the number $n$ at various trophic levels result in automatic changes in the model structure. By using a subset of the state variables of the generalized biological model we can simulate various ecosystems. For example, if the component number $n$ is assigned to 1 for nutrient, phytoplankton and zooplankton and to 0 for all other

functional groups, the generalized biological model will represent the NPZ model. When the component number of detritus is assigned to 1 in the previous configuration, the generalized model will be a NPZD model. If the component number is assigned to be 2 for all the trophic levels above, the generalized biological model will be a doubled NPZD model. The potential combinations and actual structures of the generalized biological model can be very large. The model can be adaptive, i.e. the state variables, model structures and parameter values can change in response to field measurements, ecosystem function and scientific objectives.

This model has been coupled with HOPS. An application of this forecasting system to the Monterey Bay area is underway to study biological response to upwelling events at ecosystem level.

### 4.2.1 Monterey Bay Application

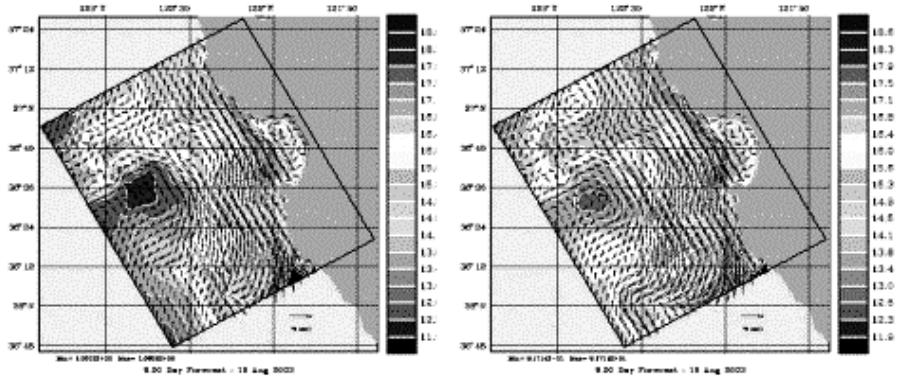

*Figure 6: Simulated surface water temperature (℃) before (left) and after (right) non-automated adaptation on August 11, 2003 during the AOSN-II experiment in Monterey Bay.*

The Monterey Bay ecosystem is characterized by episodic upwelling events, patchiness and filaments in biological fields resulting from upwelling jets, plumes, fronts and interactions with the California Coastal Currents. The large size mesoplankton food web generally dominates in upwelling centers and plumes whereas the microbial food web prevails in the adjacent oceanic waters. Succession in food web structure between upwelling and relaxation periods has also been observed. To adapt the generalized biological model to this specific ecosystem, 10 state variables were considered in the simulation, including the microbial food web ($NH_4^+$, picophytoplankton, microzooplankton, bacteria, dissolved organic carbon (DOC) and particulate organic carbon (POC) and the mesoplankton food web ($NO_3^-$, diatoms, mesozooplankton and large sinking detritus). In addition to these 10 functional state variables, 4 auxiliary variables were simulated as well, prokaryote, eukaryote and total chlorophyll and bioluminescence.

This data-driven physical and biological prediction system was applied during the AOSN-II field experiment. Remote and *in situ* sensors and platforms including multiple satellite images, drifters, gliders, moorings, AUV and ship-based data [1] were deployed to collect data in real time. These data were assimilated into

numerical models and daily predictions of the ocean fields and uncertainties were issued. 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, it was necessary to adapt several parameters of the physical ocean model 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 in Figure 6, the new values for wind mixing clearly modified surface properties and improved the temperature fields and corresponding currents.

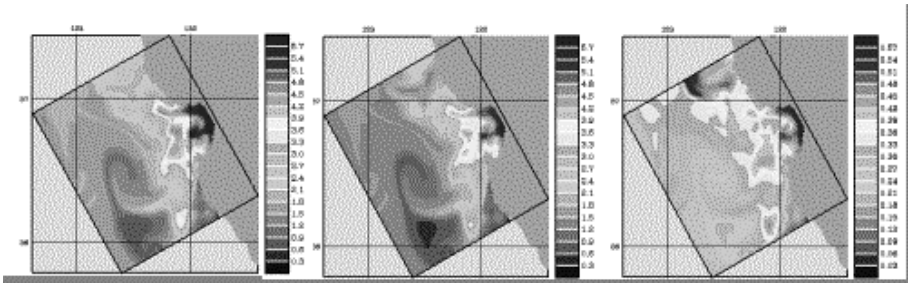

*Figure 7: Simulated field of total (left), eukaryote (middle) and prokaryote (right) chlorophyll (μg/l) on August 11, 2003 in Monterey Bay during an upwelling event.*

The generalized biological model and parameter values have been configured to adapt to the Monterey Bay system. Historical data have been mapped onto the simulation grids by using objective analysis; they were then used to initialize the biological simulation. The simulation was started on August 6, 2003, and stopped on August 11, 2003, i.e. a 5-day simulation. The preliminary results show that physical processes are the key factor in determining biological dynamics and distribution. While primary production is linked to upwelling events, the distribution of biological field is essentially by currents and eddies. An anticyclone was simulated offshore from Monterey Bay. Filaments and fronts in biological distributions can be observed accordingly in Figure 7.

### 4.3 Acoustical-Biological Measurement Models

Acoustical-biological measurement models involve the reversal of the underwater sound wave scattering process, thereby allowing estimation of the biological population size and species distribution (zooplankton population) from the backscatter spectrum [2,35,44]. This scattering reversal process (acoustical inversion) allows estimation of the expected value of the zooplankton size and species distribution as well as estimation of the error of this inversion process.

Estimation of the error of such an inversion process is useful in the data assimilation framework adopted in this project. In addition, such error estimation is needed for the development of adaptive sampling methods (§4.1), which aim to optimally reduce the uncertainty in the field estimates. The present acoustical-biological measurement methods allow for this to be implemented in practice for acoustical-biological-physical estimation via judicious adjustment of the real-time

acoustic sensing capability. Initial progress towards this goal and the technical details of the acoustic-biological models are reported in [35].

## 5. CONCLUSIONS

This paper provides an overview of a DDDAS for rapid adaptive interdisciplinary ocean forecasting. Information technology allows the development of an Internet-based distributed system that enables the seamless integration of field and remote observations, dynamical, measurement and error models, data assimilation schemes and adaptive sampling strategies for the effective estimation of oceanic fields and their uncertainties. Important components of the system already developed or envisioned for the near term have been described in some detail. Illustrative examples of interdisciplinary modeling and forecasting and progress to date towards an integrated ocean science DDDAS were presented. Automated adaptive modeling and adaptive sampling, fully coupled physical-acoustical-biological oceanography and grid computing application aspects are recommended for further study.

## ACKNOWLEDGEMENTS

This work was funded in part from NSF/ITR (under grant EIA-0121263), and from DoC (NOAA via MIT Sea Grant) (under grant NA86RG0074). We thank the AOSN-II team. The HU results were funded in part from NSF/ITR (under grant 5710001319) and ONR (under grant N00014-01-1-0771, N00014-02-1-0989 and N00014-97-1-0239).

## REFERENCES

1. AOSN II/Monterey Bay 2003. http://www.mbari.org/aosn/MontereyBay2003.htm.
2. Berman, M.S., Green, J.R., Holliday, D.V. & Greenlaw, C.F. (2002). Acoustic determination of the fine-scale distribution of zooplankton on Georges Bank. *Mar. Ecol. Prog. Ser.*
3. Buizza, R. & Montani, A. (1999). Targeting observations using singular vectors. *Journal of the Atmospheric Sciences*, 56(17), 2965-2985.
4. Chang, R.C. (2003). *The Encapsulation of Legacy Binaries using an XML-Based Approach with Applications in Ocean Forecasting*, M.Eng. in Electrical Engineering and Computer Science thesis, Massachusetts Institute of Technology.
5. Chervenak, A., Deelman, E., Foster, I., Guy, L., Hoschek, W., Iamnitchi, A., Kesselman, C., Kunszt, P., Ripenu, M., Schwartzkopf, B., Stocking, H., Stockinger, K. & Tierney, B. (2002). *Giggle: A framework for constructing scalable replica location services*. In Proc. of SC2002 Conference, November 2002. IEEE CS/ACM
6. Common Component Architecture Forum: http://www.cca-forum.org/.
7. Common Object Request Broker Architecture: http://www.corba.org/.
8. Cornillon, P., Gallagher, J., & Sgourosy, T. (2003). OPENDAP: Accessing data in a distributed, heterogeneous environment. *Data Science Journal*, 2(5), 164–174, November 2003.
9. Curtin, T.B., Bellingham, J.G., Catipovic, J. & Webb, D. (1993). Autonomous Oceanographic Sampling Networks. *Oceanography*, 6(3), 86-94.
10. Darema, F. et al. (2000). *NSF Sponsored Workshop on Dynamic Data Driven Application Systems.* Technical report, National Science Foundation, http://www.cise.nsf.gov/cns/darema/dd_das/dd_das_work_shop_rprt.pdf.
11. Earth System Grid (ESG). https://www.earthsystemgrid.org/.
12. eXtensible Markup Language: http://www.w3.org/XML/.
13. Evangelinos, C., Chang, R., Lermusiaux, P.F.J. & Patrikalakis, N.M. (2003). *Rapid real-time interdisciplinary ocean forecasting using adaptive sampling and adaptive modeling and legacy codes: Component encapsulation using XML*. In P.M.A. Sloot, D. Abramson, A. Bogdanov, J.J. Dongarra, A. Zomaya & Y. Gorbachev (Eds), Computational Science - ICCS 2003, International

Conference on Computational Science 2003, volume 2660 of Lecture Notes in Computer Science, (pp. 375–384). Springer.
14. Evangelinos, C., Chang, R., Lermusiaux, P.F.J. & Patrikalakis, N.M. (2004). Web-Enabled Configuration and Control of Legacy Codes. *International Journal of Cooperative Information Systems*. In press.
15. Foster, I. & Kesselman, C., (Eds.) (1999). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann
16. Geiger, S.K. (2004). *Legacy Computing Markup Language (LCML) and LEGEND – LEGacy Encapsulation for Network Distribution*, S.M. in Ocean Engineering thesis, Massachusetts Institute of Technology.
17. Gloor, M., Fan, S-M., Pacala, S. & Sarmiento, J. (2000). Optimal sampling of the atmosphere for purpose of inverse modeling: A model study.*Global Biogeochemical Cycles*, 14(1), 407-428.
18. Guo, D. (2004). *Automated feature extraction in oceanographic visualization*. M.Sc. in Ocean Engineering and EECS thesis, Massachusetts Institute of Technology.
19. Guo, D., Evangelinos, C., & Patrikalakis, N.M. (2004). *Flow feature extraction in oceanographic visualization*. In D. Cohen-Or, L. Jain, & N. Magnenat-Thalmann, (Eds), Proc. of Computer Graphics International Conference, CGI 2004, Crete, Greece, Los Alamitos, CA, June 2004. IEEE Computer Society Press.
20. Haidvogel, D., Arango, H., Hedstrom, K., Malanotte-Rizzoli, A.B.P. & Shchepetkin, A. (2000). Model evaluation experiments in the North Atlantic basin: Simulations in nonlinear terrain-following coordinates. *Dyn. Atmos. Oceans*, 32, 239–281.
21. Houstis, C., Lalis, S., Christophides, V., Plexousakis, D., Vavalis, E., Pitikakis, M., Kritikos, K., Smardas, A. & Gikas C. (2002). *A service infrastructure for e-Science: the case of the ARION system*. In Proc. of the 14th Intern. Conf. on Advanced Information Systems Engineering (CAiSE 2002), E-Services and the Semantic Web workshop (WES2002), Toronto, Canada. Number 2512 in Lecture Notes in Computer Science, (pp. 175–187). Springer.
22. Java Explorer: Enables web deployment of data explorer visualization solutions. http://www.sdsc.edu/dx/JavaExplorer.htm.
23. Java Native Interface: http://java.sun.com/j2se/1.4.2/docs/guide/jni/.
24. Larson, J.W., Jacob, R.L., Foster, I. & Guo, J. (2001). The Model Coupling Toolkit. Technical Report ANL/CGC-007-0401, Argonne National Laboratory, April 2001.
25. Lermusiaux, P.F.J. (1997). *Error Subspace Data Assimilation Methods for Ocean Field Estimation: Theory, Validation and Applications*, PhD thesis, Harvard University.
26. Lermusiaux, P.F.J. (2001). Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay. *J. Marine Systems*, 29, 385-422.
27. Lermusiaux, P.F.J., Evangelinos, C., Tian, R., Haley, P.J., McCarthy, J.J., Patrikalakis, N.M., Robinson, A.R. & Schmidt, H. (2004). *Adaptive coupled physical and biogeochemical ocean predictions: A conceptual basis.* In M. Bubak et al. (Eds), Computational Science - ICCS 2004, International Conference on Computational Science 2004, Lecture Notes in Computer Science, Springer. In press.
28. Lermusiaux, P.F.J. & Robinson, A.R. (1999). Data assimilation via Error Subspace Statistical Estimation. Part I: Theory and schemes. *Month. Weather Rev.*, 127, 1385–1407.
29. Lermusiaux, P.F.J., Robinson, A.R., Haley, P.J. & Leslie, W.G. (2002). *Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation.* In Proc. of the OCEANS 2002, (pp. 795–802). MTS/IEEE, Holland Publications.
30. NCAR Command Language and NCAR Graphics. http://ngwww.ucar.edu/.
31. Open visualization Data eXplorer (OpenDX). http://www.opendx.org/.
32. Patrikalakis, N.M. Poseidon: *A Distributed Information System for Ocean Processes* http://czms.mit.edu/poseidon/
33. Patrikalakis, N.M., Abrams, S.L., Bellingham, J.G., Cho, W., Mihanetzis, K.P., Robinson, A.R., Schmidt, H. & Wariyapola, P.C.H. (2000). *The digital ocean.* In Proc. of Computer Graphics International, GCI ’2000, Geneva, Switzerland, (pp. 45–53). Los Alamitos, CA: IEEE Computer Society Press.
34. Palmer, T.N., Gelaro, R., Barkmeijer, J. & Buizza, R. (1998). Singular vectors, metrics, and adaptive observations. *American Meteorol. Soc.*, February 1998.
35. Renard, B. (2003). *Inversion of Acoustic Zooplankton Measurement for Adaptive Physical-Biological Ocean Forecast.* M.Sc. in Ocean Engineering thesis, Massachusetts Institute of Technology.

36. Robinson, A.R. Harvard: *Ocean Prediction System (HOPS)* http://oceans.deas.harvard.edu/HOPS/HOPS.html.
37. Robinson, A.R. (1999). Forecasting and simulating coastal ocean processes and variabilities with the Harvard Ocean Prediction System. In C. Mooers (Ed.), *Coastal Ocean Prediction*, AGU Coastal and Estuarine Studies Series. (pp. 77–100). American Geophysical Union.
38. Robinson, A.R. & Glenn, S.M. (1999). Adaptive Sampling for Ocean Forecasting. *Naval Research Reviews*, 51(2), 28-38.
39. Robinson, A.R. & Sellschopp, J (2000). Rapid assessment of the coastal ocean environment, in N. Pinardi & J.D. Woods (Eds), *Ocean Forecasting: Conceptual Basis and Applications*, (pp. 203-232), Springer-Verlag.
40. Robinson, A.R. & the LOOPS Group (1999). *Realtime Forecasting of the Multidisciplinary Coastal Ocean with the Littoral Ocean Observing and Predicting System (LOOPS).* In Proc. Of the Third Conference on Coastal Atmospheric and Oceanic Prediction and Processes, 3-5 November 1999, New Orleans, LA, (pp. 30-35), American Meteorological Society.
41. Roy, J. & Ramanujan, A. (2001). XML schema language: Taking XML to the next level. *IT Professional*, 3, 37–40.
42. Schmidt, H. & Tango, G. (1986). Efficient global matrix approach to the computation of synthetic seismograms. *Geophys. J. R. Astr. Soc*.,. 84.
43. Singh, G., Bharathi, S., Chervenak, A., Deelman, E., Kesselman, C., Manohar, M., Patil, S. & Pearlman, L. (2003). *A metadata catalog service for data intensive applications*. In Proc. of SC2003 Conference, Phoenix, AZ, November 15-21 2003. IEEE CS/ACM.
44. Stanton, T.S. & Chu, D., (2000). Review and recommendations for the modelling of acoustic scattering by fluid-like elongated zooplankton: euphausiids and copepods. *ICES J. Mar. Sci*, 57, 793-807.
45. Terekhov, A. & Verhoef, C. (2000). The realities of language conversions. *IEEE Software*, 111–124.
46. The Globus Project: http://www.globus.org/.
47. Virtual Network Computing (VNC). http://www.realvnc.com/.
48. XML Schema Specification: http://www.w3.org/XML/Schema/.