Chapter 14
Toward Dynamic Data-Driven Systems for Rapid Adaptive Interdisciplinary Ocean Forecasting

N. M. Patrikalakis, P. F. J. Lermusiaux, C. Evangelinos, J. J. McCarthy, A. R. Robinson, H. Schmidt, P. J. Haley, S. Lalis, R. Tian, W. G. Leslie, and W. Cho

Abstract The state of the ocean evolves and its dynamics involves transitions occurring at 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. This chapter sets 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 acoustic measurements. The technical goal is to build a dynamic modeling and instrumentation system based on advanced infrastructures, distributed/grid computing, and efficient information retrieval and visualization interfaces, from which all these are incorporated into the Poseidon system. Importantly, the Poseidon system combines a suite of modern legacy physical models, acoustic models, and ocean current monitoring data assimilation schemes with innovative modeling and adaptive sampling methods. 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 optimized sampling of key dynamics, and maintain overall coverage. With adaptive modeling, model improvements dynamically select the best model structures and parameters among different physical or biogeochemical parameterizations. The dynamic coupling of models and measurements discussed here, and embodied in the Poseidon system, represents a Dynamic Data-Driven Applications Systems (DDDAS). Technical

N. M. Patrikalakis · P. F. J. Lermusiaux (✉) · C. Evangelinos · H. Schmidt · P. J. Haley · S. Lalis · W. G. Leslie · W. Cho
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: pierrel@mit.edu

J. J. McCarthy · A. R. Robinson · R. Tian
Harvard University, Cambridge, MA, USA

F. Darema et al. (eds.), *Handbook of Dynamic Data Driven Applications Systems*,
https://doi.org/10.1007/978-3-031-27986-7_14

and scientific progress is highlighted based on examples in Massachusetts Bay, 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

14.1 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 dynamics, biogeochemical interactions, 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 [12, 34, 35, 41, 53] have given rise to complete Ocean Prediction systems [49] 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 [29, 45, 62].

Poseidon [44] is one such distributed computing-based system, within the general framework of Dynamic Data-Driven Applications Systems (DDDAS) [67]. It brings together advanced modeling, observation tools, and field and parameter estimation methods for oceanographic research. The Poseidon system is aimed to address three main objectives: (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 allows users high-level interaction without the need to manually interact with the complexity of the underlying distributed heterogeneous software and hardware resources. The Poseidon system will 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 system employs the Harvard Ocean Prediction System (HOPS) [48], 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 [49], and analyses indicate that accurate real-time operational forecast capabilities were achieved. Error Subspace Statistical Estimation (ESSE) [39], the advanced data assimilation DA scheme of HOPS that provides an estimate of the dominant uncertainty modes in the forecast [35, 36], is central to the Poseidon system's stated goal of adaptive modeling and sampling [25, 37, 38]. The architecture of Poseidon is being designed based on HOPS, while also keeping open 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, which is computing and data-intensive, is also an important driving force behind the architectural design decisions.

In the remainder of this chapter, Sect. 14.2 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 dynamic adaptive sampling and adaptive modeling [38]. Section 14.3 illustrates interdisciplinary ocean modeling and forecasting applications, including generalized biological modeling and objective, non-automated adaptive sampling and adaptive modeling. Section 14.4 discusses the design of the new computational components Poseidon and the initial accomplishments in distributed/grid computing and implementing user interfaces. Section 14.5 provides summary comments on the work presented in this chapter.

14.2 Overview of Dynamic Data-Driven System Architecture

The Poseidon system architecture aims to bring together field and remote observations, dynamic 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 (see Fig. 14.1) focuses mainly on key modules or components that lead to large gains in efficiency. In general, complex software is thus not rewritten, but only modified or updated so as to allow efficient and adaptive distribution. By allowing for interdisciplinary interactions (see Sect. 14.3), linking physics computations 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 is discussed in Sects. 14.3.2 and 14.3.3), and of observational strategies when the most useful data are collected based on ocean field and error forecasts (adaptive sampling, Sect. 14.3.1) affording Poseidon as a dynamic data-driven system [11].

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). Poseidon is not tied to any ocean model but its specifics are currently tailored to HOPS. A schematic description of ESSE is shown in Fig. 14.1.

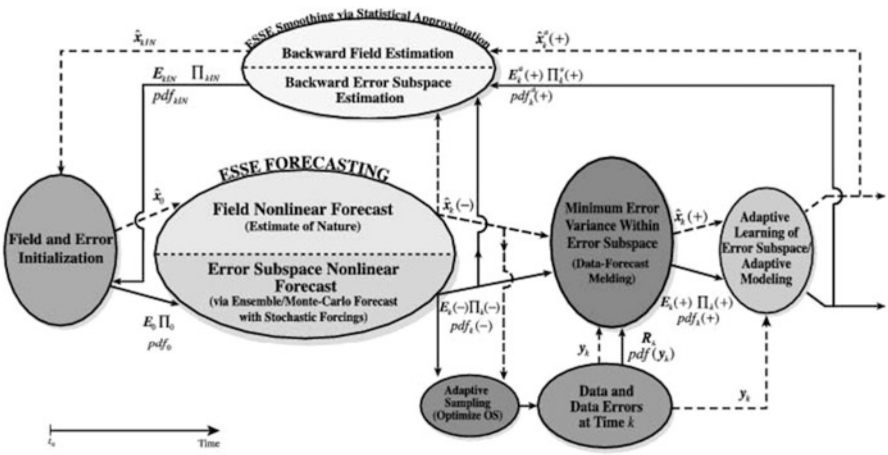

Fig. 14.1 The ESSE schematic workflow. (Adapted from [40])

The Poseidon system builds on ESSE (Sect. 14.4.1) and applies it for the data assimilation of an ocean modeling system including interdisciplinary interactions between physical oceanography, biological oceanography, and ocean acoustics. The application is a rapid prediction of interdisciplinary ocean fields and uncertainties. The Dynamic Data Driven Applications Systems (DDDAR)-based [11] Poseidon system employs autonomous adaptive models for physics and biology, allowing adaptation for parameter values, model structures, and state variables, and requires 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 [21, 22] (Sect. 14.4.2.2), which 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* technologies [18] (Sect. 14.4.1) 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 (Sect. 14.4.2.1) 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 [26]. Finally, Poseidon provides lightweight and user-friendly Web interfaces (Sect. 14.4.2.1) for remote access, control, and visualization (Sect. 14.4.2.2).

14.3 Real-Time Interdisciplinary Modeling and Forecasting and Process to Date Toward an Ocean Science DDDAS

This section illustrates interdisciplinary ocean modeling and forecasting applications, during which objective adaptive sampling [51] and non-automated physical adaptive modeling [33] result. The adaptive sampling was tested in Monterey Bay in real time [24] and enabled by an improved Poseidon-based distribution of the ensemble of parallel ESSE computations (Sect. 14.4.1). The adaptive modeling was manually evaluated in real time during the same experiment. In order to allow automated (*dynamic data-driven*) 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 the Monterey Bay test case (shown in Fig. 14.2). Finally, the preliminary development of acoustical-biological measurement models to estimate biological properties from acoustical sensing is outlined in this chapter.

14.3.1 Objective Adaptive Sampling Using ESSE

With adaptive sampling, the most useful data collected are based on the 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 accord with the constraints, available forecasts, and past data (e.g., [3, 20, 31, 46, 50]). 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 adaptive sampling [32, 51, 52], field and error forecasts can be combined with a priori experience to intuitively choose future sampling. An example of this comes from the Autonomous Ocean Sampling Network (AOSN-II) [10] 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, freshwater (Fig. 14.2 top left) toward the Monterey Bay Peninsula. The temperature and salinity error fields (Fig. 14.2 top right and bottom left) from a 450-member ensemble (computed using the first version of the distributed ESSE scheme, see Sect. 14.4.1) 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 the collected information, and constrained by operational limitations, a sampling pattern (Fig. 14.2 bottom right) was devised for the research vessel Pt. Lobos.

Several different methodologies for obtaining the areas of interest for targeted observations (e.g., breeding vectors, singular vectors, ESSE, feature extraction)

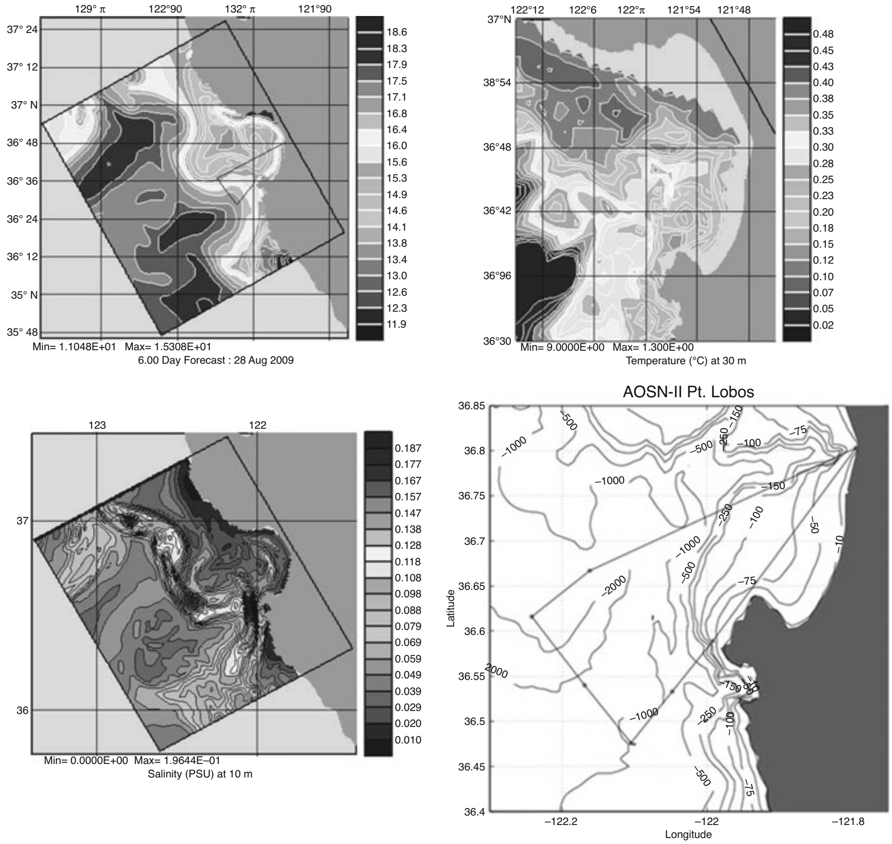

Fig. 14.2 From left to right, top: Surface temperature forecast and temperature error forecast; from left to right, bottom: salinity error forecast and adaptive sampling pattern for Pt. Lobos during AOSN-II, August 26, 2003

were 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 available assets and corresponding constraints sought to enhance relevant data.

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: for example, the pycnocline and thermocline (i.e., the water layer of steep density and temperature gradients) 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 multi-reflections between a quiescent sea and a flat sediment bottom are features likely to generate undesired sonar echoes.

14.3.2 Generalized Biological Modeling and Non-Automated Physical Adaptation

Dynamic data-driven adaptive modeling and real-time forecast of marine ecosystems [33] is an increasing research opportunity 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, and 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 perturbations, necessitates 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.

The authors have developed a preliminary version of a generalized, flexible biological model specifically designed for adaptive modeling and real-time ecosystem forecast (Fig. 14.3). Marine ecosystems function through a series of highly integrated interactions between biota, habitats, and dynamic links among food web

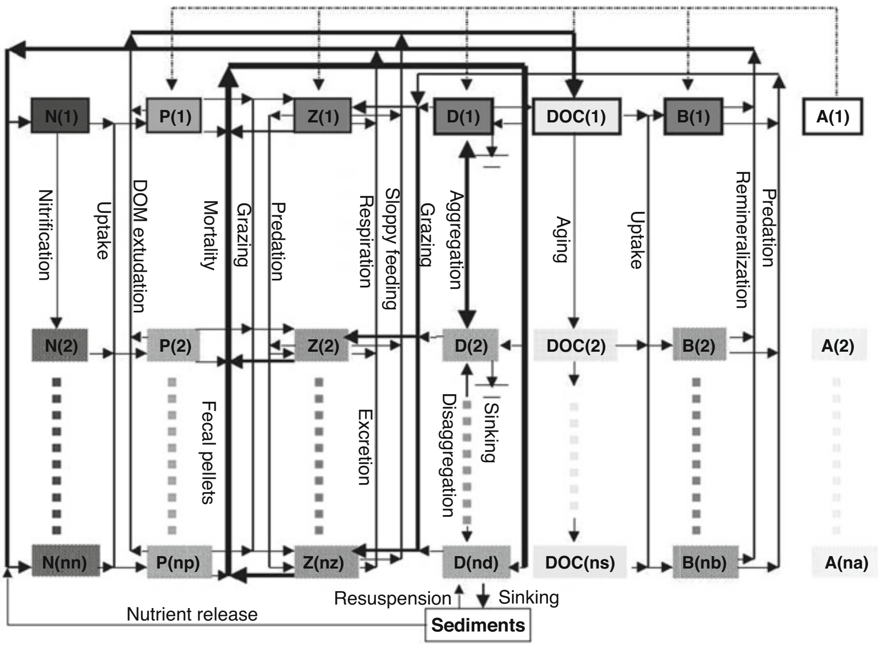

Fig. 14.3 Generalized biological model. N Nutrients; P Phytoplankton; Z Zooplankton; D Biogenic detritus; DOM Dissolved organic matter; B Bacteria; A Auxiliary state variables; *nn*, *np*, *nz*, *nd*, *ns* and *na* are the total numbers of state variables of the functional groups

components. Based on the trophic and biogeochemical dynamics for data collected $i = 1.,,,, n$, 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. However, in the generalized biological model developed by the authors, 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 software, 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, one 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 Nutrient-Phyroplancton-Zooplancton (NPZ) model. When the component number of detritus is assigned to 1 in the previous configuration, the generalized model will be an NPZ-detritus (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.

In the application, the state variables, model structures, and parameter values can change, at execution time, in response to field measurements, ecosystem function, and scientific objectives. All of these components of the model can be driven dynamically by data inputs. The Poseidon-DDDAS model has been coupled with HOPS. The forecasting system application includes the Monterey Bay area to study biological response to upwelling events at ecosystem level.

14.3.2.1 Monterey Bay Application

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^+, 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.

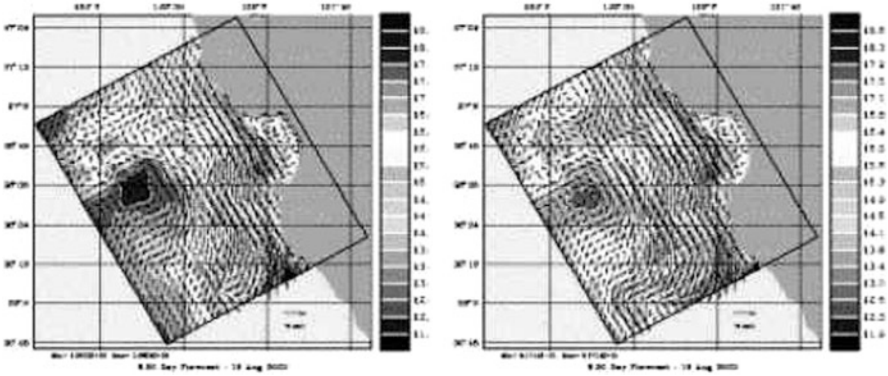

Fig. 14.4 Simulated surface water temperature (°C) before (left) and after (right) non-automated adaptation on August 11, 2003 during the AOSN-II experiment in Monterey Bay

The dynamic-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, autonomous underwater vehicle (AUV), and ship-based data [1] were deployed to collect data in real time. These data were dynamically assimilated into the 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 measured atmospheric fluxes to the upper layers of this model. As shown in Fig. 14.4, the new values for wind mixing clearly modified surface properties and improved the temperature fields and corresponding currents.

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, which was then used to initialize the biological simulation. The simulation was started on August 6, 2003, and stopped on August 11, 2003, making for 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 determined by currents and eddies. An anticyclone was simulated offshore from Monterey Bay. Corresponding filaments and fronts in biological distributions can be observed in Fig. 14.5.

In summary, DDDAS concepts were applied to physical parameterizations at execution time, and in real time [38]. The next steps improve the execution-time optimization of the physical parameterization and apply such DDDAS ideas for biogeochemical modeling in real time.

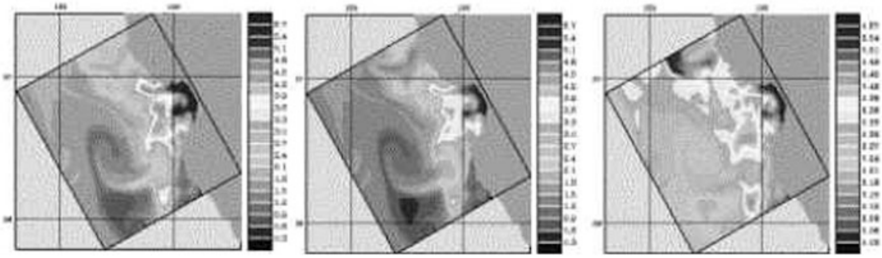

Fig. 14.5 Simulated field of total (left), eukaryote (middle), and prokaryote (right) chlorophyll ($\mu g/l$) on August 11, 2003 in Monterey Bay during an upwelling event

14.3.3 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, 47, 57]. The 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 the inversion process.

Estimation of the error of such an inversion process is useful in the acoustic modeling and assimilation framework [29, 62]. In addition, such error estimation is needed for the development of adaptive sampling (Sect. 14.3.1 and [61, 62, 64]), which aims to optimally reduce the uncertainty in the field estimates. The present acoustical-biological measurement methods allow for implementation in practice for acoustical-biological-physical estimation via judicious adjustment of the real-time acoustic sensing capability. Initial report and the technical details of the acoustic-biological models are reported in [47].

14.4 Components of the Poseidon System: Architecture Design and Progress

This section shows initial accomplishments in distributed/grid computing, user interfaces, and overall design of an automated DDDAS. The present design parameters are based 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 [65, 66, 62]. The subsequent subsections present some of the research issues, the resulting design developed by the authors, and implementations of the architecture of the evolving system.

14.4.1 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 hundred). The large number of forecasts 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 approaches in 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; and (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 and cloud computing-based framework [68] 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 [59] Toolkit). Such approaches are needed to accommodate the dynamic computation and data requirements of environments such as the present DDDAS-based Poseidon system.

The low-level computational strategy applied here was shaped by the often-conflicting targets of (i) maximizing computational performance, (ii) maintaining programming investment, and (iii) accommodating the needs of software developers. To avoid the resulting major discontinuity in code development [58], the present work uses the constituent domain science routines themselves rather than transforming them into the subroutine form suitable for classical component [7, 8] or Java agent-based distributed computing [26, 28] (which would also require more effort to integrate with a Globus-based Grid computing environment). Component interaction thus generally takes place via file input/output (I/O) within automated workflows. To address performance issues, on the other hand, the authors' strategy employed parallel approaches (using Message Passing Interface (MPI) and coupling frameworks [30]) for the 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 software in an elegant and efficient manner using function pointers and mixed-language programming.

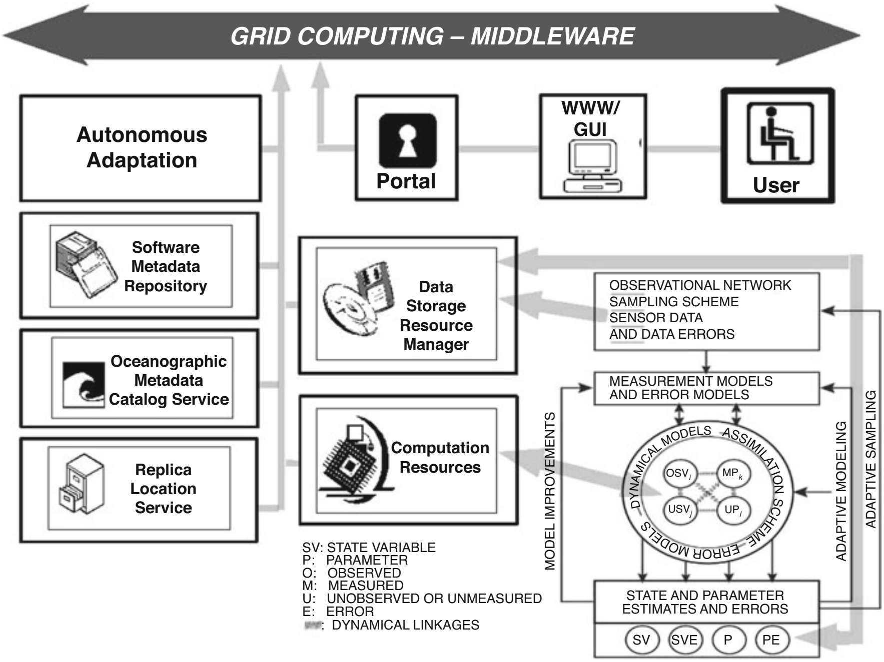

Fig. 14.6 An overview of the functional components of the Poseidon system and how it improves the efficiency and capability of the HOPS/ESSE system (schematized in the lower right corner)

A high-level view of the grid computing-based system architecture of Poseidon is provided in Fig. 14.6. 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, see also [38]).

The goal of the computational components (rectangular boxes in Fig. 14.6) 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 [13], allowing us to capitalize on new developments in that area. Grid computing will provide 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/Open Data Access Protocol (DAP) [9] enabled Data Storage Resource Managers while their associated location (and that of any valid cached copies) is recorded in a Replica Location Service (RLS [6]). The same data grid assets store the results of forecasts. The metadata for any observational and simulation datasets will be stored in an Oceanographic Metadata Catalog Service (MCS [56]), allowing for searching for datasets based on their content rather than their filename. The combination of MCS and RLS allows 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 (e.g., 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 Sect. 14.4.2.1) used for remote configuration of the computational tasks.

14.4.2 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, minimizing the underlying computational and management complexity from ocean scientist users. While a significant part of the intricacy of the use of grid computing middleware can be behind Web-enabled Computational Portals and Problem Solving Environments, the complicated (build- and run-time) configuration of the actual interdisciplinary computational components in the Dynamic Data Driven Poseidon Application System remains a challenge. Ubiquitous access (from remote sites, e.g., on ships) via a lightweight graphical user interface (GUI) was determined to be 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: such as for low-bandwidth connections and heterogeneous low-end mobile clients.

14.4.2.1 Generic Web User Interfaces

The process of designing Poseidon had to address the fact that HOPS (like other ocean applications, e.g., for physical oceanography ROMS [23] or for ocean acoustics OASES [55]) are, like most scientific applications, legacy[1] programs. The software 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

[1] The 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 being developed mainly in Fortran and in any case are command line and not GUI-driven.

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, is 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 these challenges is the extra complication of configuring the rebuilding of the code (e.g., via a set of preprocessing definitions specified in a Makefile, each with its own dependencies on others and on runtime configuration options) and one arrives at a scenario that is not suited for remote use over the Web.

The Poseidon system, as all DDDAS environments, requires efficient data interfaces and data handling. After examining various ways of dealing with this issue without costly changes for developers [58], and keeping in mind that the Poseidon system should allow for future handling of non-HOPS components without excessive recoding, the decision was made to avoid changing the codes or generating specialized GUIs; instead opting to describe their functionality and requirements – essentially "software metadata" – using the eXtensible Markup Language (XML) [14]. Thus, for practical purposes, a computer-readable manual for the codes was created, with information useful for checking option/parameter correctness (type, range, and dependencies) and producing properly formatted input files, scripts, Makefiles, and command lines. The authors developed a hierarchy of XML Schemata [54, 63] for their software metadata descriptions, attempting to cover as general a set of legacy applications as possible beyond the HOPS and acoustics binaries in Poseidon [4, 15, 16, 17]. A prototype Java-based tool, called LEGEND (LEGacy Encapsulation for Network Distribution) [19], was designed using a repository of software metadata and associated schemata, which automatically generates a validating GUI to produce scripts for building and running the binaries and allows for controlling their grid or local execution. Results of the XML Schemata approach are presented in [17].

14.4.2.2 Remote Visualization and Feature Extraction

The graphical output from the individual components of the Poseidon system is based on a variety of software: for example, *National Center for Atmospheric Research* (NCAR) Graphics [42] 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 [60]), 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 Network Common Data Form (NetCDF) self-describing portable file format that HOPS and most ocean modeling codes use. Based on scientists' usage patterns, three different visualization approaches are pursued as part of the present work:

1. To cover the standard set of 2D horizontal and vertical slice-based visualizations ocean scientists always look at, the LEGEND-configurable shell scripts are employed here, that automatically generate Web pages with the required results embedded as images. Such scripts use the existing tools (NCAR Graphics), thus leveraging the robust corpus of model-specific visualization work.
2. For more interactive and capable visualization work, Open Data Explorer (OpenDX) from IBM [43] and Java Explorer [27], are employed which (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 allows users to 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 improve situation awareness. Without feature extraction, the human operator needs to visually identify important dynamical events, which is vital for human-directed adaptive sampling. Visualization is an aspect that is important in DDDAS environments where both the model and the user are enabled to control the measurement process. The authors have been developing a suite of tools to automatically identify oceanic flow features such as eddies/gyres and upwelling and graphically present these results to enhance the effectiveness of the human forecaster and operations planner. For the more involved problem of vortex (eddy) identification an efficient two-stage algorithm has been developed that first identifies vortex cores and then locates the boundaries of the closed streamline region around them [21, 22]. The same tools, appropriately modified, can serve as key components in the fully automated adaptive sampling loop as part of a dynamic data-driven observation system.

14.5 Conclusions

This chapter provides an overview of a DDDAS-based system for rapid adaptive interdisciplinary ocean forecasting as implemented in the Poseidon system. 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. The work presented in this chapter describes important components of the system available for extensions and further development. Also presented are illustrative examples of interdisciplinary modeling and forecasting for DDDAS integration into ocean science research. This chapter highlighted automated adaptive modeling and sampling, fully coupled physical-acoustical-biological oceanography, and grid computing applications that afford to monitor ocean activity.

Acknowledgments This work was funded in part from NSF/ITR (under grant EIA-0121263), from DoC (NOAA via MIT Sea Grant) (under grant NA86RG0074), and from SMART/CENSAM initiative [5] (CENSAM, SMART, 2008). The authors thank the AOSN-II and MURI-ASAP teams. The ocean modeling results were funded in part from NSF/ITR (under grant 5710001319) and ONR (undergrantN00014-01-1-0771, N00014-02-1-0989 and N00014-97-1-0239). The authors also thank Dr. Darema for her support of their DDDAS research.

References

1. AOSN II/Monterey Bay 2003. http://www.mbari.org/aosn/MontereyBay2003.htm.
2. Berman, M.S., Green, J.R., Holliday, D.V. and Greenlaw, C.F. Acoustic determination of the fine-scale distribution of zooplankton on Georges Bank. Mar. Ecol. Prog. Ser.2002.
3. Buizza, R. and Montani, A. Targeting observations using singular vectors. Journal of the Atmospheric Sciences, 56(17), 2965–2985, 1999.
4. Chang, R.C. The Encapsulation of Legacy Binaries using an XML-Based Approach with Applications in Ocean Forecasting, M.Eng. thesis in Electrical Engineering and Computer Science, Massachusetts Institute of Technology,2003.
5. Center for Environmental Sensing and Modeling (CENSAM), http://censam.mit.edu/.
6. Chervenak, A., Deelman, E., Foster, I., Guy, L., Hoschek, W., Iamnitchi, A., Kesselman, C., Kunszt, P., Ripenu, M., Schwartzkopf, B., Stocking, H., Stockinger, K. and Tierney, B.Giggle: A framework for constructing scalable replica location services. In Proc. of SC2002 Conference, November 2002, IEEECS/ACM.
7. Common Component Architecture Forum: http://www.cca-forum.org/.
8. Common Object Request Broker Architecture: http://www.corba.org/.
9. Cornillon, P., Gallagher, J., and Sgourosy, T. OPENDAP: Accessing data in a distributed, heterogeneous environment. Data Science Journal, 2(5), 164–174, November2003. https://earthdata.nasa.gov/collaborate/open-data-services-and-software/api/opendap. Accessed, Aug 2020.
10. Curtin,T.B., Bellingham, J.G., Catipovic, J. and Webb, D. Autonomous Oceanographic Sampling Networks. Oceanography, 6(3), 86–94, 1993.
11. Darema, F.*et al.* NSF Sponsored Workshop on Dynamic Data Driven Ap- plication Systems. Technical report, National Science Foundation, 2000. http://www.cise.nsf.gov/cns/darema/dddas/dddasworkshoprprt.pdf.
12. Deleersnijder, E. and Lermusiaux, P.F.J.(Guest Eds.). Multi-Scale Modelling: Nested Grid and Unstructured Mesh Approaches, Editorial. Ocean Dynamics, 58, 335–336, Springer, 2008. DOI https://doi.org/10.1007/s10236-008-0170-5.
13. Earth System Grid (ESG). https://www.earthsystemgrid.org/.
14. eXtensible Markup Language: http://www.w3.org/XML/.
15. Evangelinos, C., Chang, R., Lermusiaux, P.F.J. and Patrikalakis, N.M. 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 and Y. Gorbachev (Eds), International Conference on Computational Science 2003 (ICCS 2003), volume 2660 of Lecture Notes in Computer Science, (pp. 375–384). Springer, 2003.
16. Evangelinos, C., Chang, R., Lermusiaux, P.F.J. and Patrikalakis, N.M. Web-Enabled Configuration and Control of Legacy Codes. International Journal of Cooperative Information Systems, 2004.

17. Evangelinos,C.,Lermusiaux,P.F.J.,Geiger,S.,Chang,R.C.,andPatrikalakis,N.M.,Web-Enabled Configuration and Control of Legacy Codes: An Application to Ocean Modeling. Ocean Modeling, 13, 197–220, 2006.
18. Foster, I. and Kesselman, C.,(Eds.) The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, 1999.
19. Geiger,S.K.LegacyComputingMarkupLanguage(LCML)andLEGEND–LEGacyEncapsulationforNetworkDistribution,S.M.inOceanEngineeringthesis,MassachusettsInstitute of Technology,2004.
20. Gloor,M.,Fan,S-M.,Pacala, S. and Sarmiento ,J. Optimal sampling of the atmosphere for purpose of inverse modeling: A model study. Global Biogeochemical Cycles, 14(1), 407–428, 2000.
21. Guo, D. Automated feature extraction in oceanographic visualization. M.Sc. in Ocean Engineering and EECS thesis, Massachusetts Institute of Technology, 2004.
22. Guo, D., Evangelinos, C., and Patrikalakis, N.M. Flow feature extraction in oceanographic visualization. In D. Cohen-Or, L. Jain, and N. Magnenat-Thalmann, (Eds),Proc. of Computer Graphics International Conference, CGI 2004, Crete, Greece, Los Alamitos, CA, June 2004. IEEE Computer Society Press.
23. Haidvogel, D., Arango, H., Hedstrom, K., Malanotte-Rizzoli, A.B.P. and Shchepetkin, A. Model evaluation experiments in the North Atlantic basin: Simulations in nonlinear terrain-following coordinates. Dyn. Atmos. Oceans, 32, 239–281, 2000.
24. Haley, P.J. Jr., Lermusiaux, P.F.J., Robinson, A.R., Leslie, W.G., Logutov, O., Cossarini, G., Liang, X.S., Moreno, P., Ramp, S.R., Doyle, J.D., Bellingham, J., Chavez, F., Johnston, S.. Forecasting and Reanalysis in the Monterey Bay/California Current Region for the Autonomous Ocean Sampling Network-II Experiment. Special issue on AOSN-II, Deep Sea Research, Part II. ISSN 0967-0645, DOI: https://doi.org/10.1016/j.dsr2.2008.08.010, 2009.
25. Heaney, K. D., Gawarkiewicz, G.,Duda, T.F., and Lermusiaux, P.F.J. Non-linear Optimization of Autonomous Under sea Vehicle Sampling Strategies for Oceanographic Data-Assimilation. Special issue on "Underwater Robotics", Journal of Field Robotics, Vol. 24(6), 437–448, 2007.
26. Houstis, C., Lalis, S., Christophides, V., Plexousakis, D., Vavalis, E., Pitikakis, M., Kritikos, K., Smardas, A. and Gikas C. 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, 2002.
27. Java Explorer: Enables web deployment of data explorer visualization solutions. http://www.sdsc.edu/dx/JavaExplorer.htm.
28. Java Native Interface: http://java.sun.com/j2se/1.4.2/docs/guide/jni/.
29. Lam, F.P, Haley, P.J. Jr., Janmaat, J., Lermusiaux, P.F.J., Leslie, W.G., and Schouten, M.W. At-sea Real-time Coupled Four-dimensional Oceanographic and Acoustic Forecasts during Battlespace Preparation 2007. Special issue of the Journal of Marine Systems on "Coastal processes: challenges for monitoring and prediction", Drs. J.W. Book, Prof. M. Orlic and Michel Rixen (Guest Eds.). DOI: https://doi.org/10.1016/j.jmarsys.2009.01.029, 2009.
30. Larson, J.W., Jacob, R.L., Foster, I. and Guo, J. The Model Coupling Toolkit. Technical Report ANL/CGC-007-0401, Argonne National Laboratory, April2001.
31. Lermusiaux, P.F.J. Error Subspace Data Assimilation Methods for Ocean Field Estimation: Theory, Validation and Applications, PhD thesis, Harvard University, 1997.
32. Lermusiaux,P.F.J.Evolvingthesubspaceofthethree-dimensionalmultiscaleoceanvariabil- ity: Massachusetts Bay. J. Marine Systems, 29, 385–422, 2001.
33. Lermusiaux, P.F.J, Evangelinos, C., Tian, R., Haley, P.J., McCarthy, J.J., Patrikalakis, N.M., Robinson,A.R.,andSchmidt,H.AdaptiveCoupledPhysicalandBiogeochemicalOceanPredictions: A Conceptual Basis. Refereed invited manuscript, F. Darema (Ed.), Lecture Notes in Computer Science, 3038, 685–692, 2004.
34. Lermusiaux, P.F.J., Malanotte-Rizzoli, P., Stammer, D., Carton, J., Cummings, J. and Moore, A.M., Progress and Prospects of U.S. Data Assimilation in Ocean Research. Oceanography, Special issue on Advances in Computational Oceanography, T. Paluszkiewicz and S. Harper, Eds., Vol. 19, 1, 172–183, 2006.

35. Lermusiaux, P.F.J., Chiu,C.-S., Gawarkiewicz, G.G., Abbot, P., Robinson, A.R., Miller, R.M., Haley, P.J., Leslie, W.G., Majumdar, S.J., Pang, A. and Lekien, F. Quantifying Uncertainties in Ocean Predictions. Refereed invited manuscript. Oceanography, Special issue on Advances in Computational Oceanography, T. Paluszkiewicz and S. Harper (Office of Nava lResearch), Eds., Vol. 19, 1, 92–105, 2006.
36. Lermusiaux, P.F.J. Uncertainty Estimation and Prediction for Interdisciplinary Ocean Dynamics. Refereed manuscript, Special issue of on Uncertainty Quantification. J. Glimmand G. Karniadakis, Eds. Journal of Computational Physics, 217, 176–199, 2006.
37. Lermusiaux, P.F.J., Haley, P.J.Jr., and Yilmaz, N.K. Environmental Prediction, Path Planning and Adaptive Sampling: Sensing and Modeling for Efficient Ocean Monitoring, Management and Pollution Control. Sea Technology, Vol. 48(9), 35–38, 2007.
38. Lermusiaux, P.F.J, Adaptive Modeling, Adaptive Data Assimilation and Adaptive Sampling. Refereed invited manuscript. Special issue on Mathematical Issues and Challenges in Data AssimilationforGeophysicalSystems:InterdisciplinaryPerspectives.C.K.R.T.JonesandK. Ide, Eds. Physica D, Vol 230, 172–196, 2007.
39. Lermusiaux, P.F.J. and Robinson, A.R. Data assimilation via Error Subspace Statistical Estimation. Part I: Theory and schemes. Month. Weather Rev., 127, 1385–1407, 1999.
40. Lermusiaux, P.F.J., Robinson, A.R., Haley, P.J. and Leslie, W.G. 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, 2002.
41. Logutov, O.G. and Lermusiaux, P.F.J. Inverse Barotropic Tidal Estimation for Regional Ocean Applications. Ocean Modeling, 25, 17–34. doi:https://doi.org/10.1016/j.ocemod.2008.06.004, 2008.
42. NCAR Command Language and NCAR Graphics. https://ncar.ucar.edu/. Access Aug, 2020.
43. Open visualization Data eXplorer (OpenDX). http://www.opendx.org/.
44. Patrikalakis, N.M. Poseidon: A Distributed Information System for Ocean Processes http://czms.mit.edu/poseidon/
45. Patrikalakis, N.M., Abrams, S.L., Bellingham, J.G., Cho, W., Mihanetzis, K.P., Robinson, A.R., Schmidt, H. and Wariyapola, P.C.H. The digital ocean. In Proc. of Computer Graphics International, GCI '2000, Geneva, Switzerland, (pp. 45–53). Los Alamitos, CA: IEEE Computer Society Press, 2000.
46. Palmer,T.N., Gelaro, R., Barkmeijer, J. and Buizza, R. Singular vectors, metrics, and adaptive observations. American Meteorol. Soc., February1998.
47. Renard, B. Inversion of Acoustic Zooplankton Measurement for Adaptive Physical- Biological Ocean Forecast. M.Sc. in Ocean Engineering thesis, Massachusetts Institute ofTechnology, 2003.
48. Robinson, A.R. Harvard: Ocean Prediction System (HOPS) http://oceans.deas.harvard.edu/HOPS/HOPS.html.
49. Robinson, A.R. 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, 1999.
50. Robinson, A.R. and Glenn, S.M. Adaptive Sampling for Ocean Forecasting. Naval Research Reviews, 51(2), 28–38, 1999.
51. Robinson, A.R. and Sellschopp, J. Rapid assessment of the coastal ocean environment, in N. Pinardi and J.D. Woods (Eds), Ocean Forecasting: Conceptual Basis and Applications, (pp. 203–232), Springer-Verlag, 2000.
52. Robinson,A.R. and the LOOPS Group. Real time 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 November1999,NewOrleans,LA,(pp.30–35),AmericanMeteorologicalSociety, 1999.
53. Robinson, A.R. and Lermusiaux, P.F.J. Prediction Systems with Data Assimilation for Coupled Ocean Science and Ocean Acoustics, Proceedings of the Sixth International Conference on Theoretical and Computational Acoustics (A. Tolstoy, et al., editors), World Scientific Publishing, 325–342. Refereed invited Keynote Manuscript, 2004.

54. Roy, J. and Ramanujan, A. XML schema language: Taking XML to the next level. ITProfessional, 3, 37–40, 2001.
55. Schmidt, H. and Tango, G. Efficient global matrix approach to the computation of synthetic seismograms. Geophys. J. R. Astr. Soc. 84, 1986.
56. Singh,G.,Bharathi,S.,Chervenak,A.,Deelman,E.,Kesselman,C.,Manohar,M.,Patil, S. and Pearlman, L. A metadata catalog service for data intensive applications. In Proc. of SC2003 Conference, Phoenix, AZ, November 15–21 2003. IEEE CS/ACM, 2003.
57. Stanton, T.S. and Chu, D. Review and recommendations for the modelling of acoustic scattering by fluid-like elongated zooplankton: euphausiids and copepods. ICES J. Mar. Sci, 57, 793–807, 2000.
58. Terekhov, A. and Verhoef, C. The realities of language conversions. IEEE Software, 111–124, 2000.
59. The Globus Project: http://www.globus.org/.
60. Virtual Network Computing (VNC). http://www.realvnc.com/.
61. Wang, D., Lermusiaux, P.F.J., Haley, P.J., Leslie, W.G., and Schmidt, H. Adaptive Acoustical-Environmental Assessment for the Focused Acoustic Field-05 At-sea Exercise, Proceedings of IEEE/MTS Oceans'06 Conference, Boston, MA, September 18–21. 6pp, 2006.
62. Wang, D., Lermusiaux, P.F.J., Haley, P.J., Eickstedt, D., Leslie, W.G., and Schmidt, H. Acoustically Focused Adaptive Sampling and On-board Routing for Marine Rapid Environmental Assessment.SpecialissueofJournalofMarineSystemsonCoastalprocesses:challenges for monitoring and prediction, Drs. J.W. Book, Prof. M. Orlic and Michel Rixen (Guest Eds). DOI: https://doi.org/10.1016/j.jmarsys.2009.01.037, 2009.
63. XML Schema Specification: http://www.w3.org/XML/Schema/.
64. Xu, J., Lermusiaux, P.F.J., Haley, P.J. Jr., Leslie, W.G., and Logutov, O.G. Spatial and Temporal Variations in Acoustic propagation during the PLUSNet'07 Exercise in DabobBay. AcousticalSocietyofAmerica,ProceedingsofMeetingsonAcoustics(POMA).155thMeet- ing, Vol. 4. 11pp. DOI: https://doi.org/10.1121/1.2988093, 2008.
65. Yilmaz, N.K., Evangelinos, C., Patrikalakis, N.M., Lermusiaux, P.F.J., Haley, P.J., Leslie, W.G., Robinson, A.R., Wang,D., and Schmidt, H. Path Planning Methods for Adaptive Sampling pling of Environmental and Acoustical Ocean Fields, Oceans 2006, 6pp, Boston, MA, 18–21 Sept. 2006. https://doi.org/10.1109/OCEANS.2006.306841.
66. Yilmaz, N.K., Evangelinos, C., Lermusiaux, P.F.J., and Patrikalakis, N.M. Path Planning of Autonomous Underwater Vehicles for Adaptive Sampling Using Mixed Integer Linear Programming. IEEE Transactions, Journal of Oceanic Engineering. Vol. 33 (4), pp. 522–537. DOI: https://doi.org/10.1109/JOE.2008.2002105, 2008.
67. Darema, F.: Dynamic data driven applications systems: A new paradigm for application simulations and measurements. In: ICCS. Springer (2004)
68. Evangelinos, C., P.F.J. Lermusiaux, J. Xu, P.J. Haley, and C.N. Hill, 2011. Many Task Computing for Real-Time Uncertainty Prediction and Data Assimilation in the Ocean. IEEE Transactions on Parallel and Distributed Systems, Special Issue on Many-Task Computing, I. Foster, I. Raicu and Y. Zhao (Guest Eds.), 22, doi: https://doi.org/10.1109/TPDS.2011.64