Progress toward autonomous ocean sampling networks

Thomas B. Curtina,*,1, James G. Bellinghamb,2

1 5461, North Four Mile Run Drive, Arlington, VA 22205, USA
2 Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039-9644, USA

ARTICLE INFO

Keywords:
Autonomous Ocean sampling Network

ABSTRACT

The goals of the Autonomous Ocean Sampling Network (AOSN) are reviewed and progress toward those goals is assessed based on results of recent, major field experiments. Major milestones include the automated control of multiple, mobile sensors for weeks using spatial coverage metrics and the transition from engineering a reliable data stream to managing the complexities of decision-making based on the data and the possibilities of timely feedback.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

The Autonomous Ocean Sampling Network (AOSN) concept (Curtin et al., 1993) leverages autonomous mobile platforms and assimilative dynamic models to observe and predict dynamic ocean fields. While advances in autonomous underwater vehicles have enabled distributed observation of dynamic processes by fleets of robotic vehicles, continuous fields must still be realized from a limited number of discrete observations. Dynamical models can interpolate and extrapolate observations deterministically, and thus can generate continuous realizations of ocean fields from discrete measurements. Realizations based on statistical models can provide continuity for random variables. The temporal evolution of three-dimensional ocean fields results from both deterministic and stochastic processes, and thus even with perfect observations, prediction skill will deteriorate in time. The great challenges of AOSN revolve around learning how to better sample the ocean field, and improving the skill of assimilative models for synthesis and prediction of the evolution of those same fields.

Historically, maps of ocean fields have been based on arrays of fixed-point (moorings, stations) and Lagrangian (floats) data, and, more recently, on quasi-synoptic cross-sections (tow-yos). The time scale of experiment design, deployment, data acquisition, and analysis was years. Such experiments yielded time series that revealed a broadband cascade in the ocean energy spectrum. High-resolution, local maps associated with specific expeditions onboard research vessels yielded initial insights into processes governed by spatial gradients. Observations from spacecraft and with surface radar provided powerful new mapping capability for near-surface ocean fields. The current Argo array provides for the first time persistent, comprehensive subsurface observations on a global scale. The sampling resolution of this Lagrangian array, however, is controllable only by initial seeding. Plans to expand and stabilize an ocean observing system have been formulated and are slowly being implemented.

Pragmatic calculations using true ocean time and space scales and the real cost and complexity of in-situ observations show that a systemic challenge of reducing error in ocean field estimation is sparse sampling. The AOSN initiative (Curtin et al., 1993; Curtin and Bellingham, 2001) was launched to develop new tools and methodologies to address the sparse sampling problem and reduce errors in ocean field estimation to enable definitive hypothesis testing. The premise is that there is a significant advantage to adapting the distribution of observations on time and space scales comparable to the processes driving the variability, and to persisting long enough to accumulate robust statistics on coherence scales. This premise is supported by the growing literature on targeted observations spanning a wide range of disciplines including meteorology.

For this introduction, we draw on experience from the AOSN II field program in August 2003 in Monterey Bay. We also consider follow-through from that field program, as manifested in the Adaptive Sampling and Prediction (ASAP) program and the Monterey Bay 2006 field program (MB06). AOSN efforts have been directed toward three principal contributions:

(1) Multiple, mobile sensors that can resolve synoptic fields and spatial gradients to a desired level of precision: Mapping of transient spatial fields with minimum error requires network-class autonomous vehicles that are in a practical cost-size envelope (Curtin et al., 2005). Staying within this envelope enables enough in-water testing within a realistic budget to
achieve acceptable reliability and ultimate use of such vehicles. Since mobile, platforms with persistence are slow due to energy constraints, their trajectories can be greatly influenced by ambient currents. If positioned poorly, they can be placed disadvantageously; for example, clustered in a region far from an evolving area of interest. Thus the performance of an AOSN is intimately connected with feedback control and decision-making processes.

(2) Control of mobile sensor arrays in a feedback loop with response time sufficient to stabilize and reduce error in evolving mapped fields: This loop not only provides a means to constrain errors in mapping gradients with multiple vehicles using local linearization, but also provides the flexibility to respond quickly to unexpected anomalies. Underlying this need is a fundamental design question: where should control authority reside in the system? Communications are not always readily available, and consequently mobile platforms must have some level of control authority. Balancing this is the reality that an individual vehicle will not be privy to information generated by other AOSN components, and thus may make non-optimal decisions if not controlled centrally. Finding the right balance of autonomy for individual assets and investing in the right level of communications to support collective planning remain challenging problems.

(3) Decision making that provides a range of options for the experimenter faced with uncertainty: Such options are easy to understand when directly connected to testing hypotheses or assessing the skill of predictive tools such as numerical models. Recalling that AOSN activities involve collaboration of teams of PIs, the availability of well-framed options helps building consensus on the best course of action when intuition is weak, there is disagreement among experts with similar objectives, or there are a variety of competing objectives. Practical experience from the AOSN II and MB06 field programs taught that different goals can often be accommodated within sampling plans. Thus the premature down-select of objectives can be counterproductive. Once the various sampling needs are understood, it is often possible to take advantage of the multiplatform nature of the observation system to satisfy multiple investigative needs.

The sensitivity of AOSN performance to real-time decision making creates many demands, not all of which are technical. AOSN, as a system of systems, depends on individual elements operating as components of an integrated system (Fig. 1). Many of the components deployed to date have been developed by and are operated by individual research groups, each with their own agendas. In a traditional field program, investigators depend
primarily on data produced by their assets for their subsequent scientific analyses, perhaps collaborating with other investigators to pool data. Within the context of an integrated system, the performance of the overall observation system may impose requirements on platforms that are not aligned with the research interests of the operating principal investigator. For example, a modeler may want to deploy observational assets to monitor boundary conditions at the periphery of the domain that may compromise sampling of processes in the interior. Such circumstances impose a requirement that core AOSN observation elements be operated collectively while insuring that individual investigators’ interests are represented, and data be available to all.

An AOSN system must capture and communicate the state of the observational system, the stages of data processing, the estimated state of the ocean based on observations, the synthesized ocean as realized by assimilative models, and the prediction of future states of the ocean. Most important is the evaluation of uncertainties in all of these factors. The system must also provide a framework for understanding risk and consequences for different possible courses of action.

2. State-of-the-art

In current AOSN systems, first-generation mobile sensors are operating reliably and persistently, and can be considered automated if not autonomous. In parallel with the maturation of hardware capabilities, a growing understanding of how to design effective ocean observing systems has come. Efforts in the mid-1990s primarily focused on getting individual hardware elements to function properly. In the late 1990s, field programs began to operate multiple observation system elements, and the complexity of the trade space for observation system design began to come into focus. The AOSN II experiment in 2003 demonstrated that individual elements, including models, were mature enough to be operated as a collective system. As part of the ASAP program in 2006, control of an array of platforms (gliders) constrained to a fixed sampling pattern for a month was demonstrated with no person in the loop (Paley et al., 2008). However, at the same time the sensitivity of system performance to its cyber infrastructure created a need to improve frameworks for sharing, exploring, and understanding data and model output in a timely fashion (Godin, 2006).

A common thread of growing importance is the critical role of data and knowledge management. Decision-making processes remain human intensive. Knowledge of the system includes scheduling of available assets, resource constraints, and key metadata for each observation. As observations increase, synthesizing and interpreting the information become more challenging. A variety of state variables are being extracted from a diversity of measurements on a heterogeneous mix of platforms. The priority assigned to observing or deriving particular state variables depends on particular objectives. In a large field program, diverse investigators will generate an equally diverse set of priorities. Thus the decision framework encompasses a multi-dimensional space.

Analysis of the 2003 experiment provided spatial and temporal statistics on variability that allows us to quantify synoptic performance of different survey system configurations. The integration of observations into models for the purpose of comparing outputs or for ensemble forecasting raises other issues including synchronization of system processes (e.g., comparison of model results from identical time evolutions), data assimilated as a function of time, assimilation techniques used, boundary conditions imposed, and underlying dynamical assumptions and parameterizations.

Sophisticated methods for multi-objective optimization are available, and will be implemented in future AOSN systems. Serious debate continues on the merits of adaptive sampling versus following a fixed observation sequence (e.g., occupying repeated stations). The a priori survey decision process is driven by available information, which can range from a nearly complete absence of information other than oceanographic intuition, to knowledge of spatial and temporal scales, to an understanding of patterns of variability. The reward of adaptive sampling lies in the ability to respond to and refine understanding of processes under study. The risk is that the process of adaptation may result in a data set that is difficult to interpret. Thus the integrity of the reconstruction process of the total field is a key to the perceived utility of adaptive sampling. How accurate is the understanding of the state of the fluid as it is evolving in space and time? This judgment determines whether adaption risk is worth the potential return.

Evidence for a payoff in predictive skill has been growing. For example, Montani et al. (2007) calculate that short-range Atlantic storm tracking prediction errors (up to day 2) are reduced on average by 15%, with a maximum error reduction of about 37%, if observations coincide with the regions over which the singular vectors have been optimized. Gelaro et al. (2000), in an analysis of North Pacific winds, conclude that the early stages of error growth in most numerical weather forecasts are dominated by a relatively small number of unstable structures, and that preferentially reducing analysis errors that project onto these structures can produce significant improvements in forecast skill. But such projections must be done with attention to detail. In analyzing the impact of adaptive observations in a case of poorly forecast North Atlantic cyclogenesis, Bergot et al. (1999) concluded that targeted observations showed great promise, but that current assimilation systems, such as 3DVAR, require all the structure of the target to be well sampled to have a significant beneficial effect; sampling only the extremum does not suffice. In the ocean, Lermusiaux (2007) has used adaptive modeling approaches based on simplified maximum likelihood principles to calibrate parameter values and model structures, and shows that error estimates, ensemble sizes, error subspace ranks, covariance tapering parameters, and stochastic error models can be calibrated by such quantitative adaptation.

The evaluation of sampling strategies depends on the ability to produce quantitative metrics for sampling performance, which in turn depends on having a clear sense of what aspects of the observed field are important. For example, the root-mean-square error of the measured field is a common metric used to evaluate the performance of an observation system. However, gradients in scalar quantities may have a critical importance, in, for example, the calculation of current from the derivative of the density field. In these cases the estimation will be sensitive to not just the error in the sampled density field, but also in the ordering of measurements of the density. On other occasions, the important quantities to measure may be fluxes or budgets of specific parameters such as heat or nutrients. A number of techniques exist for determining the sensitivity of meteorological model forecasts to specific assimilated observations (e.g., Bishop et al., 2001). On average, targeted observations based on these techniques have been shown to improve forecast skill for state variables. When risk is included as a metric, state variable may be associated with different utility functions and conditioned by expected values. When competing objectives are at issue, equilibria based on game theory may be the best approach. The most useful system in an uncertain environment will include all these tools as options to be exercised.

Even with the large number of observational assets deployed in the AOSN II experiment, decisions were made in the face of...
significant uncertainty. Sensor precision, data density, and modeling errors all contributed to the uncertainty. Observations were sparse compared to the spatial and temporal scales of variability. To interpolate between observations and extrapolate in time, three models were run in parallel on an operational cycle: the Regional Ocean Model System (ROMS), the Harvard Ocean Prediction System (HOPS), and the Navy Coastal Ocean Model (NCOM).

ROMS is a free-surface, hydrostatic, primitive equation ocean model that uses stretched, terrain-following coordinates in the vertical and orthogonal curvilinear coordinates in the horizontal. It was based on the S-coordinate Rutgers University Model (SCRUM) (Song and Haidvogel, 1994). Current ROMS features include high-order advection schemes; accurate pressure-gradient algorithms; several subgrid-scale parameterizations; atmospheric, oceanic, and benthic boundary layers; biological modules; radiation boundary conditions; and data assimilation. HOPS is a primitive equation dynamical model supported by data-gridding routines, initialization and assimilation field preparation routines, visualization software, data preparation codes, and topography-conditioning software (Robinson, 1999). NCOM is also a primitive equation model with a free surface. The physics and numerics of NCOM are based largely on the Princeton Ocean Model (POM) (Blumberg and Mellor, 1987), with some aspects from the Sigma-2-level Model (Martin, 2000). The model has been used for modeling of estuaries, coastal regions, and open oceans. Additional modeling details and model comparisons can be found at [http://www.ocean-modeling.org/docs.php](http://www.ocean-modeling.org/docs.php) and at [http://www.mbari.org/oasni/](http://www.mbari.org/oasni/). Models, like sensors, have characteristic error sources. Using several model outputs helps to bound modeling precision, but evaluating model accuracies has proven to be challenging.

In the AOSN-II experiment, data management emerged as a central issue. While investigators agreed to share data and to make data visualization products available on the web as quickly as possible, the process was time consuming. In an attempt at standardization, a data format convention was selected, and the responsibility placed on data originators to translate their output to those formats; however, this proved unrealistic. In meeting these familiar challenges, much was learned. Productive collaboration resulted in a greater understanding of both methodology and phenomenology. This volume captures a representative subset of that understanding, and many of the papers imply strategies for effective adaptive sampling.

Davis et al. (2008) analyze the trade-offs in routing strategies for underwater gliders whose speed through the water is often less than ambient ocean current speeds. For steady velocity fields, a ray-based trajectory model provides a general approach for minimizing the travel time across strong current shears. The methodology for traversing eddies rapidly will be of great utility in operational areas with high mesoscale energy. The more general sampling problem is addressed using an objective mapping skill metric (minimum error variance) based on the measurement covariance matrix (Breherton et al., 1975). In theory, this metric is elegant but there are many practical limitations to its implementation. The study of Davis et al. (2009) is grounded in practical realities, and presents workable solutions to limitations in communication, computation, and multiple platform interaction. Of most significance, however, is the underlying motivation to maintain a certain level of informed intuition in the sampling and analysis process. The process starts by defining ideal sampling tracks and choosing off-track penalty weights and bias vectors. Control is based on the skill gradient with respect to the glider’s next position if glider separations are on the order of the correlation scale (the intuition behind the ideal sampling tracks). If separations are much larger than the correlation scale, control based on physical separation is more effective in maintaining uniform coverage (Leonard et al., 2007). Advantages of the simple skill-gradient control method include manageable implementation, efficiency in steering toward defined tracks, and the ability to coordinate arrays of multiple vehicles to maintain spacing. Adaptive sampling in this context occurs through the reformulation of new ideal tracks by an intelligent analyst.

Chao et al. (2008) document the performance of the California coast and nested ROMS model, and describe the challenges of delivering a timely, operational ocean forecast. The ROMS, HOPS, and NCOM that were run on an operational cycle during the AOSN-II experiment often disagreed significantly in their forecast fields, limiting their use in guiding adaptive sampling. Possible causes include different data assimilation procedures, run-time synchronization, and boundary condition differences, as well as differing subgrid-scale parameterizations. Predictive skill was measured by a weighted average of the bias, a root-mean-square error, and a pattern correlation coefficient of model estimates. For hypothesis testing, the forecasts of competing models were compared based on their respective data–forecast misfits. For each candidate model, uncertainty bounds can be computed based on the small sample sizes because error variances often converge faster than covariances (Lermusiaux et al., 2004).

Lermusiaux (2007) concludes that quantitative adaptive filters may not be as useful in oceanography as they are in engineering applications where the number of independent observations is typically large when compared to the number of control parameters. However, adaptive schemes can still be quite effective in a number of oceanic applications where prior estimates can often be quite far from reality. The adaptation of models, error models, and assimilation schemes is then required. The use of adaptive schemes in oceanography is recent and many research questions remain (Zhang et al., 2007). Quantifying and automating the learning process of both the ocean researcher and operators in the ocean environment should become more and more fruitful in the years to come.

Shulman and Paduan (2008) quantify the benefit of assimilating surface current radar data even when high-resolution wind-stress fields are available. The critical factor is the horizontal divergence in the surface velocity field that drives interior motions. Shadden et al. (2008) numerically integrate HF-radar-derived surface currents to map the flow into regions of separable dynamics (Lagrangian coherent structures). This information can help maximize the persistence of a drifter within a given domain, which could be an important factor in a sampling decision based on an area coverage metric.

Liang and Robinson (2009) demonstrate how the complex gradients of the wind-driven coastal ocean can be partitioned into organized structures and how energy cascades from the large-scale structure to the small-scale structure through mixed barotropic/baroclinic instabilities off Point Sur during upwelling events and in northern Monterey Bay during relaxations. The associated surface eddy kinetic energy, driven by upwelling wind events and nonlinear instabilities during summer, comes to full closure in the deep ocean in late summer and fall, moving offshore as far as 127 W where the vertical shear flow is transformed to the vertical mean flow through nonlinear processes associated with baroclinic instability (Haney and Hale, 2001). Such upshifting and downshifting of energy across the wavenumber spectrum suggest that sampling resolution be adapted on a number of time scales from seasonal to synoptic.

On the meteorological synoptic scale, Doyle et al. (2008) analyze the forcing associated with upwelling favorable winds and find that offshore wind-stress curl patterns influenced by coastal topography may produce surface divergences comparable...
to those at the coastline. Not surprisingly, the 3-km resolution model has more realistic structure than the 9-km one. On the 1–10-day time scale, coastal ocean adaptive sampling may best be conditioned by gradients in forecast wind fields if upwelling is the principal variable of interest.

On the 10-day to 1-month time scale, Ramp et al. (2008) have documented the effect of the offshore meanders and eddies of the California Current System on modulating the seaward excursion of upwelled water masses. This modulation suggests a nearshore sampling strategy related to larger offshore structures in the flow. Such interactions are also critical in determining the boundary location and conditions for data assimilative, nested ocean models.

On diurnal time scales, Rosenfeld et al. (2008) and Wang et al. (2008) address issues in including tides with other forcing mechanisms in nested, data-assimilating, primitive equation ocean models. The amplitudes and phases of the eight major tide constituents are well-simulated; however, the error in barotropic tidal current exceeds 30%. Furthermore, model-derived barotropic tidal currents cannot be validated over large spatial scales using long time series of HF-radar-derived surface currents due to the small-scale variability introduced by internal tides. Surface tidal currents are shown to be sensitive to small changes in stratification. A model sensitivity analysis constrained by a maximum acceptable error in tidal current could be used to prescribe associated density field requirements for assimilation. Such dynamically driven adaptive sampling relies on sophisticated models to couple the velocity and density fields on diurnal time scales.

Proper sampling becomes even more critical for understanding and prediction when biological and chemical fields as well as fields of physical variables are considered. Johnston et al. (2008) discuss a persistent (days), extensive (30–60 km) structural feature observed in coastal upwelling zones: thin layers (<5 m) of high phytoplankton concentration. Although current shear, stratification, and isopycnal compression contribute to thin-layer formation, their high productivity may be due to successful exploitation of an ecological niche. The interplay between biological and physical effects in causing non-uniform plankton distributions remains an intriguing question that will only be answered through more sophisticated, four-dimensional sampling. An example of such sampling is described in Moline et al. (2008) who exploit the differences in bioluminescence flash kinetics between dinoflagellates and zooplankton to estimate the relative abundances of the two groups in a given domain sampled by an autonomous vehicle.

### 3. Prospectus

AOSN system development sits squarely on the interface between science and engineering and is multi-disciplinary in its potential applications. A central aim has been to maintain a productive spiral in which new sampling tools have been developed to further understand ocean processes, and greater understanding has then led to further refinement of the tools. Much of the initial engineering progress was reported in a special issue of IEEE Journal of Oceanic Engineering (Curtin and Bellingham, 2001). This current volume continues to report progress with emphasis on the associated ocean science. Advances in science accomplished with AOSN-developed tools are now widely published in less-concentrated formats. Much empirical evidence supports a 10-year time scale to develop reliable, in-situ, ocean instrumentation. The tools in the AOSN system are no exception. First-generation hardware is now working reliably and second-generation prototypes are being designed and built. As the system evolves further, software will be an ever more dominant component in the development cycle.

AOSN progress is built on advances in global telemetry, the global positioning system, the internet, web-based applications, lightweight sensors, low-power microprocessors, compact memory, higher-capacity batteries, and stronger composite materials. All of these component technologies are likely to continue advancing rapidly, providing unprecedented opportunities for ocean sampling. As the technology becomes more empowering, the challenges associated with the decision-making process, both human and automated and combinations of the two, will grow. How will networks of mobile sensors be managed effectively to advance both science and address societal needs? How will return on investment be articulated and quantified to motivate capitalization of these assets?

The nature of interactions of investigators has changed greatly over a decade of AOSN research. Perhaps the greatest change came in 2003 during the AOSN II experiment. AOSN PIs agreed to let an operational team control the disposition of all the observational assets and to share all data among team members. Further, graphical data and model products were published to an open web site immediately. The contributions of individual investigators were protected by instituting policies that outlined responsibilities of data users to acknowledge, or in some cases, include as co-authors, the generators of data. With these protections in place and supported by the data system, data from the experiment were released on the web within months of the experiment. In 2006 these same arrangements were adopted as a matter of course. Further, publicly accessible web sites were employed for sharing data and analysis products, and for discussing new operational plans. In effect, the AOSN program promotes a culture, supported by appropriate policies, that encourages data accessibility. This is a promising development, as openness coupled with recognition greatly increases the impact of observing system operations, and rewards the investigators whose contributions made the AOSN possible.

### References


