

Nonlinear Optimization of Autonomous Undersea Vehicle Sampling Strategies for Oceanographic Data-Assimilation

.....

Kevin D. Heaney

*Ocean Acoustical Services and Instrumentation Systems, Inc.
11006 Clara Barton Dr.
Fairfax Station, Virginia 22039*

Glen Gawarkiewicz and Timothy F. Duda

*Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543*

Pierre F. J. Lermusiaux

*Division of Engineering and Applied Sciences
Harvard University
Pierce Hall, 29 Oxford St.
Cambridge, Massachusetts 02138*

Received 2 May 2006; accepted 22 December 2006

The problem of how to optimally deploy a suite of sensors to estimate the oceanographic environment is addressed. An optimal way to estimate (nowcast) and predict (forecast) the ocean environment is to assimilate measurements from dynamic and uncertain regions into a dynamical ocean model. In order to determine the sensor deployment strategy that optimally samples the regions of uncertainty, a Genetic Algorithm (GA) approach is presented. The scalar cost function is defined as a weighted combination of a sensor suite's sampling of the ocean variability, ocean dynamics, transmission loss sensitivity, modeled temperature uncertainty (and others). The benefit of the GA approach is that the user can determine "optimal" via a weighting of constituent cost functions, which can include ocean dynamics, acoustics, cost, time, etc. A numerical example with three gliders, two powered AUVs, and three moorings is presented to illustrate the optimization approach in the complex shelfbreak region south of New England. © 2007 Wiley Periodicals, Inc.

1. INTRODUCTION

The problem of accurately predicting the ocean temperature and salinity fields for use by sonar performance prediction algorithms is addressed (Abbot & Dyer, 2002; Heaney & Cox, 2006; Robinson & Lermusiaux, 2004; Robinson, Lermusiaux & Sloan, 1998). Current meso-scale oceanographic models have incorporated much of the relevant physics so that if initial and boundary conditions are known well, these models can provide useful predictions of the ocean temperature, salinity, and velocity fields. Recent developments in data assimilation (optimal interpolation, extended Kalman filtering, and error subspace methods) have shown that models can predict the ocean state better with the inclusion of field data, and the use of models can improve mapped fields generated from measured data.

The challenge to this situation is limited resources. Working in the ocean is expensive and technically challenging. Methodologies that determine ideal ways to measure the ocean with the fewest wisely placed sensors are needed. An integrated optimization-assimilation-modeling system is being developed to perform optimal ocean sampling and ocean prediction. The system involves the combination of three subsystems or algorithms: (1) an ocean model with data-assimilation capabilities [in our case the Harvard Ocean Prediction System (HOPS) (Haley, Lermusiaux, Leslie & Robinson, 2006; Lermusiaux et al., 2006; Patrikalakis et al., 2004; Robinson et al., 1998)], with the Error Subspace Statistical Estimation algorithm (Lermusiaux, 2004) (ESSE)); (2) an ensemble of measurement platforms (i.e., moorings, gliders, powered autonomous undersea vehicles); and (3) a nonlinear constrained global optimization algorithm, which determines the optimal placement of these sensors. In this paper, we present the measurement subsystems and the nonlinear constrained global optimization subsystem.

The optimization approach is an application of the Genetic Algorithm (GA) (Goldberg, 1989). The GA solves complex global optimization (minimization in our case) problems by generating a population of individuals (possible solutions) and within each successive population of individuals (iteration) choosing the survivors based upon the value of the complex cost-function surface. Mutation and crossover, as well as other analogies to natural selection, are implemented to prevent the solution leading to a local minimum. The implementation of the GA solu-

tion requires a model for the search parameter space and the definition of a cost function to be minimized. The parameter space will involve a vector for each platform defining its search geometry. Examples of vector values for a sensor deployment are initial position, initial direction, range of sample path, and number and direction of turns. Once a sensor deployment scheme is defined (an individual in the GA nomenclature) the cost function is evaluated using the values from the oceanographic model that this particular deployment of sensors would sample. The cost function is evaluated for this individual, and then the GA generates another set of search laydowns based upon these results. In order to define the global cost function for the GA to optimize, a set of scalar cost function constituents has been developed. The solutions of the GA optimization are only as optimal as the cost function that is defining them. It is a current research program to build and define the best set of cost functions for this problem.

Significant work has been done in the past 15 years on developing and deploying AUV sampling networks (Curtin, Bellingham, Catipovic & Webb, 1993). The issues associated with optimal deployment and oceanographic sampling have also been covered (Bellingham & Wilcox, 1996; Wilcox, Bellingham, Zhang & Baggeroer, 2001). Real-time, adaptive experiments using AUVs and ocean models have been conducted. Other powerful techniques for adaptive control of underwater assets based on coverage metrics are described in Leonard (Leonard et al., 2006; Ogren, Fiorelli & Leonard, 2004; Paley, Zhang & Leonard, 2006).

The outline of this paper is as follows. In Section 2 we use measurements and a fully data-assimilated model from the ONR Shelfbreak PRIMER Experiment (Linder, Gawarkiewicz & Pickart, 2004) as the test-bench for the new methods we have developed. This includes a description of the ocean model runs that were used for optimization. In Section 3 the Genetic Algorithm optimization for multiple sensor deployment is presented and applied to these data-driven model runs. Section 4 is the conclusion, where plans for the future testing of this algorithm are discussed.

2. MODEL PROBLEM DEFINITION

In this paper, a data-driven model of the ocean in the Middle Atlantic Bight region (shelfbreak south of

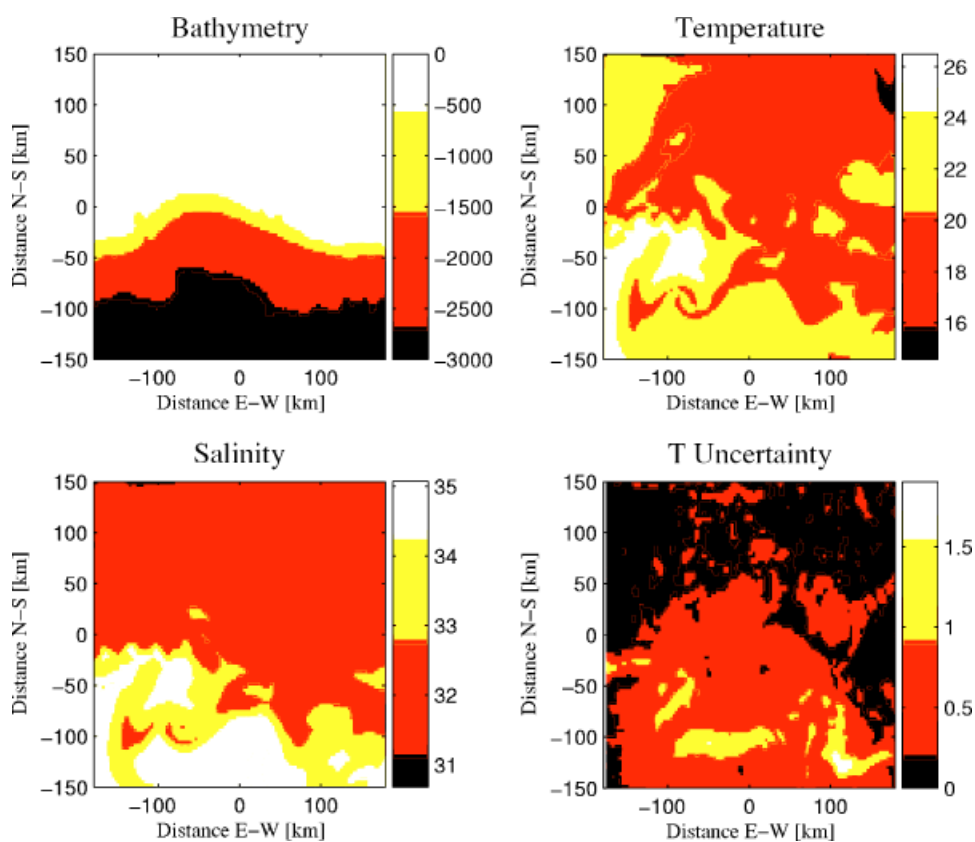


Figure 1. Illustration of the HOPS-ESSE simulations with assimilation of data from the ONR Shelfbreak PRIMER Experiment. Shown are the bathymetry (m) and horizontal snapshots of the surface temperature ($^{\circ}\text{C}$) and salinity (PSU) on day 30 and of the modeled temperature uncertainty of the surface temperature ($^{\circ}\text{C}$) on day 15.

New England) will be used to illustrate the new optimal adaptive sampling approach. This is a region of high oceanographic variability that has been studied by US oceanographers and ocean acousticians for the past 15 years (Apel et al., 1997; Gawarkiewicz et al., 2004). The shelfbreak is located 100 km offshore and interactions of the shelfbreak front, Gulf Stream meanders, and wind-driven response to storms create a complicated environment. The bathymetry for the region is shown in Figure 1(a). The model grid has been rotated to align with the local isobaths.

Using measured environmental data (temperature, salinity, and density) from the Shelfbreak PRIMER experiment from July 2006, a 30 day simulation of the shelfbreak south of New England was generated using the Harvard Ocean Prediction System (HOPS) along with data assimilation via the Error Subspace Statistical Estimation (ESSE) algorithm.

A significant amount of measured data went into this model forecast, including a week of SeaSoar high-resolution sampling, which resolved the fields in terms of both the spatial and temporal correlation scales Gawarkiewicz et al., (2004). The resulting surface temperature and salinity fields (for day 30) are shown in Figure 1 (upper right and lower left panels). The remnants of a slope eddy are visible in the surface temperature and salinity maps (at $y = -50$ km N/S, $x = -100$ km E/W). The surface expression of the shelfbreak front is visible as the high temperature/salinity gradient line at 0 km N/S in the western half of the model domain. In the eastern half, the shelfbreak front extends much further offshore, up to 100 km seaward relative to the western portion of the model domain. To illustrate the HOPS-ESSE estimate of ocean field uncertainties, the standard deviation of the surface temperatures of the en-

semble of model simulations on day 15 is shown in the lower right panel of Figure 1. On the shelf, the surface temperature uncertainty is around 0.2 to 0.4°C. Uncertainties are largest (above 1°C) around the surface expression of the shelfbreak front and its eddies, from about $y=0$ to -100 km, with a band of localized uncertainty near $y=-100$ km N/S, where the shelfbreak front extends far offshore. Such complex time- and depth-dependent ocean fields (temperature, salinity, velocity) and their uncertainties will be used in the estimation of an optimal sampling strategy in the next section.

3. GENETIC ALGORITHM OPTIMIZATION

A numerical experiment was undertaken to demonstrate the use of the nonlinear optimization algorithm. The task is to optimally deploy eight sensors for a period of 1 week in this simulation. For this problem, we optimize the deployment of the following assets: three fixed moorings, three slow moving (30 cm/s) gliders, and two powered REMUS (Remote Environment Monitoring UnitS Autonomous Undersea Vehicles developed by Woods Hole Oceanographic Institution) AUVs. The REMUS vehicles are only deployed for the initial 24 h of operations. The optimization algorithm is tasked with determining the best way to deploy all of these sensors to optimally sample the simulated ocean variability.

The specifications of each platform for this particular numerical example are as follows:

- Moorings—fixed position (x,y) temperature sensor for 7 days
- Gliders—drifting temperature sensor (with a 0.3 m/s velocity in addition to the local currents) deployed for 7 days. Input parameters are x , y , and direction vector.
- REMUS powered AUV vehicles—temperature sensor with a velocity of 4 m/s deployed for 1 day (two 12 h legs with a turn). Search parameters: deployment x , y , and direction vector and turn at 12 h.

This ensemble of platforms leads to a parameter search of 23 independent variables. The sampling constraints for each platform were coded in MATLAB and the cost function was defined, for use within the MATLAB Genetic Algorithm Toolbox.

3.1. Cost Function Constituents

There are currently five constituent cost functions, evaluated for a specific sensor path (x) . The first is the oceanographic variability. The specific function is the negative average standard deviation [$f = -\text{std}(T(x))$] of the measured temperature field for each sensor. Intuitively, the GA optimization (minimization) leads us to sensor sample patterns which sample the most dynamic ocean regions including fronts and eddies. Note that this function, as well as many of the other cost function constituents used, is not positive definite. The second cost function is the oceanographic temperature range (related to the strongest dynamical features). The negative average of the maximum minus the minimum measured temperature for each platform is used [$f = -|\max(T(x)) - \min(T(x))|$]. This function seeks regions where there is a very strong change in the temperature field—an indication of strong fronts or other ocean features. The third constituent cost function is the negative of the integrated uncertainty along each platform path [$f = -\text{sum}(U(x))$]. The ESSE scheme outputs an estimate of the uncertainty for each region and for each field, in space and time. This uncertainty-based function seeks to place sensors at positions where the uncertainty estimate is the largest, helping reduce the largest uncertainties with subsequent measurements. The fourth cost function is the transmission loss (TL) sensitivity. For each platform path, the acoustic field is computed for the first day of the measurement and for the final day of the measurement (using the model-only range-dependent sound speed field). The negative average TL difference is the value of the cost function. This function seeks to measure places where the sound speed field changes in such a way as to impact the ocean acoustic propagation. Regions of high ocean variability, where the acoustic field is not affected, are not important for our use of the ocean model as an input to sonar performance prediction. The final cost function option is a distance potential, where each platform is punished for close proximity (approximately 1.5 to 2 times the baroclinic Rossby radius of deformation in the region, or 20 km in this situation) to other platforms, by the inverse of the range. This function maintains sensor distance to physically intuitive lengths and minimizes possible collisions between sensor platforms.

The global scalar cost function is what the Genetic Algorithm seeks to minimize. It is a function of

the search parameters, in this case the positions and directions of the sensor platforms. In defining the “optimal” deployment configuration of this diverse suite of sensors, it is up to the user to linearly combine the defined cost-function components to build a global scalar cost function for the GA to minimize. This permits the user to weigh ocean dynamics, ocean variability, model uncertainty, TL sensitivity, etc. and produce a platform deployment plan that is consistent with this view of optimal. The cost function E we will use is a weighted combination of the five cost functions described above. The weighted, normalized cost function $E(x)$ is computed as follows:

$$E(x) = \sum_i \frac{W_i f_i(x)}{\sigma(f_i)} \quad (1)$$

where W_i are weights and $\sigma(f_i)$ are normalization functions, each explained in the next paragraph. The underlying principle is that for a specific survey plan, or sensor geometry (x), a variety of constituent functions (f) can be used at the discretion of the user, each evaluated from the modeled values of the field. In real-time, with data coming in, a residual vector can be computed between the measured and the modeled ocean fields can be included. This will reduce the value of measurements where the model is correctly predicting the true ocean, even in dynamic places.

The primary input by the user are the weights W_i . Prior to a genetic algorithm run, a large sampling of the space is conducted and the variance (energy) of each of the cost functions is computed. The cumulative cost function in E is then obtained by summing the weighting value of each cost function ($W_i f_i$) divided by the prior computation of the standard deviation of $\sigma(f_i)$. This normalization takes care of the difference in magnitudes of the various cost functions as well as the units. Each cost function is a dimensionless, normalized scalar, permitting the user defined weighted sum to be the global cost function used for optimization. Many options exist for the constituent functions. The five that we have used, which were introduced above, are next described in more detail.

3.1.1. Oceanographic Variability $\sigma(T)$

The first constituent cost function is the oceanographic variability. It is computed by taking the

negative of the standard deviation of the measured temperatures. The underlying principle here is that we want to measure where the ocean is changing as a function of space and time. Regions with oscillating temperature fields or strong fronts will yield high standard deviations and the GA will prefer to sample these locations. To illustrate the GA search for this function, a simplified optimization with a single REMUS vehicle is computed. The results are shown in Figure 2 (upper-left panel) for day 30 of the ocean model simulation. The REMUS vehicle was placed by the GA such that it transected the remnant of a warm-core eddying feature and then moved into colder water. The standard deviation of the measured temperature for this path was 2.851° , which is large relative to other possible paths in the model domain. This solution agrees well with our intuition for where the most oceanographic variability can be expected.

3.1.2. Oceanographic Dynamic Range (Feature Strength)

The oceanographic dynamics cost function measures the temperature drop across the sensor path. Specifically it is the negative of the $\text{Max}(T) - \text{Min}(T)$. This cost function is based upon the notion that we want to measure where the ocean has the strongest changes in temperature. This cost function works effectively as a front detector. The results for the single REMUS case are shown in Figure 2 with the temperature differential across the path of -8.7° . This path is very similar to the path found using only oceanographic variability, in Section 3.1.1, except that the path is moved slightly north to capture the cold water just north of the eddy.

3.1.3. Oceanographic Uncertainty

One of the outputs of statistical ocean circulation models (from a model ensemble), such as the HOPS-ESSE system, is the forecasted uncertainty for each specific location and variable as a function of time. In order to improve accuracy, measurements should be taken at positions (and times) that reduce these uncertainties the most. This problem has been addressed for example by Yilmaz (Yilmaz, 2005; Yilmaz, Evangelinos, Lermusiaux, & Patrikalakis, 2006) and Lermusiaux (Lermusiaux, 2006). The definition of the cost function is the negative of the integrated uncertainty across the path. This makes the mini-

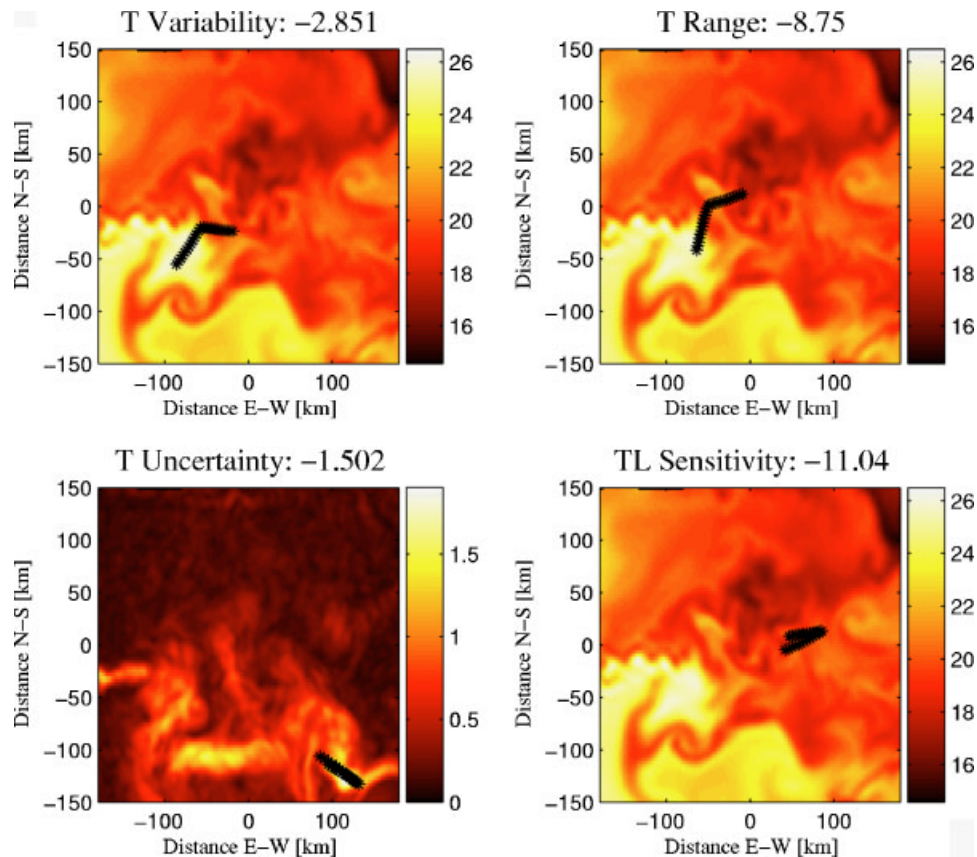


Figure 2. Single REMUS results for ocean variability, ocean dynamic range, ocean uncertainty and transmission loss sensitivity. The color maps are temperature (top two and bottom bottom right, °C) with color axis: [15 26] and temperature uncertainty (bottom left, °C) color axis: [0 2].

mum of the cost function the path in which the highest uncertainty is measured. The results for a single REMUS deployment are shown in Figure 2 (lower left panel). The REMUS sample position is overlaid on the uncertainty map for day 15 of the HOPS-ESSE simulation. Clearly the highest uncertainty at the surface is where the shelfbreak front extends a large distance offshore. The uncertainty cost function complements the previous two cost functions. The solution is to run the REMUS through the peak of the uncertainty map both ways. The REMUS vehicle path doubles back on itself to resample the region of highest temperature uncertainty. This could be eliminated by applying the distance potential constraints to regions where a single sensor has previously gone. Current implementation of the distance potential is between separate sensors.

3.1.4. Temperature Minimum

Following optimal adaptive sensor planning and data assimilation, one of the final products of the oceanographic model is the prediction of sound speed fields for acoustic computations. Acoustic energy refracts away from warm water and therefore propagation is best in regions of a sound channel axis or a sound minimum. An efficient proxy for acoustic propagation, and therefore a candidate for a cost function, is the minimum temperature measured by a particular sensor.

3.1.5. Transmission Loss Sensitivity

The final product of the anti-submarine-warfare (ASW) system that uses oceanographic predictions is a sonar performance prediction. It is the stated goal

of this work that the oceanography should be accurate enough to permit accurate acoustic propagation modeling (Abbot & Dyer, 2002; Heaney & Cox, 2006; Robinson, Abbot, Lermusiaux & Dillman, 2002; Robinson and Lermusiaux, 2004). Ocean acoustic transmissions are commonly characterized by their transmission loss (TL) as a function of range (Jensen, Kuperman, Porter & Schmidt, 1997). TL is the attenuation of an acoustic signal due to geometrical spreading, absorption, and scattering. TL is sensitive to the oceanographic environment (sound speed) as well as the sediment, which along with the surface forms the boundaries of the ocean sound channel. Experimentally, TL is measured by taking the received acoustic intensity (RL) and subtracting (in dB) the source level (SL) within any prescribed frequency band. Thus $TL = RL - SL$. Often in the literature, TL is positive ($SL - RL$). With accurate acoustic predictions in mind, we seek to include acoustic propagation sensitivity in the cost function for the platform deployment plan. Clearly regions where the acoustics are insensitive to the exact nature of the oceanographic variability need not be measured.

To begin to investigate this problem, the transmission loss (TL) was computed using the parabolic equation [Navy Standard RAM (Collins, 1993)] for each platform path (a 40 km path was used for the fixed moorings). The three-dimensional sound speed field and bathymetry (Figure 2) were used to generate two 2-D environments for the acoustic propagation computation. The TLs computed using the sound speed profile (SSP) from day 1 and day 5 were then compared (difference averaged over range and depth), to compute a single measure of the acoustic sensitivity to oceanographic variability. This method should be normalized so that bathymetric differences and geo-acoustic differences between regions are removed from the sensitivity computation. In order to compare transmission loss and not be dominated by the interference of multipath [which Dyer (Dyer, 1970) showed should lead to a standard deviation around 5.6 dB], we performed range smoothing (10% of the range). This is equivalent to a 20 Hz bandwidth average at 200 Hz (Harrison & Harrison, 1995). Range smoothing is significantly more efficient than computing the broadband acoustic field. The cost function metric is the average of the point-by-point difference for regions with TL less than 110 dB. This last threshold is put in to limit the regions of variability to be those where we expect

signal. Note that in the nulls of a TL field, the differences could easily be 10–15 dB, and this result is insignificant to the problem.

The GA result for a single sensor using only the TL cost function is shown in Figure 2 (lower right panel). This is not a location of high surface dynamic oceanography nor of significant surface uncertainty. This is because the TL depends on the vertical variations (not shown) of the temperature, salinity, and pressure, and not so much on the surface values of these properties (fields at depths were not used in the previous GA examples). The value of the cost function for this path is -11.041 dB, which is a significant average TL difference (remember we are doing the equivalent of a 20 Hz band averaging). To illustrate the magnitude of this TL difference, in many environments, the TL drops as a function of range (cylindrical spreading) according to $10 \log(r)$. In this environment, a 10 dB difference in TL is an order of magnitude difference in range of detection of a quiet target (all other variables being fixed). Thus a target would be detectable at 20 km rather than 2 km for a 10 dB error in prediction.

To diagnose what happened to produce an average TL difference of 11 dB, the range- and depth-dependent oceanography along the path of the vehicle was examined. The mixed layer is deeper (30–50 m) at $t=0$, and there is the presence of the shelf-break front for the last 30 km of the path. The shelf-break front is not affecting the net TL difference, however. It is clear that the deepened mixed layer at $t=0$ leads to the situation where energy is refracted towards the bottom and stripped out of the water column. With a shallower mixed layer at $t=5$ days, and the presence of some warm water near the sea-floor at 20–25 km, the sound is able to survive the bottom interactions and propagation over the shelf into deeper water.

This example illustrates the need to incorporate acoustic sensitivity, or at least three-dimensional ocean fields, in any oceanographic measurement and model system, usually referred to as Ocean Observing and Prediction System (Lermusiaux, 2006). If this particular portion of the environment is not well modeled, acoustic predictions can be off by as much as 15 dB.

3.1.6. Distance Potential

In the situation where multiple platforms are deployed it is important to apply a constraint to the

deployment scheme that does not permit the platforms to overlap each other. Without this sort of constraint, the GA will attempt to put all of the sensors in the same location, which minimizes the cost function. This constraint can be applied in two different ways. The first is in the sensor deployment parametrization. A scheme could be devised to move a sensor if another one was already there. This leads to unphysical dependencies between variables in the search space, in addition to being a challenge to program. A simpler approach, the one taken here, is to apply a distance potential and incorporate this into the cost function (Leonard et al. 2006). Simply stated, the distance potential cost function penalizes any particular individual (a specific deployment scheme of multiple platforms) if two or more of the platforms come within a certain range of each other. If the platforms stay beyond the specified potential distance, there is no penalty. The potential distance was chosen to be 1.5 times the baroclinic Rossby radius. The ocean is expected to be coherent within this range and, if another sensor has measured this area, there is no value in adding a second one. The equation for the distance potential cost function is:

$$f(r_{ij}) = \frac{r_o}{r_{ij}} + 1 \quad (2)$$

where r_o is the distance potential width and was taken to be 20 km. The distance between each sensor pair for all times is r_{ij} . The distance potential therefore asymptotically approaches 1 for large r_{ij} , and grows exponentially when $r_{ij} < r_o$. By adding the distance potential to the cost function, the GA solution naturally chooses solutions that have platforms remaining beyond the distance potential width r_o .

3.2. Multiple-Platform Combined Cost Function Run

Prior to running the multi-platform, multi-parameter cost function, we perform a simple examination of the topography of the cost function. In general the cost function is a function of multiple parameters (launch position, launch direction, turns) and multiple vehicles. To examine the form of the cost function, we will search over the initial position of a single platform, which samples a 20 km path oriented due North. The cost functions associated with the temperature dynamics, the minimum tem-

perature, and the integrated uncertainty, as well as the weighted sum of the normalized cost functions, are shown in Figure 3. With uniform weighting, the topology of each cost function is visible in the combined cost function. In particular, the shelf break front minima in the STD T panel around ($y = 0$ km, $x = -150$ km) and the strong T-Uncertainty minima ($y = -120$ km, $x = 110$ km) are visible.

With the cost function constituents now defined, a run of the GA for a full suite of eight sensor platforms and a linear combination of all five cost functions is conducted. The sensor suite includes three moorings, simultaneously deployed with three gliders for 5 days and two REMUS vehicles (for 24 h only). Each of the first five cost functions ocean variability, ocean dynamics, ocean uncertainty, TL sensitivity, and sensor range potential, was given an equal weighting ($W_i = 1$ for all i). The sensitivity of the solution to this weighting can be evaluated through multiple GA runs, or can be estimated from Figure 3. The ocean variability (and ocean dynamic range, not shown) cost function is deepest near the remnant of the gulf stream eddy; the ocean T uncertainty is focused on the shelf-break front and the temperature minima is generally spread north-east of the front.

The full GA solution is shown in Figure 4. The proof that this is the optimal sensor deployment will require a significant amount of Monte Carlo modeling and ocean data assimilation and is left for future work. The results are qualitatively appealing, however. Note that all three moorings have been placed directly on the shelf-break front (which passes over them during the 5 days of deployment). The REMUS vehicles both pass over manifestations of the break and the gliders track large gradients with time. The distance potential also appears to be performing its task of maintaining sensor placement beyond a Rossby radius (15 km in this example). The temporal dependence of the ocean field strongly influences the location of the moorings (which are fixed in space) and the gliders (which are slowly drifting and are driven by currents). The REMUS vehicles, on the other hand, are quickly moving and therefore are insensitive to time-dependent oceanographic effects.

Note that every sensor begins its survey in cold water and ends the survey in warm water, highlighting the effect of the min/max value in the cost function as well as the standard deviation. The exception to this is platform 4 (a 24 h REMUS deployment), which does the reverse.

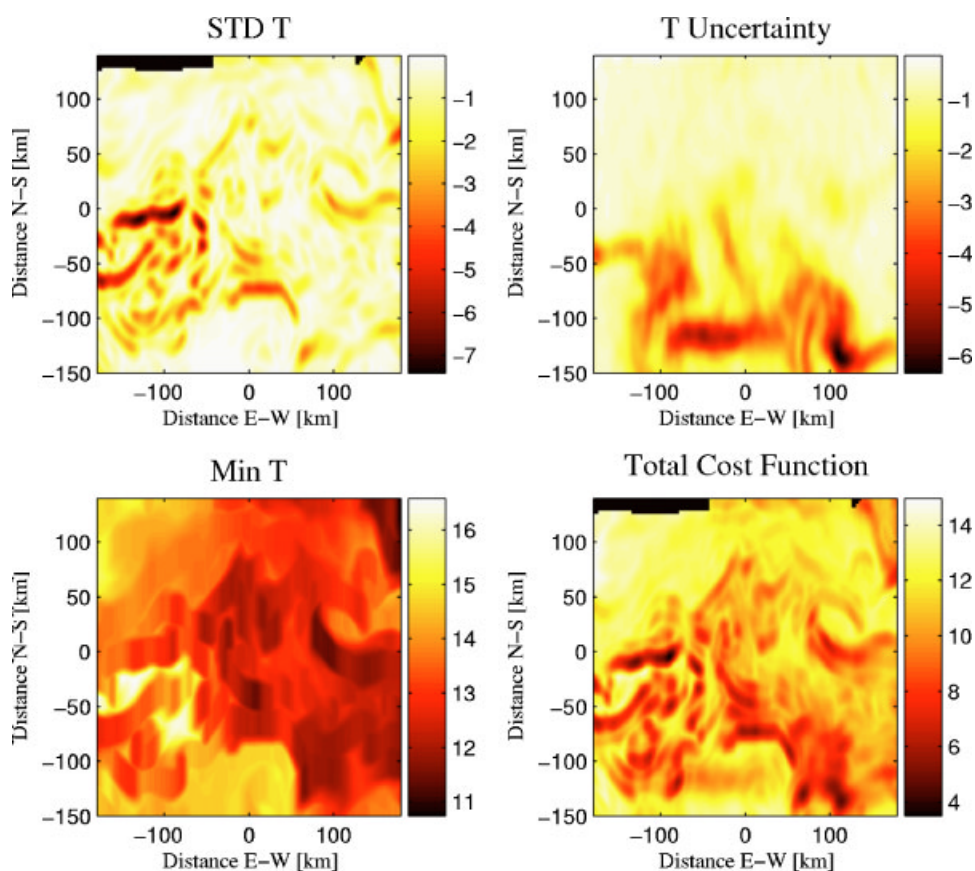


Figure 3. Cost function topology for the standard deviation of the measured temperature on day 30, the integrated uncertainty on day 15, and the minimum measured temperature on day 30. The normalized, combined cost-function is plotted in the lower right panel. The only independent variables in the computation are the single platform starting position.

4. CONCLUSION

By combining model forecasts, nonlinear optimal sensor deployment and data assimilation, a coupled approach to optimally estimating the ocean environment is presented. The technique involves a nonlinear constrained optimization utilizing the Genetic Algorithm. The technique was applied to a simulated example involving the placement of eight measurement platforms in a dynamic region. Five different constituent cost functions were developed and applied separately to this problem. The functions are ocean variability, ocean dynamic range, ocean uncertainty, sensor distance potential, and transmission loss sensitivity. A major advantage of this technique rests in the ability to add multiple normalized scalar

cost functions, providing the user with the ability to weight the final cost function based upon his own specific needs.

One of the significant questions to be answered in model-based optimization is how “optimal” is the solution. Without an extensive set of measurements this is difficult to verify. Note that this is still true if a model is not formally utilized in the planning: one cannot quantitatively determine if chosen measurement sites are ideal without additional (large) data sets. Another question relates to the optimality of the solution computed by the GA algorithm. Even though the algorithm is global and designed to avoid local minima, its final solution is not guaranteed to be the optimum of the chosen cost function. In addition, the final generation of the GA depends on the rules

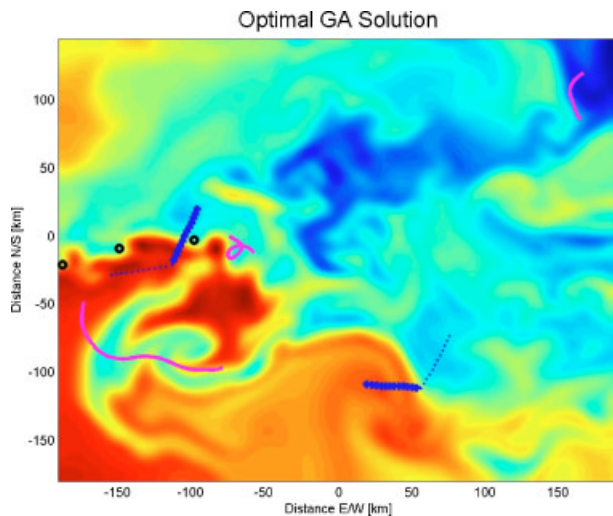


Figure 4. Genetic Algorithm solution for three moorings, three gliders, and two REMUS vehicles. Note how the moorings (o) are located near the shelf-break front, the REMUS vehicles and the gliders cross over major fronts. The blue tracks are REMUS vehicles and the magenta tracks are gliders.

chosen to determine how new generations are formed during the optimization process. Nonetheless, there are counter-arguments to these two questions. First, the use of computational models to predict oceanic and acoustic states is quantitative and is now beginning to provide more skill than guessing at what the future ocean and acoustic conditions will be. Similarly, the sampling strategies provided by our GA approach so far are intuitively pleasing, combine all of the features we wish to optimize, are consistent with the deployment constraints and are obtained automatically and efficiently by a computer.

The examples provided here to illustrate the approach all used computationally simulated rather than experimentally derived fields. To further evaluate the capabilities and validity of the approach, real at-sea exercises should be carried out. Such exercises would sample the ocean using GA-based path planning and the GA would define its cost function based on the real ocean measurements and the data-driven real-time model predictions of ocean and acoustic fields and uncertainties. Pretest surveys would be required to provide initial conditions for the model. With the ocean model up and running, a flexible sampling architecture would then be necessary. This would hopefully include various surface ships for de-

ploying fixed moorings and deploying and recovering AUVs. Glider deployment can be done at the start of the test and the gliders can be guided via satellite communications. Redeploying the gliders with surface ships is a significant plus, but for large operational areas this can be very difficult. Acoustic propagation measurements, both as data for assimilation and/or post-test verification, would also be part of this demonstration test. A comprehensive set of oceanographic measurement (moorings/air XBTs) could be used to validate the results, after the test. The suite of measurements will provide the opportunity, after the test, to perform a numerical study to determine the value added to the ocean prediction for each measurement.

The example field measurement optimization that has been described here was designed to sample variability and to sample regions where the model is poorly constrained. This is only one of many possible approaches to survey design. Because the ocean supports many physical and biogeochemical processes, optimization of field measurements is not uniquely defined because various scientific objectives would each be consistent with different sets of optimization criteria. Thus, a suitable oceanographic question must be posed before an optimal survey can be computed.

Note that only features having scales resolved by the model can enter into the optimization. Consistency between data and model is important, and can determine the role of unresolved features (such as nonlinear internal waves). This information can then be used in continuing data collection, filtering, and assimilation.

In the absence of a reliable model, the situation may arise where the importance of a specific dynamical effect or feature is to be determined, sometimes called the classification process. An example is determining the scale length of dominant features. In these cases, the statistical functions used in the example optimization do not provide the necessary information, and other cost functions would need to be employed. Estimating spatial scales using data from fixed and moving platforms required careful analysis and is a problem in its own right. Methods have been developed for such estimation (Zhang, Bellingham & Baggeroer, 2001), and for determining the measurement capabilities of moving platforms for certain types of processes (scales) (Wilcox et al., 2001). The capabilities of platforms as defined in prior work can be coupled with knowledge of the resolved energy-containing

scales (perhaps through an iterative or feedback method) to constrain the optimization process. For example, the distance potential constituent function may be adjusted based on this, or an AUV/glider survey generation algorithm (i.e., lawnmower pattern generator) could be tuned appropriately. In this paper we have presented a methodology for automated optimal sensor planning for data-assimilation in dynamical ocean modeling. The solution involves performing a Genetic Algorithm optimization simultaneously on a deployment of multiple sensors, with the ability of using multiple platform types. The versatility of the approach is the ability to use multiple user defined scalar cost-functions with user defined weights. Thus the user has the control to define what is meant by “optimal” and the algorithm search for a deployment scheme which optimizes the user defined global cost function.

ACKNOWLEDGMENTS

We appreciate the useful comments from the anonymous reviewers. P.F.J.L. thanks P. J. Haley and W. G. Leslie for discussions. This work was funded under a Small Business Innovative research (SBIR) contract through the Office of Naval Research, Space and Naval Warfare Center (SPAWAR-San Diego), contract grant numbers N00039-06-c-0063 (Heaney and Duda), N00014-05-1-0413 (Gawarkiewicz), and N00014-05-1-0335 and N00014-05-1-0370 (Lermusiaux).

REFERENCES

- Abbot, P., & Dyer, I. (2002). Sonar performance predictions incorporating environmental variability. Paper presented at the Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance, Lerici, Italy.
- Apel, J.R., Badiey, M., Chiu, C.-S., Finette, S., Headrick, R., Kemp, J., et al. (1997). An overview of the 1995 swarm shallow-water internal wave acoustic scattering experiment. *IEEE Journal of Ocean Engineering*, 22(3), 465–480.
- Bellingham, J.G., & Wilcox, J.S. (1996). Optimizing auv oceanographic surveys. Paper presented at the AUV '96.
- Collins, M. (1993). A split-step pade solution for the parabolic equation method. *Journal of the Acoustical Society of America*, 93(4), 1736–1742.
- Curtin, T.B., Bellingham, J.G., Catipovic, J., & Webb, D. (1993). Autonomous oceanographic sampling networks. *Oceanography*, 6(3), pp. 86–94.
- Dyer, I. (1970). Statistics of sound propagation in the sea. *Journal of the Acoustical Society of America* 48(1), 337–345.
- Gawarkiewicz, G.G., Brink, K.H., Bahr, F., Beardsley, R.C., Caruso, M., Lynch, J.F., et al. (2004). A large-amplitude meander of the shelfbreak front during summer south of new england: Observations from the shelfbreak primer experiment. *J. Geophysical Research*, 109, 3006.1–3006.16.
- Goldberg, D.E. (1989). *Genetic algorithms in search, optimization and machine learning*. Boston: Addison-Wesley.
- Haley, P.J., Lermusiaux, P.F.J., Leslie, W.G., & Robinson, A.R. (2006). The Harvard ocean prediction system (hops).
- Harrison, C.H., & Harrison, J.A. (1995). A simple relationship between frequency and range averages for broadband sonar. *Journal of the Acoustical Society of America*, 97(2), 1314–1317.
- Heaney, K.D., & Cox, H. (2006). A tactical approach to environmental uncertainty and sensitivity. *IEEE Journal of Oceanographic Engineering*, 31(02), 356–367.
- Jensen, F.B., Kuperman, W.A., Porter, M.B., & Schmidt, H. (1997). *Computational ocean acoustics*. New York: AIP Press.
- Leonard, N.E., Paley, D., Lekien, F., Sepulchre, R., Fratantoni, D.M., & Davis, R.E. (2006). Collective motion, sensor networks and ocean sampling. *IEEE Special Issue on Network Control Systems*, in press.
- Lermusiaux, P.F.J. (2004). Application of the error subspace statistical estimation (esse) system to real-time error forecasting, data assimilation and adaptive sampling off the central california coast during aosn-ii. Paper presented at the 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Seattle, WA.
- Lermusiaux, P.F.J. (2006). Adaptive sampling, adaptive data assimilation and adaptive modeling. In C.K.R.T. Jones & K. Ide (Eds.). Special issue of the *J. of Inverse Problems* on “Mathematical issues and challenges in data assimilation for geophysical systems: Interdisciplinary perspectives.”
- Lermusiaux, P.F.J., Chiu, C.-S., Gawarkiewicz, G.G., Abbot, P., Robinson, A.R., Miller, R.N., et al. (2006). Quantifying uncertainties in ocean predictions. In T. Paluskiewicz & S. Harper (Eds.). *Oceanography, Special issue on “Advances in computational oceanography,”* 19, 92–105.
- Linder, C.A., Gawarkiewicz, G.G., & Pickart, R.S. (2004). Seasonal characteristics of bottom boundary layer detachment at the shelfbreak front in the middle atlantic bight. *Journal of Geophysical Research*, 109(C), 3049–3058.
- Ogren, P., Fiorelli, E., & Leonard, N.E. (2004). Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment. *IEEE Journal of Transactions on Automatic Control*, submitted.
- Paley, D., Zhang, F., & Leonard, N.E. (2006). Cooperative control for ocean sampling: The glider coordinated

- control system. *IEEE Transactions on Control Systems Technology*, in press.
- Patrikalakis, N.M., McCarthy, J.J., Robinson, A.R., Schmidt, H., Evangelinos, C., Haley, P.J., et al. (2004). Towards a dynamic data driven system for rapid adaptive interdisciplinary ocean forecasting. In F. Darema (Ed.), *Dynamic data driven application systems*, Netherlands: Kluwer Academic Publishers.
- Robinson, A.R., Abbot, P., Lermusiaux, P.F.J., & Dillman, L. (2002). Transfer of uncertainty through physical-acoustical-sonar end-to-end systems: A conceptual basis. Paper presented at the *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, Lerici, Italy.
- Robinson, A.R., & Lermusiaux, P.F.J. (2004). Prediction systems with data assimilation for coupled ocean sciences and ocean acoustics. In A. Tolstoy (Ed.). *Theoretical and computational acoustics* (pp. 325–342).
- Robinson, A.R., Lermusiaux, P.F.J., & Sloan, N.Q. (1998). Data assimilation. In K.H. Brink & A.R. Robinson (Eds.), *The sea* (Vol. 10, pp. 541–593). New York: Wiley.
- Wilcox, J.S., Bellingham, J.G., Zhang, Y., & Baggeroer, A.B. (2001). Performance metrics of oceanographic surveys with autonomous underwater vehicles. *IEEE Journal of Oceanographic Engineering*, 26(4), 711–725.
- Yilmaz, N. (2005). Path planning of autonomous underwater vehicles for adaptive sampling. Unpublished Ph.D. thesis, Massachusetts Institute of Technology, Cambridge.
- Yilmaz, N., Evangelinos, C., Lermusiaux, P.F.J., & Patrikalakis, N.M. (2006). Milp path planning of auvs for adaptive sampling. *IEEE Journal of Oceanographic Engineering*, under review.
- Zhang, Y., Baggeroer, A.B., & Bellingham, J.G. (2001). Spectral-feature classification of oceanographic processes using an autonomous underwater vehicle. *IEEE Journal of Oceanographic Engineering*, 26(4), 726–741.