

# An Evaluation of Sampling Path Strategies for an Autonomous Underwater Vehicle\*

Andres Mora, Colin Ho, and Srikanth Saripalli

**Abstract**—A critical problem in planning sampling paths for autonomous underwater vehicles is balancing obtaining an accurate scalar field estimation against efficiently utilizing the stored energy capacity of the sampling vehicle. Adaptive sampling approaches can only provide solutions when real-time and a priori environmental data is available. Through utilizing a cost-evaluation function to experimentally evaluate various sampling path strategies for a wide range of scalar fields and sampling densities, it is found that a systematic spiral sampling path strategy is optimal for high-variance scalar fields for all sampling densities and low-variance scalar fields when sampling is sparse. The random spiral sampling path strategy is found to be optimal for low-variance scalar fields when sampling is dense.

## I. INTRODUCTION

Autonomous underwater vehicles (AUV) are mobile robotic platforms that are utilized for environmental sampling, sensing and are capable of operating in a multitude of underwater aquatic environments [3]. In research they are commonly used to carry hydrological, geophysical, and/or biological sensor payloads that are used to gather remote and in-situ data to aid the study of open and closed aquatic bodies [10], [4].

An autonomous underwater vehicle is generally described as being capable of operation without any human controller; thus it must be able to traverse the aquatic environment on a desired path through autonomous navigation and control [6]. An example of an autonomous underwater vehicle is the one used in this paper to collect the data that was then processed to generate scalar fields (see Figure 1).

The goal of environmental sampling and sensing is to generate an accurate estimate of the underlying scalar field(s) over an area of interest. An estimate is generated by interpolating the sensing data collected over the area being studied. In the case of aquatic environments, a mobile vehicle can be used to collect the necessary sensing data. However the sensing platform is limited in range by the amount of stored energy it carries aboard. The problem of environmental sensing then becomes determining how one obtains an accurate estimation of the scalar field of interest while being constrained by the amount of stored energy the vehicle can utilize in order to sample the scalar field.

The purpose of this paper is to experimentally evaluate sampling strategies based upon estimation accuracy and en-



Fig. 1. Modified OceanServer IVER2 AUV at Lake Pleasant.

ergy consumption. This evaluation is conducted with a cost-evaluation function that considers multiple parameters, and can assign priorities to each factor. The evaluation takes into account real world and simulated isotropic, anisotropic scalar fields, as well as varying sampling densities to determine which sampling strategy is optimal for different scalar field types, for a range of sampling densities. The sampling path strategies evaluated in this study were systematic and stratified random sampling distributions with spiral and lawn mower sampling paths.

Through utilizing the cost-evaluation function, it is found that the systematic spiral sampling path strategy best optimizes the energy consumption with the scalar reconstruction error compared to the other sampling path strategies evaluated. Additionally, it is found that the systematic spiral path consumes the least amount of energy out of the sampling path strategies evaluated.

## II. RELATED WORK

In the fields of robotics, hydrology, geology, and geostatistical sciences, optimal sample collection and path planning are an active area of research [7]. There is much prior and current work ongoing in the domain of path planning and sampling optimization for autonomous vehicles. For example, adaptive sampling algorithms have been developed that can direct the path of single or multiple autonomous underwater vehicles in towards locations of high probable data yield [8], and can be used in conjunction with existing sensor networks [12]. Additionally, energy optimal paths can be computed based upon known and sensed external variables such as ocean currents [11]-[9], and static or dynamic obstacles [1].

However all of these planning schemes require a priori knowledge of the environment or real time sensing and data feedback. There are many AUV deployments scenarios

\*This work was supported in part by NSF grant XYZ

A. Mora, C. Ho, and S. Saripalli are with the School of Earth and Space Exploration, Arizona State University, 550 E. Tyler Mall, Tempe, AZ, USA, 85287 {amoravar, srikanth.saripalli, colinho} at asu.edu

where there is a lack of a priori data on the environment, and the AUV is equipped with sensors that do not provide real time data (such as taking physical water samples) [10]. Therefore, in many real world deployments, the paths chosen are often simple lawnmower patterns that may or may not be optimal for the situation they are being utilized [3], [10]. The evaluation conducted in this paper is targeted at looking at autonomous underwater vehicle path planning from an experimental field scientists data sampling perspective. The goal is to comparatively evaluate various sampling path strategies using a cost-evaluation function that can optimize multiple parameters. These results from this evaluation will aid choosing the best sampling path strategy for unknown scalar fields, with no real time data feedback.

### III. EXPERIMENTAL APPROACH

The evaluation consisted of conducting AUV sampling paths over various scalar fields to generate estimations of those scalar fields, and the energy consumption of the sampling path taken. Each of the simulated sampling paths and their corresponding scalar field estimation and energy consumption are then used as inputs for a cost-evaluation function that is used to comparatively determine the optimal sampling path.

#### A. Assumptions

The following assumptions are made:

- 1) The simulated AUV travels at a constant velocity and is capable of navigating to the desired sampling locations.
- 2) The vehicles total energy consumption is based upon the total distance it travels, and the total angle it turns.
- 3) The underlying scalar field being sampled has little or no temporal variation while being sampled.
- 4) There is no real-time access to the sampled data; the samples can only be accessed offline.

#### B. Sampling Strategies

The two main sampling strategy types that are evaluated are systematic sampling, and stratified random sampling. Systematic sampling is defined as sampling from an area of interest at regularly spaced intervals. Prior research has shown that having equilateral sampling grids is only optimal when the scalar field variation is isotropic [7]. Therefore stratified random sampling was chosen as an additional sampling method to be comparatively evaluated. Stratified random sampling is conducted by splitting the desired sampling area into grid of equal sized sub-areas, in which samples are chosen from a random location in each sub-area. Stratified random sampling often produces a weighted mean with less variability than a standard random sampling. For each type of sampling strategy, lawn mower and spiral path patterns are evaluated. Both are simple path patterns which are commonly used for surveying. Example paths of the four sampling strategies are shown in Figure 2.

### C. Underlying Scalar Field Data

The underlying scalar field was represented with both real world and simulated data. The goal was to utilize enough scalar field data to be representative of many scalar field distributions commonly seen in real world data sets. Thus, six different underlying scalar fields were generated and collected for the purposes of this evaluation and are shown in Fig. 3. The data sets derived from real world data consisted of a turbidity data set representing multi-modal data, a chlorophyll data set representing high anisotropy and variance, and a blue green algae cell count data set representing moderate anisotropy data with a high degree of spatial variance. The first simulated data set was a linearly varying distribution representing isotropic data. The second simulated data set was a normal distribution that represented moderately anisotropic data. The third and final simulated data set was a bi-modal normal distribution that represented moderate-variance anisotropic data.

### IV. EVALUATION METHODS

#### A. Estimation Evaluation

Once a sampling path has been generated, it is then used to sample the underlying scalar field. With those samples, an estimation of the underlying scalar field is generated through ordinary Kriging. Kriging is a form of linear least squares estimation that interpolates the value of a location upon a scalar field from known values at nearby locations, which in this case are the sampled locations [2]. Ordinary Kriging is a specific type of Kriging that assumes that the experimental variogram can be constructed, and that the mean of the scalar field being estimated is unknown but constant.  $Z$  represents the underlying scalar field, thus the Kriging estimation of the scalar field at any location is the weighted average of observed values. This is represented through the following equation (1).

$$z(x_0) = \sum_{i=1}^n w_i z(x_i) \quad (1)$$

The estimation variance of  $z(x_0)$  is given by

$$\sigma_k^2 = 2 \sum_{i=1}^n w_i \gamma(x_i, x_0) - \sum_{j=1}^n w_i w_j \gamma(x_i, x_j) \quad (2)$$

where  $\gamma(x_i, x_0)$  is the variogram of  $x_i$  and  $x_0$ . The weights  $w_i(x)$  are chosen such that they fulfill the unbiasedness condition (they all sum to 1), and also minimize the estimation variance. Thus, the weights are determined through the equation (3)

$$\sum_{j=1}^n w_j \gamma(x_i, x_j) + \psi = \gamma(x_i, x_0) \quad (3)$$

where  $\gamma$  is the Lagrange parameter associated with the minimization. Once the Kriging estimation was generated, its accuracy is evaluated through the computation of its integrated mean square error (IMSE) with respect to the underlying scalar field.

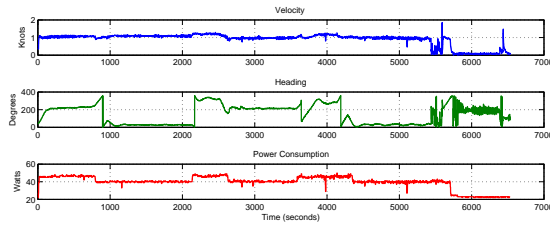


Fig. 2. Velocity, heading and power consumption used to estimate the vehicle's dynamical model.

### B. Energy Consumption Evaluation

In order to improve the simple energy consumption model used in previous studies [5], an estimation of the dynamical model of the robot based on the selection of two known inputs and one output from the data acquired was used. The two inputs of the estimation were the velocity and the heading of the vehicle measured by the on-board GPS. The output selected for the estimation of our model was the total power consumed by the vehicle during the sampling mission. The estimation consisted of first calculating the state-space model using a sub-space method and later on refining the model by computing the prediction error estimate of the state-space model.

With this estimated model, it is not required to assume that the vehicle is travelling straight at a constant velocity, thus there is no restriction with respect to the rate at which the vehicle consumes its stored energy. An increase in the energy consumed by the vehicle while it turns may be nominal or significant, depending on the design of the vehicle. The estimated model incorporates the possible effect in the increment in energy consumption a turning maneuver may represent by including the changes in the orientation of the vehicle. The velocity of the vehicle (in knots), its heading (in degrees) and the corresponding power consumption (in Watts) measured by the instruments on-board are presented in Figure 2.

In order to estimate and validate the model, the estimation is based on half of the entire data set and it is later validated using the remaining half of the data set acquired by the vehicle's instruments (GPS, XXYY, etc.).

The resulting dynamic model obtain from the estimation follows the standard state-space model convention as presented by equation 4:

$$\begin{cases} x(t+Ts) = Ax(t) + Bu(t) + Ke(t) \\ y(t) = Cx(t) + Du(t) + e(t) \end{cases} \quad (4)$$

where  $A = \begin{bmatrix} 0.99495 & -0.069517 \\ 0.021859 & 1.0813e-5 \end{bmatrix}$ ,  $B = \begin{bmatrix} -0.0032466 & 0.86892 \end{bmatrix}$ ,  $C = \begin{bmatrix} 126.37 & 6.4396 \end{bmatrix}$ ,  $D = \begin{bmatrix} 0 & 0 \end{bmatrix}$ ,  $K = \begin{bmatrix} 0.0070135 \\ 0.0090162 \end{bmatrix}$ , and  $x(0) = \begin{bmatrix} -0.14306 \\ -0.36083 \end{bmatrix}$ . The data acquisition frequency was set to  $2Hz$ .

The resulting power based on the estimated dynamical model is compared against the half of the data set used to validate the model as shown in Figure 3.

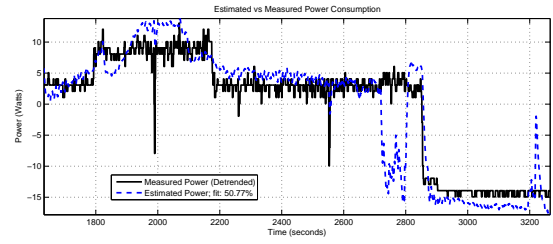


Fig. 3. Estimated vs Measured Power Consumption.

### C. Cost-Evaluation Function

The cost-evaluation function is a general evaluation method that allows for quantitative comparative analysis for multiple input parameters. It normalizes and then assigns each input parameter a weight, allowing for the function to prioritize each parameter in respect to one another. These weights must all sum to one. In the case of this evaluation, we are investigating which sampling strategy is optimal for both estimation error and energy consumption. Therefore the parameters that were selected to be inputs for the cost-evaluation function were the total distance travelled  $d$ , the total energy consumed  $E$ , and the IMSE for the specific sampling path taken  $\epsilon$ . Our proposed cost-evaluation function is as follows,

$$E(p) = \sum_{i=p} W_d N_d d_i + W_E N_E E_i + W_\epsilon N_\epsilon \epsilon_i \quad (5)$$

where  $W_d$ ,  $W_E$ , and  $W_\epsilon$ , are the weighted factors that provide specific priorities between the total distance travelled, the total energy consumed, and the IMSE for the sampling path taken.  $N_d$ ,  $N_E$ , and  $N_\epsilon$  are normalization factors assigned constant values to normalize each of their corresponding parameters and eliminate their dimensions.

## V. EXPERIMENTAL RESULTS AND DISCUSSION

### A. Underlying Scalar Field Data Generation

The six different scalar fields used in this evaluation were either generated from simulated equations, or from a real world dataset. The isotropic scalar field used for this evaluation was generated by the following equation.

$$I(x, y) = \frac{x + y}{0.1} + 20 \quad (6)$$

The first simulated anisotropic scalar field was generated by using a standard normal distribution and then scaling it by a factor of 50 and offsetting by +20. The multi-modal anisotropic field was generated by simply adding together two equally scaled inverted normal distributions with differing  $x - y$  locations for their means. The real world dataset used for generating the turbidity, chlorophyll and blue green algae scalar fields was created by sampling a local freshwater lake with an autonomous underwater vehicle carrying a water quality monitoring sonde. The contiguous underlying scalar field was then generated by interpolating the samples using ordinary kriging.

The AUV to used to collect the samples was a heavily modified OceanServer IVER2 platform, which carries a YSI 6600V2 water quality monitoring sonde. The sonde sampled the in-situ turbidity, chlorophyll, and blue green algae, among a variety of other measurements.

All of the samples were taken at Lake Pleasant, a large freshwater reservoir in Arizona. The approximate coordinates of the sampling location are 33°51'55.66 N and 112°17'45.01 W. The sampling locations are shown in Fig. 3. The AUV collected 2258 samples over an area of approximately 70[m] x 100[m]. A subset of the samples from 30-70[m] northing and 30-70[m] easting was selected, since it was the region containing the highest density of samples.

### B. Estimation Error Computation

To compute the estimation error, an estimate of the underlying scalar field was first generated for each of the sampling strategy types. These estimates were generated across varying sampling grid sizes in order to characterize a range of sampling densities from low to high. The sampling grid dimensions and total sample size are listed in Table I.

The IMSE for each of the stratified random sampling strategy based estimations was the average of fifteen iterations, in order to generate a stable value. The systematic sampling estimation was iterated five times at five separate evenly distributed path seeding locations to remove biasing in error results. These seeding locations are shown in Fig. 4.

### C. Estimation Error Result

The IMSE error for each of the scalar varying from low (20 samples), moderate (56 samples) and high (120 samples) sampling density are displayed alongside one another in the Table II on the following page.

Through looking at the IMSE versus number of sample plots for the various underlying scalar fields, a number of correlations are apparent. Firstly, both strategies demonstrate a reduction in IMSE as the number of samples increases, as one would expect. Additionally, it is found that the systematic sampling strategy minimizes error for the real world data sets across all sampling densities. The stratified random sampling strategy minimizes error for the low variance isotropic and anisotropic 2 data sets. Finally, for the anisotropic 1 data set, the systematic sampling strategy significantly minimizes error for low sampling densities, and the stratified random sampling strategy minimizes error for moderate and high sampling densities.

### D. Energy Consumption Computation

The energy consumption of each of the sampling path patterns was computed using (4). First, the distance and total angle turned was computed for each of the sampling path types through the range of sampling densities shown in Table I. For each of the stratified random sampling paths, the total distance travelled and total cumulative angle turned was computed as the averages of 50 iterations of each sample grid size. Finally (4) was applied using  $k_\theta$  values from a range of 0 to 2. This range characterizes vehicles for which a 90° turn

correlates to consuming the energy equivalent to travelling an additional distance of 0[m] to 2[m].

### E. Energy Consumption Results

The average total energy consumption versus the  $k_\theta$  for each of the sampling path strategies is shown in Fig. 9. The resulting total energy versus  $k_\theta$  plots were then fitted to linear equations to quantify their relationship. The resulting energy consumption model equations for each of the sampling path strategies are shown below in Table III.

It is shown in the energy consumption model equations, that the even spiral sampling path strategy is the most energy efficient, and through viewing Figure 9, it is easily apparent.

### F. Cost-Evaluation Computation

The cost-evaluation function was computed with the all the weighted factors set equal to 1/3, thus weighting all of the evaluation parameters equally. To compute the energy for each sampling path,  $k_\theta$  was set to 0.1. The function was applied to each of the sampling strategies for all the underlying scalar fields, at 20, 56, and 120 samples. This was done such that a comparative analysis of each of the sampling strategies at low, moderate and high sampling densities could be conducted.

### G. Cost-Evaluation Results

The results of the cost-evaluation are listed in Table IV. For a low sampling density the systematic spiral sampling method is optimal for all of the data sets except for the isotropic scalar field. For a moderate sampling density the systematic spiral strategy is optimal for the anisotropic 1, turbidity, and chlorophyll scalar fields. The random lawnmower and random spiral strategies are optimal for the anisotropic 2 and isotropic scalar fields respectively. Finally, for a high sample density the random spiral strategy is optimal for the isotropic and both of the anisotropic scalar fields. The systematic spiral strategy was found to be optimal for the turbidity, chlorophyll and blue green algae scalar fields. The strategy that was optimal for the most types of scalar fields and densities was found to be the systematic spiral strategy. In general, a spiral path strategy was found to be the optimal in almost all of the cases.

### H. Discussion of Results

The results from this cost-evaluation are specific to the weights given to the cost-evaluation function. However these results do provide insight into which sampling strategies are optimal for different types of scalar fields, given the assumption that distance travelled, total energy consumption and IMSE are considered equal factors in optimality.

For isotropic scalar fields, the random spiral sampling strategy was found to be the optimal sampling strategy for all sampling densities.

It was found that in low variance scalar fields, such as the single/multi-modal anisotropic scalar fields used in this evaluation, the systematic spiral strategy is optimal for coarse sampling densities, and the random spiral strategy is optimal for high sampling densities.

In high variance scalar fields such as the turbidity, chlorophyll and blue green algae data sets, the systematic spiral strategy was optimal for almost every case.

## VI. CONCLUSIONS

We experimentally evaluated four different sampling strategies through a cost-evaluation function. Assuming that optimality was defined as minimizing the total path distance, the total energy consumed, and the IMSE, it was found that the systematic spiral sampling strategy was the most optimal strategy overall. Thus if one were given the task to choose a sampling strategy given no a priori knowledge of the underlying scalar field and no access to the sampled data in real time, the best sampling strategy to choose would be a systematic spiral path, rather than a standard systematic lawnmower path. It was also demonstrated that the cost-evaluation function is a useful tool that can be used to quantify overall optimality for multiple sampling strategy parameters.

In the future we plan on further experimentally validating the results of this evaluation. Additionally, the cost-evaluation function can be utilized to create an online path optimization planner that chooses path options in real time.

## REFERENCES

- [1] C. V. Caldwell, D. D. Dunlap, and E. G. Collins. Motion planning for an autonomous underwater vehicle via sampling based model predictive control. In *OCEANS 2010*, pages 1–6. OCEANS, September 2010.
- [2] J. P. Delhomme. Kriging in the hydrosiences. In *Advances in Water Resources*, volume 1, pages 251–266, September 1978.
- [3] A. L. Forrest, H. Bohm, B. Laval, E. Magnusson, R. Yeo, and M. J. Doble. Investigation of under-ice thermal structure: small auv deployment in pavilion lake, bc, canada. In *OCEANS 2007*, volume 1, pages 1–9, 2007.
- [4] C. R. German, D. R. Yoerger, M. Jakuba, T. M. Shank, C. H. Langmuir, and K. ichi Nakamura. Hydrothermal exploration with the autonomous benthic explorer. *Deep Sea Research Part I: Oceanographic Research Papers*, 55(2):203 – 219, 2008.
- [5] C. Ho and S. Saripalli. Where do you sample? - an autonomous underwater vehicle story. In *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*, pages 119 –124, September 2011.
- [6] M. Kumagai, T. Ura, Y. Kuroda, and R. Walker. New auv designed for lake environment monitoring. In *Underwater Technology, 2000. UT 00. Proceedings of the 2000 International Symposium on*, pages 78–83, 2000.
- [7] A. B. McBratney, R. Webster, and T. M. Burgess. The design of optimal sampling schemes for local estimation and mapping of regionalized variables i: Theory and method. In *Computers and Geosciences*, volume 7, pages 331–334, June 1981.
- [8] D. O. Popa, A. C. Sanderson, R. J. Komerska, S. S. Mupparapu, D. R. Blidberg, and S. G. Chappel. Adaptive sampling algorithms for multiple autonomous underwater vehicles. In *Autonomous Underwater Vehicles, 2004 IEEE/OES*, pages 108– 118. IEEE/OES, June 2004.
- [9] R. N. Smith, A. Pereira, Y. Chao, P. P. Li, D. A. Caron, B. H. Jones, and G. S. Sukhatme. Autonomous underwater vehicle trajectory design coupled with predictive ocean models: A case study. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 4770–4777. IEEE, May 2010.
- [10] C. R. . Stoker, D. Barch, J. Farmer, M. Flagg, T. Healy, T. Tengdin, H. Thomas, K. Schwer, and D. Stakes. Exploration of mono lake with an rovs: a prototype experiment for the maps auv program. In *Autonomous Underwater Vehicle Technology, Proceedings of the 1996 Symposium on*, pages 33–40, June 1996.
- [11] J. Witt and M. Dunbabin. Go with the flow: Optimal auv path planning in coastal environments. In *Australian Conference on Robotics and Automation, 2008*, 2008.
- [12] B. Zhang and G. S. Sukhatme. Adaptive sampling for estimating a scalar field using a robotic boat and a sensor network. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3673–3680. IEEE, April 2007.