Where Do You Sample? - An Autonomous Underwater Vehicle Story

Colin Ho and Srikanth Saripalli

***Abstract*— We present an experimental evaluation of various sampling path strategies for an Autonomous Underwater Vehicle. Both systematic and stratified random sampling path strategies were evaluated based upon their estimation accuracy for isotropic and anisotropic scalar fields, as well as the relative energy consumption. We present results from several experimental trials that shows that the stratified random sampling strategy minimizes estimation error for denser sample distributions, and the systematic sampling strategy minimizes estimation error for sparser sample distributions. Finally, we experimentally show that the systematic spiral path sampling strategy is the most energy efficient.**

Keywords- Sampling, AUV, kriging

I. INTRODUCTION

Autonomous underwater vehicles are robotic platforms that are capable of traversing in underwater aquatic environments, typically with the purpose of environmental sampling and/or sensing. In research they are commonly used to carry hydrological, geophysical, or biological sensors which are used to aid in the study aquatic bodies of all sorts.

An autonomous underwater vehicle must be able to operate without any human controller; thus it must be able to navigate and control itself in order to traverse the aquatic environment in a controlled and desirable path. Autonomous underwater vehicles (AUV) are flexible research platforms that can be used for a variety of exploration purposes, such as environmental sensing. For example, the vehicle used in this paper to collect the samples that were used to generate a scalar field data set is shown in figure 1. The goal of environmental sensing is to obtain an accurate estimate of underlying scalar fields of interest. The estimate generated is based upon sensing data collected throughout the area of interest. However, the platform being used to collect the sensing data is limited by the amount of stored energy that is carried aboard the vehicle. The sensing platform can only travel and sense as far as its stored energy capacity will allow it to. The question then becomes how one obtains the most accurate estimation of the scalar field while being limited by the stored energy constraints of the vehicle being used to sample the scalar field.

The goal of this paper is to experimentally evaluate various sampling strategies based upon their relative estimation accuracy and energy consumption. The evaluation considers both modeled and actual isotropic and anisotropic scalar fields to see if sampling strategy optimality is dependent upon the scalar field being considered. Additionally, the various sampling strategies are analyzed to see if they vary in their relative optimality based upon the sampling density.

Fig. 1. Modified OceanServer IVER2 AUV

The sampling path strategies evaluated were systematic sampling and random lawnmower and spiral sample paths.

Through the results of the estimation error evaluation, we hypothesize that the systematic sampling strategies provide better estimation errors for both isotropic and anisotropic scalar fields for sparse sampling densities, and the random stratified sampling strategies provide better estimation errors for dense sampling densities.

Furthermore the energy consumption evaluation found that the systematic spiral path was the most energy efficient sampling strategy out of those compared. However, the stratified random spiral path is more energy efficient if the energy required to turn the vehicle is significantly higher than the nominal energy required to propel the vehicle forward.

II. RELATED WORK

Optimal sample path planning and optimal sampling has been an active research field in robotics, as well as the hydrological, geological and geostatistical sciences [1].

For autonomous underwater vehicles, much research exists on developing various path planning schemes that optimize for various parameters, such as the estimation of the underlying scalar field based upon a priori knowledge through existing sensor networks [2], or the energy efficiency of the path based upon ocean currents and dynamic obstacles [3]. These optimizations primarily focus on the theoretical aspects of path planning optimization. This evaluation conducted in this paper

is targeted at looking at autonomous underwater vehicle path planning from an experimental field scientist's data sampling perspective. The goal is to experimentally evaluate various sampling path strategies based on how well each strategy estimates scalar fields at various sampling densities, and how much relative energy consumption each sampling strategy takes. These results will hopefully aid in choosing which sampling strategy would work best for sampling an unknown scalar field.

Though we were not able to find a conclusive trend between the sampling strategies and estimation error, we were able to find that the most energy efficient sampling strategy in our comparison is the systematic spiral path strategy.

III. Experimental Approach

The experiment consisted of simulating an AUV sampling a lake with various sampling strategies to generate estimations of the underlying scalar field. The various sampling strategies were then evaluated based upon the accuracy of the estimation generated, as well as their relative energy consumption.

A. Assumptions

We make the following assumptions

- The simulated vehicle travels at a constant velocity and that it has a waypoint controller that is capable of navigating it to the precise desired sampling locations.
- The vehicle's power consumption is dependent on the distance it travels, as well as the total angle it turns.
- The underlying scalar field that is being sampled is assumed to have little/no temporal variation while being sampled.

B. Sampling Strategies

The two main types of sampling strategies that are evaluated are systematic sampling, and stratified random sampling. Systematic sampling is a standard form of sampling in which the desired sampling area is sampled at regular intervals. However equilateral sampling grids are only optimal where variation is isotropic [1]. 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 experimental and simulated data. The goal was to use scalar field data that was reflective of all types of real world data. The experimental data used was a blue green algae cell count data set which represented highly anisotropic data with a high

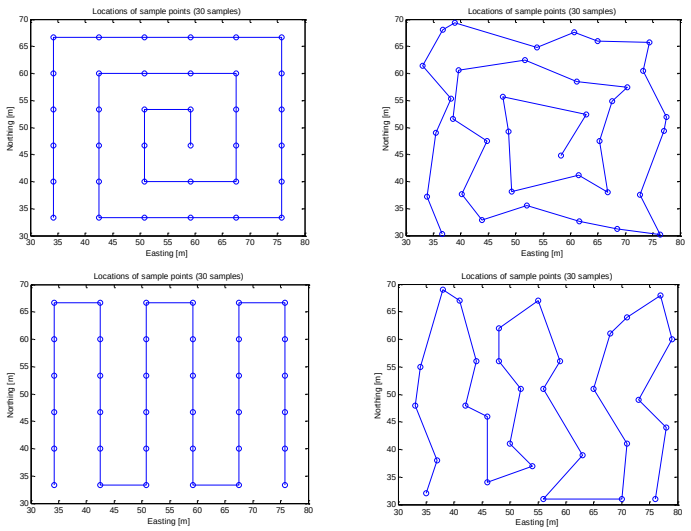

Fig. 2. The four sampling path strategies evaluated

degree of spatial variance. The first simulated data set was a normal distribution that represented low-variance anisotropic data. The other simulated data set was a linearly varying distribution representing isotropic data.

D. Estimation Evaluation

Once the underlying scalar field was sampled, an estimation of the scalar field was generated through the use of ordinary Kriging. Kriging is a geostatistical technique which utilizes linear least square estimation to interpolate the value of a random field at a location from known values of the random field at nearby locations [4]. Ordinary kriging is a form of kriging which operates on the assumption that the mean of the field is unknown and constant, and that the experimental variogram of the field can be estimated. The kriging estimate of z, the underlying scalar field, is a weighted average of the observed values $z(x_1)$, $z(x_2)\ldots z(x_n)$ in it, thus the estimation of the scalar field at an location x_o is

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

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

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

Where $\gamma(x_i, x_0)$ is the variogram between x_i and x_0. The weights $w_i(x)$ are chosen such that they sum to 1 to fulfill the unbiasedness condition, and also minimize the estimation variance. This is done though the equation

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

Where ψ is the Lagrance parameter associated with the minimization.

The accuracy of the kriging generated estimations was evaluated through computing the integrated mean square error (IMSE), and the average percent absolute error.

E. Energy Consumption Evaluation

The energy consumption of each sampling path was evaluated with a simple total energy consumption model. This model is dependent upon two parameters, the distance travelled and the total angle turned. The vehicle is assumed to have constant power consumption while travelling at a constant velocity, thus the total energy consumed can be determined by the distance travelled. However when the vehicle turns while travelling, it increases power consumption due to increased drag. This increase may be nominal or significant depending of the type of vehicle. Thus the total energy consumed is modeled with the following equation.

$$E(d,\theta) = d + k_\theta \theta \qquad (4)$$

In the equation, d represents the total distance travelled and θ represents the total angle turned. The constant k_θ is a weight that varies depending on how much travelling while turning increases power consumption. The weight k_θ can be correlated to the relation,

$$k_\theta = \frac{d_t}{90^o} \qquad (5)$$

This is interpreted as, for every 90 degree turn made, an amount of energy consumed equal to travelling a straight distance d_t is added.

IV. Experimental Results and Discussion

A. Underlying Scalar Field Data Generation

The simulated isotropic scalar field generated for use in the experiment was created by the equation

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

The simulated anisotropic scalar field was generated as a standard normal distribution which was scaled larger by a factor of 50 and then positively offset by 20. The experimental scalar field data set was generated through the sampling of a freshwater lake with an autonomous underwater vehicle with a water quality monitoring sonde, and then interpolating the samples using ordinary kriging to attain an estimation of the full scalar field.

The AUV used was a heavily modified OceanServer IVER2, which towed a YSI 6600V2 water quality monitoring sonde that measured blue green algae with a resolution of 1cell/mL among a variety of other measurements.

The samples were taken at approximately 33° 51' 55.66" N and 112° 17'45.01" W in Lake Pleasant, which is a freshwater reservoir in Arizona. The sampling path taken by the vehicle can be seen in figure 3.

The samples that could not be correlated to GPS locations due to the GPS antenna aboard the AUV dipping underwater were discarded.

A total of 3448 samples were taken over a 70m X 100m area which is plotted in figure 4.

Fig. 3. Plot of the AUV sampling path taken laid over satellite imagery from Google Earth

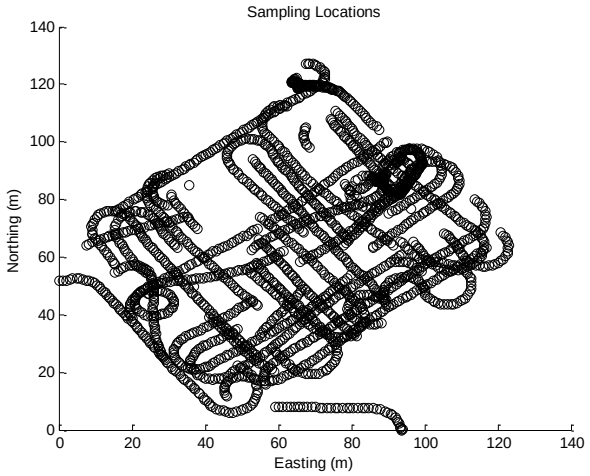

Fig. 4. Plot of the sampling locations taken

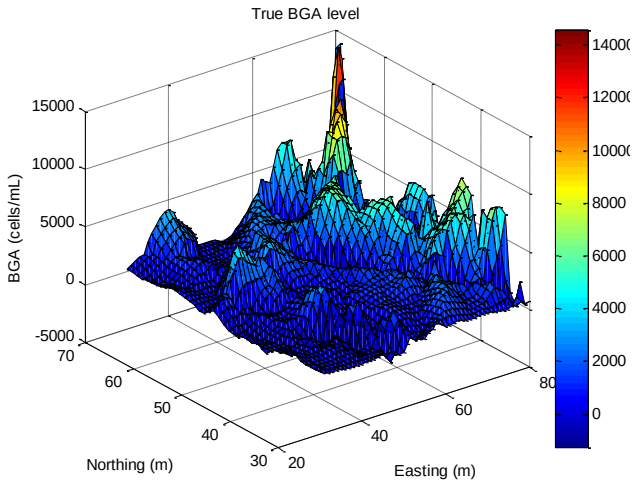

Fig. 5. Blue Green Algae data set sub-area chosen

The area from 20-80m easting and 30-70m in northing was selected as the data set to be used, since it had the highest sampling density. The resulting data is shown in figure 5.

B. Estimation Error Computation

For both of the sampling strategy types, distributions of varying grid sizes were chosen to characterize sampling grids that were both coarse and fine. These are listed in the table I below

TABLE I. SAMPLE DISTRIBUTION GRID SIZES

Grid Dimensions		**Sample count**
Size in X	***Size in Y***	
5	4	20
6	5	30
7	6	42
8	7	56
9	8	72
10	9	90
12	10	120

In order to generate the kriging estimations for each of the sampling distributions, experimental semivariograms were calculated and then modeled based upon the scalar field type being estimated. Linear and spherical variogram models were selected for the isotropic and anisotropic scalar fields respectively, since they provided the closest fit for the experimental variograms.

The stratified random sampling estimation was iterated 15 times to generate a stable average of the estimation error for each of the sampling grid dimensions.

C. Estimation Error Result

The estimation error for each of the underlying scalar fields are plotted in figure 6. Figures 6.a,b,c show the IMSE error versus the number of samples taken for each of the underlying scalar fields tested, and figures 6.d,e,f show the average absolute percent error versus the number of samples taken for each of the underlying scalar fields tested.

Looking at the various error versus number of sample plots, the systematic sampling IMSE and average percent error plots do not exhibit stable behavior, thus making it difficult to extract a useful comparative analysis. The jumps and oscillations in error can be explained by how the systematic sampling estimation was only conducted at one location for each set of sample grid sizes, while the stratified random sampling error values were taken as the average of 15 error estimation iterations. Keeping this in mind, we are still able to see some trends in the data that can be tentatively characterized. Firstly, as the number of samples increases, both strategies see a reduction in IMSE and average percent error for the isotropic and anisotropic simulated scalar fields. However this trend is not strictly followed for the BGA scalar field. When looking at the IMSE, it seems that the systematic sampling strategy maintains a lower degree of error while coarsely sampling all three scalar fields. This trend is reversed when the number of samples is increased past 100 samples. However this correlation is not terribly strong, and will require the additional simulation of higher numbers of sample points to see if this trend continues. One might wonder as to why the IMSE and average percentage error for the BGA scalar field data set was irregularly high. This is explained by the extremely high spatial variance of the data set, as well as the high range of values (0 to 14,000 cells/mL) contained in the data set. The simulated isotropic and anisotropic data sets were created to roughly model the range and distribution of the surface temperature over a body of water. Thus the range of values did not exceed from about 20-30 units, and the spatial variance was relatively low compared to the BGA data set. Therefore the errors computed for the simulated data sets are on average significantly lower than that of the BGA data set.

D. Energy Consumption Computation

To compare the energy consumption of each of the path sampling patterns, the distance and total angle turned was computed for a range of coarse and fine sample distribution grid sizes from 5 by 5 to 15 by 15. This resulted in a total range of 25 to 225 samples taken. Sampling paths of each of the four types were generated based upon the sample distribution grid sizes, and the distance and total angle turned were computed for each sampling path. For the stratified random lawnmower and spiral sampling paths, the average distance and total angle turned were computed by averaging 50 sampling paths of each sample distribution grid size.

E. Energy Consumption Results

As seen in figure 7, the distance versus number of sample plots show that the systematic (even) spiral sampling path pattern provides the shortest distance travelled, which is then followed by the systematic lawnmower path, the stratified random spiral path, and finally the stratified random lawnmower path. Over the sample range calculated, the systematic spiral path was on average 4.53% shorter than the systematic lawnmower path, 7.4% shorter than the stratified random spiral path, and 11.65% shorter than the stratified random lawnmower path.

As seen in figure 8, the total angle turned versus number of samples plot shows that the systematic sampling path methods provided the lowest total turn angle. On average the total turn angle for systematic path patterns was17.68% less than the total turn angle of the stratified random path patterns.

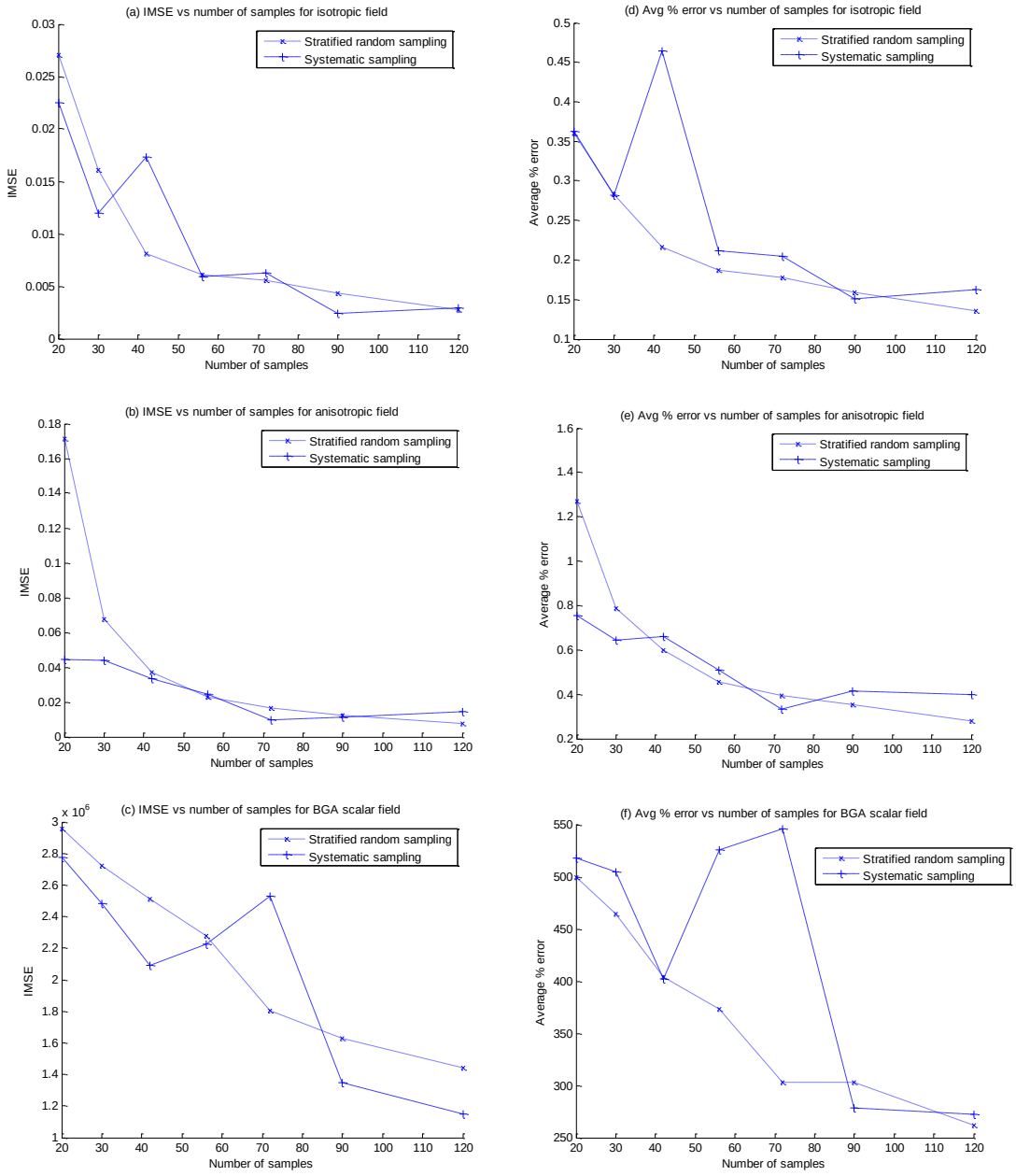

Fig. 6. Estimation errors

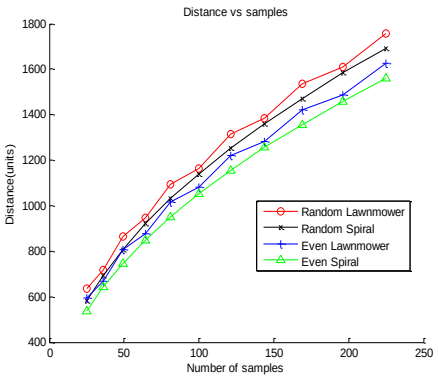

Fig. 7. Total distance vs number of samples

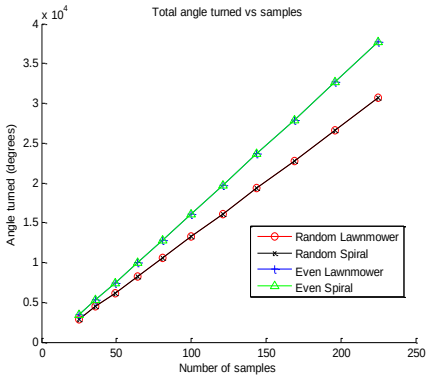

Fig. 8. Total angle turned vs number of samples

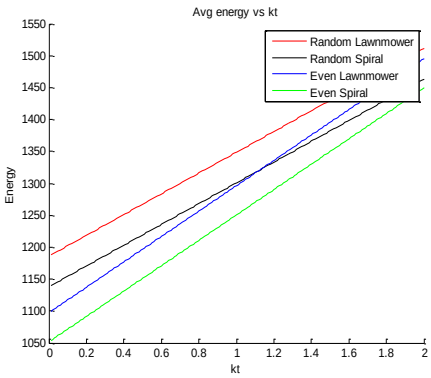

Fig. 9. Average energy consumption vs k_θ

Finally, we evaluated the energy consumption for each of the paths while the constant k_θ was varied from 0.01 to 2 in the energy consumption equation. This range characterizes vehicles for which a 90^o turn correlates to consuming the energy equivalent to travelling an additional distance of 0.01 to 2 units. This characterizes AUVs that are hydrodynamically efficient while turning (such as highly non-holonomic torpedo shaped AUVs), to AUVs that are hydrodynamically very inefficient while turning (such as holonomic multi-thruster box-shaped AUVs). The average energy consumption E(d, θ) is computed with equation 4, with d (the total distance travelled) equaling the average of all the distances, and θ (the total angle in degrees turned) equaling the average of all the total angles turned for all the sample sizes that were taken.

The total average energy consumption for the various sampling path strategies are computed for the varying k_θ values, and then plotted in figure 9. The total energy consumption for the various sampling path strategies are fitted to linear equations, and the resulting energy consumption equations are shown below in table II.

TABLE II. ENERGY CONSUMPTION MODEL EQUATIONS

Sampling path	**Model equation parameters**	
	Slope (energy/ kθ)	***intercept***
Random Lawnmower	162.6	1186
Random Spiral	162.6	1138.1
Even Lawnmower	198.73	1097
Even Spiral	198.73	1051.8

Figure 9 shows that the systematic spiral based path sampling strategy is the most energy efficient in comparison to all the other path sampling strategies over the given range of k_θ values.

F. Discussion of Results

The results from the estimation error analysis show some interesting trends, the data is not conclusive enough to make a full judgment upon these trend analyses. We will further verify these trends with additional simulations with increased sample sizes, as well as increased sampling iterations to obtain better average values. In general as the sample size increases, the estimation error reduces for the simulated anisotropic and isotropic data sets. Furthermore based on the plots we hypothesize that the stratified random sampling distribution results in a more accurate estimate for larger sample sizes (greater than 100 samples) for isotropic and anisotropic data types. Additionally, the error plots show that the for smaller sample sizes, systematic sampling distributions result in a more accurate estimate for the simulated isotropic and anisotropic scalar fields. The error plots for the BGA scalar field data also show some interesting results, such as a fluctuation in systematic sampling estimation error as the sample size increases. Since the systematic sampling estimation was only conducted in one location for each sample size configuration, we will need to obtain more data before analyzing any trends for the real world data.

The total energy consumption for the systematic spiral based path sampling strategy is on average 3.55% less than that of the systematic lawnmower sampling strategy, 3.98% less than that of the stratified random spiral sampling strategy, and 7.40% less than that of the stratified random lawnmower sampling strategy. Interestingly, when the k_θ reaches a value of about 1.1, the stratified random spiral sampling strategy consumes less energy than the systematic lawnmower path sampling strategy. The stratified random spiral sampling strategy has the lowest average energy consumption to k_θ ratio, thus if a vehicle is extremely energy inefficient at turning, a stratified random spiral sampling strategy would be the most efficient strategy.

V. CONCLUSION

We experimentally evaluated the estimation error and energy efficiency of four different sampling strategies. Through the results that were generated, we hypothesize that there is a trend between the sample size and estimation error for systematic and random stratified sampling. However this trend needs to be further explored through additional simulation with an increased number and range of sample sizes, and a higher number of estimation averages. Through the energy consumption evaluation, we postulate that the systematic spiral path sampling strategy is the most energy efficient, consuming 3.55% less energy than a standard systematic lawnmower sampling strategy.

In the future we plan on evaluating additional sampling strategies, and adaptive sampling path planning methods. We also plan on experimentally verifying these path sampling strategies through deploying our own autonomous underwater vehicle on each of these sampling paths strategies in a nearby lake.

ACKNOWLEDGMENT

The authors would like to acknowledge the following individuals for their support and help: Alexander Kafka and Brance Hudzietz.

REFERENCES

[1] A.B. McBratney, R. Webster, T.M. Burgess, “The design of optimal sampling schemes for local estimation and mapping of of regionalized variables--I: Theory and method,” Computers & Geosciences, Volume 7, Issue 4, 1981, Pages 331-334.

[2] B. Zhang, G.S. Sukhatme, "Adaptive Sampling for Estimating a Scalar Field using a Robotic Boat and a Sensor Network," *Robotics and Automation, 2007 IEEE International Conference on* , vol., no., pp.3673-3680, 10-14 April 2007

[3] J. Witt, M. Dunbabin, “Go with the Flow: Optimal AUV Path Planning in Coastal Environments,” Australian Conference on Robotics and Automation, 2008.

[4] J. P. Delhomme, Kriging in the hydrosciences, Advances in Water Resources, Volume 1, Issue 5, September 1978, Pages 251-266.