Mobile robot sensing for environmental applications

Amarjeet Singh[1], Maxim A. Batalin[1], Michael Stealey[1], Victor Chen[1], Mark H Hansen[2], Thomas C. Harmon[3], Gaurav S. Sukhatme[4], and William J Kaiser[1]

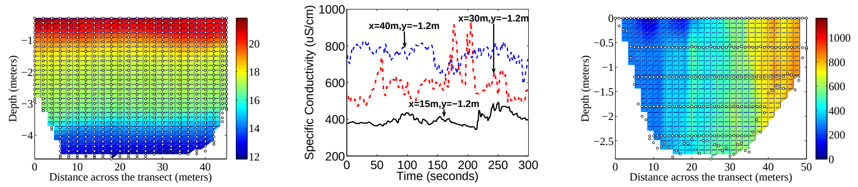
[1] Department of Electrical Engineering, University of California, Los Angeles
[2] Department of Statistics, University of California, Los Angeles
[3] School of Engineering, University of California, Merced
[4] Computer Science Department, University of Southern California

Summary. This paper reports the first application of iterative experimental design methodology for high spatiotemporal resolution characterization of river and lake aquatic systems performed using mobile robot sensing systems. Both applications involve dynamic phenomena spread over large spatial domain: 1) Characterization of contaminant concentration and flow at the confluence of two major rivers displaying dynamics due to flow of the water; and 2) Characterization of rapidly evolving biological processes such as phytoplankton dynamics in a lake system. We describe the development and application of a new general purpose method for mobile robot sensing in such environments - Iterative experiment Design for Environmental Applications (IDEA). IDEA introduces in-field adaptation of mobile robotic sensing system. Analysis of the complex spatial and temporal structures associated with each observed environment is presented. Detailed characterization of the observed environment using IDEA methodology is used as an informed prior to improve the performance of the existing adaptive experimental design approaches for mobile robotic systems - stratified adaptive sampling and hierarchical non-stationary Gaussian Processes.

1 Introduction

A broad class of environmental sensing applications, both terrestrial and aquatic, require observing an environment that displays significant heterogeneity in both space and time. As an example, river observations are useful for answering the questions pertaining to the hydraulics and multi-dimensional river modeling [1], geomorphology, sediment transport and riparian habitat restoration [2]. These require high granularity measurements of coupled velocity and water quality parameters in river cross-sections. Flow of water in the rivers result in temporal heterogeneity while mixing of the two streams at a confluence zone results in spatial heterogeneity. Fig. 1 displays the temperature distribution observed in a lake and the contaminant distribution in terms of specific conductivity (in space and time) at a confluence of two rivers, displaying the high degree of heterogeneity in such environments.

Traditionally, aquatic environments were observed sparsely in both space and time because of the difficulties and the associated cost of continuously accessing these environments [3]. Over the last few years, with the development of various robotic and sensor network platforms, several prototype systems have been used for



(a) Spatial variation of temperature (^oC) at Lake Fulmor (b) Temporal variation of specific conductivity at SJ (c) Spatial variation of specific conductivity ($\mu S/cm$) at SJ

Fig. 1: Spatio-temporal heterogeneity in environmental phenomena (Points represent observation locations)

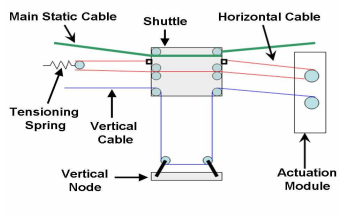
sensing the aquatic environments at different scales. Buoyed or moored platforms now exist which can provide vertical profiling capabilities with high temporal resolution over long time periods at key locations [4]. Mobile robotic platforms such as boats [4], Autonomous Underwater Vehicles [5, 6] and cabled robotic systems [7], each suited for a specific environment, have resulted in high resolution sampling in both space and time. However, literature survey reveals that this is the first time an actuated sensing system has been used for detailed characterization of applications such as phytoplankton growth in lake environment and coupled flow and water quality parameters in river environment.

In this paper, we present IDEA - Iterative experiment Design for Environmental Applications; an approach for characterizing an unknown environment under the constraints of rapidly evolving phenomena. This involves an in-field adaptation in the experiment design: to capture phenomena dynamics exploiting observations from prior models, iteratively executed experiments and the behavior of the underlying control processes (if known). We discuss how such an approach can optimize the limited campaign time and acquire data that characterizes the spatiotemporal distribution of the observed phenomena with high fidelity.

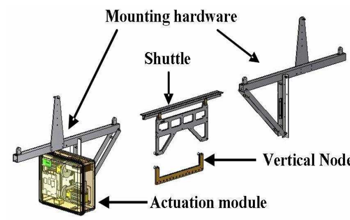
To illustrate IDEA and its validation in real world applications, we present case studies of two environmental sensing campaigns where such an approach was successfully followed and validated. The first campaign was executed at the confluence of two rivers, Merced river and San Joaquin river, in California from August 21-25, 2006 (hereafter referred to as SJ deployment). The second campaign was executed at a lake located in a sub alpine (1,600 m) coniferous forest within the San Jacinto mountains of Southern California [8] from August 28-31, 2006 (hereafter referred to as Fulmor deployment). Information collected using IDEA methodology is used to demonstrate the improvement in performance of two of the existing adaptive experiment design approaches: 1) Stratified adaptive sampling [9]; and 2) Hierarchical non-stationary Gaussian Process based modeling approach [10, 11].

2 System Description

An actuated mobile robotic system, Rapidly Deployable Networked Info Mechanical System [7] (hereafter called NIMS), was used during each of the two campaigns to navigate a sensor payload anywhere within the two dimensional cross-sections of the aquatic environments. Fig. 2a shows the schematic diagram of the system with



(a) Schematic view



(b) Major system components

Fig. 2: NIMS system with basic cable setup

the basic cable setup. It consists of three main cables: the static cable supporting the shuttle and the vertical node platform, a horizontal and a vertical cable used to control the corresponding motion of the sensor payload. Fig. 2b displays major system components that include the mounting hardware, actuation module and a shuttle. The static cable is used to support the mounting hardware which in turn is used to attach the actuation module at one end of the transect and horizontal and vertical cables at the other end.

The static cable is specified to support a maximum tension of 3700 pounds while the maximum tension on our system was close to 750 pounds. Tensioning spring at one end keeps the horizontal cable loop tight and prevents it from slipping on the spool on the other side. The vertical cable was tensioned manually. This proved to be sufficient, since the vertical cable was also kept under tension by the sensor payload. The shuttle and the vertical node platform are actuated through the horizontal and the vertical cables using two motors located inside the actuation module. Both of these motors are controlled simultaneously using a serial interface.

Several parameters of the phenomena were measured, using commercially available sensors [12–14], to validate and advance the understanding of interrelated physical, chemical and biological processes in the observed environments. These include temperature, pH, specific conductivity, Photosynthetically Active Radiation (PAR), depth, oxidation reduction potential, turbidity, Luminescent Dissolved Oxygen (LDO), nitrate, fluorometer and flow velocity among others. Additionally, a physical sampling device, designed and developed at UCLA specifically for these campaigns, was used to collect water samples for detailed lab analysis and measurement of other required parameters for which no commercial sensors are available. The device uses dual spring loaded syringes to collect water samples when actuated at any location within the two dimensional cross-section of the aquatic system.

This system architecture is an improvement from first generation system introduced in [15]. It allows for reduced node mass, increased horizontal and vertical node speed, constant and convenient access to the actuation control module and rapid yet flexible deployments. Use of cable based system enabled precise localization and ability to actuate heavy sensor payload. System design rationale and implementation details are provided in [7].

Software architecture included system for monitoring the real time data stream to provide immediate fault detection as well as to monitor the variability in the observed phenomena distribution in real time. Additionally, algorithms for adaptive ex-

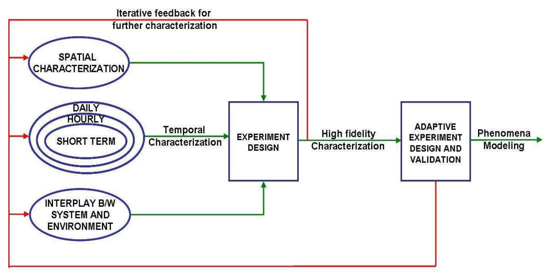


Fig. 3: Schematic view of IDEA methodology

perimental design, adjusting to the observed phenomena distribution, were integrated in the field with the existing architecture to provide autonomous system operation.

3 IDEA (Iterative experimental Design for Environmental Applications) Methodology

IDEA provides a methodology for in-field adaptation of experimental design to perform detailed characterization of the spatiotemporal distribution of the observed environment. The two successful campaigns discussed in this paper display the utility of IDEA methodology for actuating the mobile robotic system to perform such characterization. Fig. 3 displays the schematic view of IDEA methodology. Following is the detailed discussion of each module of IDEA methodology with an illustration of their general applicability to different sensing systems.

3.1 Characterizing the spatial variability

Dense spatial sampling is required to characterize the spatial distribution with high fidelity. In the case of statically distributed sensing system, this requires a dense, uniform distribution of sensor network. On the other hand, with an actuated sensing system such as NIMS, this requires performing deterministic, dense raster scans discounting the temporal variation in the phenomena distribution. In both cases, the initial spatial density can be decided based on the known characteristics of the observed phenomena. This can then be iteratively improved till it is sufficient to completely characterize the spatial variability in the observed phenomena with high fidelity. For each of our aquatic sensing applications, dense, deterministic scans (raster scans) were performed by actuating NIMS to navigate the sensor payload in the two-dimensional cross-section of the observed environment. The density of these raster scans was decided in consultation with the environmental scientists and aquatic biologists and was influenced by prior experiments in the same environment. This was found to be sufficient for high fidelity characterization based on the observed phenomena distribution.

3.2 Characterizing the temporal variability

Temporal variation in the phenomena distribution should be characterized at different scales. When using static sensor network, sampling frequency should be guided by observed temporal variation. These static sensors will provide both the short term and long term temporal variations. Though such measurements will be restricted to only a few locations in the observed environment (high resource cost makes it impractical to deploy highly dense static sensor network). In case of actuated systems, diurnal variation should be characterized by repeating the dense raster scans at different times of the day. Such scans will also provide periodic sampling checks for post deployment calibration, data integrity etc. and a baseline to characterize the performance of any adaptive sampling approach. At each of our two campaigns we performed repeated raster scans at different times of the day (and nights at Lake Fulmor), using NIMS, that enabled us to understand the mixing effects across different layers of water column.

Dwelling time at each sensing location, while performing raster scans, should be selected so as to satisfy both the requirements for observing short term temporal variations across the complete observed environment and the response time of the sensing system. For the San Joaquin deployment, dwelling time during each of the raster scans was selected to be 30 seconds to account for high variability due to constant flow in the river. On the other hand, for the physically static Lake Fulmor environment, the dwelling time for each of the raster scans was selected to be 10 seconds. Observed standard deviation at each of the sensing locations confirmed that these dwelling times for our campaign were sufficient to satisfy both the demands for sampling short term temporal distribution and the response time of the sensors.

3.3 Interplay between sensing system and observed environment

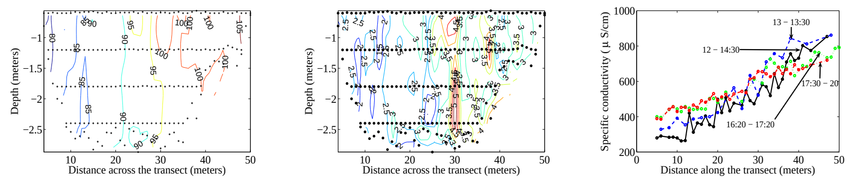
In addition to characterizing the observed environment, experiments must be designed to characterize the constraints of the sampling system such as localization drift, interference in the observed environment and others, such that the integrity of the collected data can be established. For example, during Lake Fulmor campaign we designed experiments to characterize the mixing of water across different layers of water column due to the motion of NIMS inside the water. On the other hand, during San Joaquin deployment with significant flow of the river, we designed experiments to characterize the static deflection of the suspended, pendulous sensor system while making observations inside water.

3.4 Adaptive experimental design and validation

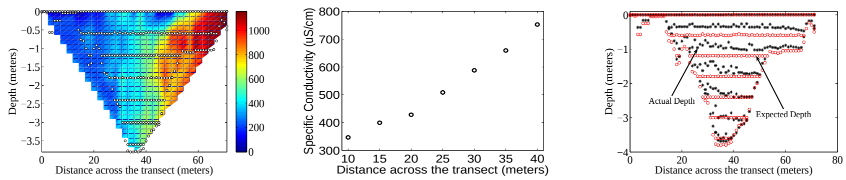
Finally, interspersed with these experiments, adaptive experiments should be designed to enable sampling with larger spatial coverage without discounting the temporal dynamics in the phenomena. During the two campaigns, we designed several experiments that adapted to the specific observed distribution by varying the sampling density and dwelling time besides others. Post deployment, the collected data should be used to validate the known understanding of the observed environments and hypothesize the deviations. The detailed characterization performed using IDEA methodology can be used as informed prior for the design of effective adaptive sampling approaches.

4 Observations following IDEA methodology

We present here the analysis from several experiments performed by actuating NIMS in the two dimensional cross-section of the aquatic environments, following the



(a) Nitrate concentration ($\mu Mol/liter$) at transect-1 (b) Standard deviation for nitrate at transect-1 (c) Diurnal variations in specific conductivity at 1.2 m



(d) Specific conductivity ($\mu S/cm$) at transect-2 (e) Mean specific conductivity during 5 minute interval (f) Characterization of localization error along y axis

Fig. 4: Experimental results from San Joaquin deployment

IDEA methodology. Piecewise bilinear interpolation is performed between the observed locations to create the surface distributions. For each surface distribution, the mean for all the observations is considered as the observed value at that location. Points in the surface distributions, wherever applicable, represent the observation locations. As per the coordinate system, x-axis represent distance along the cross-section of the aquatic environment, while y-axis represent distance along the depth ($y = 0$ represent the water surface). For the San Joaquin deployment the water stream on the near side (lower x coordinate value) is coming from the Merced River while the water stream on the far side is coming from San Joaquin river. Due to the space constraints here we only present the analysis from SJ deployment. Detailed characterization of both SJ deployment and Fulmor deployment is discussed in [16]. Additionally, only a subset of the observed variables are shown for each campaign. However, it is important to note that these are representative of the similar trends shown by all other observed parameters in each of the two campaigns.

The scientific objective for the San Joaquin campaign was to characterize the transport and mixing phenomena at the confluence of two distinct rivers - Merced river (relatively low salinity) and the agricultural drainage-impacted San Joaquin River (relatively high salinity) by observing several parameters that may control the mixing behavior of the two streams. Multiple river cross-sections were observed during the campaign. These transects were chosen based on prior observations [17] in the same environment. Sampling density was selected in consultation with the scientists to be 1 meter across the river and 0.6 meters along the depth resulting in a total of 250 and 293 observation locations respectively at the two transects. Dwelling time at each of the observed location was selected to be 30 seconds.

4.1 Spatial Trends

Fig. 1c and 4a show the spatial distribution of specific conductivity and nitrate concentration respectively. Each of these observed distributions is uniform towards each end of the cross-section while displaying dynamic variability in the mixing zone in the middle, resulting in vertical stratification. Moreover, comparing Fig. 4d, that displays the distribution of specific conductivity at the cross-section of the second transect (closer to the confluence point of the two rivers), with Fig. 1c, one can observe increased mixing of water downstream at the first transect resulting in reduced gradient in specific conductivity between the two ends of the river.

4.2 Temporal trends

Fig. 4c shows the distribution of specific conductivity at a depth of 1.2 meters during different times of the day. As the day progresses, gradient of specific conductivity between the two flows reduces. Fig. 4b shows the contour plot for the standard deviation observed at each of the sensing locations during the 30 seconds of dwelling time. As expected, maximum standard deviation (which is not significant) is in the mixing zone in the middle. To characterize the temporal variations at point locations, as would have been captured by a static sensing device, the NIMS scanning pattern included a dwell for 5 minutes at several locations along the cross-section at a depth of 1.2 meters. Fig. 1b shows this distribution at three such locations while Fig. 4e displays the mean specific conductivity observed at each of these locations during the observation time of 5 minutes. As can be observed, there was higher variation during this short time towards the middle of the river (x=30) as compared to near end of the transect (x=15). This further confirms higher standard deviation and larger dynamic range in the middle mixing zone observed during raster scans.

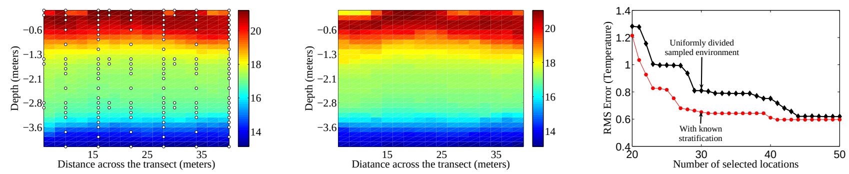
4.3 Adaptive Experiments

Based on the observed uniformity in the spatial distribution and higher temporal variability in the mixing region, an adaptive experiment was designed where the total time of the experiment was fixed at 1 hour and the dwelling time at each location (only the locations at depths 1.2 meters and 2.4 meters were chosen for this experiment) was selected based on the observed standard deviation, instead of keeping it uniform. Such an approach reduced the dwelling time towards each shore to 10 seconds while increasing the dwelling time in the middle region to as high as 80 seconds. This experiment was performed from 16:20 to 17:20 and the results can be seen in Fig. 4c, labeled accordingly. Comparing it with a raster scan performed the previous day around the same time (17:30 - 20:00), one can observe that reducing the dwelling time towards each shore end provides similar results to a deterministic scan while requiring substantially lower total observation time.

Interplay between the robotic sensing system and observed environment, in this case, involves a possible static deflection of the suspended, pendulous sensor system due to the flow of water in the river. Since this may induce a localization error along the vertical (y) axis, measures must be taken to ensure that this is precisely characterized to be accounted for in any actuation algorithm. Thus, a depth sensor was used, and integrated with the sensor payload, to directly determine depth. Fig. 4f presents a comparison of depth reported by the actuation system (containing possible deflection error), and the actual system localization determined by the integrated depth sensor.

5 Informed priors for improved adaptive approaches

Here, we discuss how two previously proposed adaptive approaches for environment sensing: (1) Stratified Adaptive Sampling [9]; and (2) Hierarchical Gaussian Process



(a) Temperature field used for Adaptive Sampling.

(b) Estimation using stratified adaptive sampling

(c) Performance improvement for GP based approach

Fig. 5: Performance improvements in adaptive algorithms when using the information about phenomena distribution at Lake Fulmor as informed prior. Points in (a) represent the locations selected by stratified adaptive sampling algorithm

Based Modeling [10]; displayed improved performance when prior information regarding the spatial distribution of phenomena, as characterized using IDEA methodology, is known as an input.

5.1 Stratified Adaptive Sampling

Several adaptive sampling approaches have been proposed in literature that hierarchically stratify the observed environment, thus performing dense sampling only in the regions of interest (regions of high phenomena variability) [9]. For this purpose, a sparse scan of the environment is required in the first stage to extract the regions of high variability. In the second stage, dense observations are performed only in these regions. If the phenomena shows considerable variability then this approach will result in poor performance. However, if some information about the spatial structure of the phenomena distribution is available as a prior, then it can be used to bias the first stage sampling accordingly.

Fig. 5a displays a typical example of temperature distribution, observed using NIMS at Lake Fulmor. This field was used as an input distribution to be observed by the stratified adaptive algorithm, discussed in [9]. When no information was known about the phenomena distribution, the algorithm made observations at 270 locations (out of a total of 522 locations) and estimated the field with root mean square error of 0.04. However, when the vertical stratification information was given as prior, the algorithm observed at only 135 locations (a significant reduction in sampling time) and estimated the distribution with root mean square error of 0.11 (a modest increase in sampling error). Points in Fig. 5a represent the observation locations selected by the algorithm when vertical stratification information was known as an input. Fig. 5b represent the field distribution estimated by making observations at this subset of locations using local polynomial fitting of second order over the observed locations.

5.2 Gaussian Process Based Modeling

A common approach in statistical methods for addressing a spatially distributed phenomena is to use a rich class of probabilistic models called Gaussian Processes (GPs) [10]. Using such models, one can quantify the informativeness of a particular location, in terms of the uncertainty about our prediction of the phenomena, given the measurements made at already visited locations. To quantify this uncertainty, we used the mutual information (MI) criterion [11]. If the phenomenon is discretized into finitely many sensing locations $\mathcal{V}$, then for a set of locations $\mathcal{P}$, visited by the mobile robot, the MI criterion is defined as:

$$\mathrm{MI}(\mathcal{P}) \equiv H(\mathcal{X}_{\mathcal{V}\setminus\mathcal{P}}) - H(\mathcal{X}_{\mathcal{V}\setminus\mathcal{P}} \mid \mathcal{X}_{\mathcal{P}}), \quad (1)$$

where $H(\mathcal{X}_{\mathcal{V}\setminus\mathcal{P}})$ is the entropy of the unobserved locations, and $H(\mathcal{X}_{\mathcal{V}\setminus\mathcal{P}} \mid \mathcal{X}_{\mathcal{P}})$ is the conditional entropy after observing at locations $\mathcal{P}$. Hence mutual information measures the reduction in uncertainty at the unobserved locations.

In particular, we first learned a non-stationary GP model, (using an extension of [11]), by maximizing the marginal likelihood [10] using a subset of temperature data at Lake Fulmor. This non-stationary process was learned by dividing the complete region into smaller sub-regions and combining the locally-stationary GPs from each of these sub regions. The two algorithms compared are when: (1) The observed environment was divided into smaller sub-spaces uniformly; and (2) The sub-spaces were selected based on horizontal stratification. For both approaches, starting from initial greedy set of 20 locations, additional 30 locations are selected, again greedily, based on the mutual information criterion given in Equation (1). Greedy selection is performed as it had been proved to provide near optimal value of information [11]. Fig. 5c compares the root mean square error in prediction after making observations at each of the selected locations for the two approaches. GP based approach with known input information regarding the horizontal stratification performs better than the approach with uniformly divided observed environment.

6 Conclusions and future work

Traditional approaches for environmental sampling are not capable of capturing the spatiotemporal dynamics of a complex phenomena, spread over large spatial extent, with high resolution. In this paper we described the Iterative experiment Design for Environmental Applications (IDEA) approach that we developed and used in two recent field campaigns for observing the aquatic system. IDEA is based on an iterative execution of experiments that reveal phenomena behavior with an objective to perform detailed characterization of the dynamics in the observed phenomena distribution. We demonstrated the effectiveness of the IDEA methodology by providing a detailed characterization of two markedly different aquatic phenomena. Literature survey reveals that this is the first time that an iterative experiment design approach is used an a mobile robot sensing system for characterizing environmental phenomena. We demonstrated that with the use of actuated mobile robotic system, it is imperative to characterize the interplay between the sensing system and the observed environment such that the integrity of the collected data can be relied upon. We illustrated this characterization through experiments based on in-field observations such as characterizing the static deflection of the suspended pendulous sensor system in an environment with significant flow Observations from such experiments, when included in creating new models for experiment design, result in increased fidelity and reliability in the collected data.

Finally, we also illustrated how accurate and detailed characterization, performed by actuating NIMS using IDEA methodology, can be used to improve the performance of existing model-based or adaptive experimental design approaches. This improved performance is enabled by using the observed spatial structure as prior information for the observed environment. In the future we plan to investigate model based adaptive approaches, that can use the specific spatial and temporal structure existing in the real environment as prior information in real time. We also plan to further advance the theoretical treatment of several environmental phenomena and incorporate these into a model based experimental design approach. These steps will then extend the IDEA system to enable it to autonomously determine the set of experiments to perform based on previously collected data and the known understanding of the observed phenomena.

Acknowledgment

This material is based upon work supported in part by the US National Science Foundation (NSF) under Grants ANI-00331481 and CCR-0120778. We also acknowledge the support of the team involved in the development of NIMS and the two campaigns discussed in this paper and Andreas Krause for the code for generating non-stationary GPs.

References

1. R. Hardy, P. Bates, and M. Anderson, "The importance of spatial resolution in hydraulic models for floodplain environments," *Journal of Hydrology*, vol. 216, pp. 124–136, 1999.
2. J. Wheaton, G. Pasternack, and J. Merz, "Spawning habitat rehabilitation i. conceptual approach and methods," *Intl. Journal River Basis Management*, vol. 2, pp. 3–20, 2004.
3. R. Gelda and S. Effler, "Metabolic rate estimates for a eutrophic lake from diel dissolved oxygen signals," *Hydrobiologia*, vol. 485, pp. 51–66(16), October 2002.
4. A. Dhariwal, B. Zhang, B. Stauffer, C. Oberg, G. S. Sukhatme, D. A. Caron, and A. A. Requicha, "Networked aquatic microbial observing system," in *IEEE ICRA*, 2006.
5. D. R. Blidberg, "The development of autonomous underwater vehicles (auvs); a brief summary,," in *IEEE ICRA*, 2001.
6. M. Kumagai, T. Ura, Y. Kuroda, and R. Walker, "A new autonomous underwater vehicle designed for lake environment monitoring," *Advanced Robotics*, vol. 16, pp. 17–26, January 2002.
7. B. Jordan, M. Batalin, and W. Kaiser, "Nims rd: A rapidly deployable cable based robot," in *IEEE ICRA*, Rome, Italy, May 2007.
8. 2006, http://www.jamesreserve.edu.
9. M. Rahimi, R. Pon, W. J. Kaiser, G. S. Sukhatme, D. Estrin, and M. Srivastavan, "Adaptive sampling for environmental robotics," in *IEEE IROS*, April, 2004, pp. 3537–3544.
10. M. Seeger, "Gaussian processes for machine learning," *Int. Jour. of Neural Systems*, vol. 14, no. 2, pp. 69–106, 2004.
11. A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research (submitted)*.
12. 2006, http://www.hachenvironmental.com.
13. 2006, http://www.satlantic.com.
14. 2006, http://www.sontek.com.
15. R. Pon, M. Batalin, J. Gordon, M. Rahimi, W. Kaiser, G. Sukhatme, M. Srivastava, and D. Estrin, "Networked infomechanical systems: A mobile wireless sensor network platform," in *IPSN'05*, 2005, pp. 376–381.
16. A. Singh, M. Batalin, M. Stealey, V. Chen, M. H. Hansen, T. C. Harmon, G. S. Sukhatme, and W. J. Kaiser, "Idea: Iterative experiment design for environmental applications," in *CENS Technical Report 67*, November 2006.
17. T. C. Harmon, R. F. Ambrose, R. M. Gilbert, J. C. Fisher, M. Stealey, and W. J. Kaiser, "High resolution river hydraulic and water quality characterization using rapidly deployable networked infomechanical systems (nims rd)," *Environmental Eng. Science*, 2006.