# IDEA: Iterative experiment Design for Environmental Applications

Amarjeet Singh
Electrical Engineering Department
UCLA
amarjeet@ee.ucla.edu

Maxim Batalin
Center for Embedded Networked Sensing
UCLA
maxim@cens.ucla.edu

Michael Stealey
Electrical Engineering Department
UCLA
stealey@ucla.edu

## ABSTRACT

This paper reports the first application of actuated sensing systems for high spatiotemporal resolution characterization of the three-dimensional environment of river and lake aquatic systems. The development of a new method and its verification in these two application areas is described. Both applications involve dynamic phenomena - one resulting from flow of the water and the other from rapidly evolving biological processes. These applications are typical environmental monitoring problems. They exemplify the key challenge in such problems - characterizing phenomena displaying spatiotemporal heterogeneity. In many such examples, the application requires a diverse array of measurements based on sensors for physical, chemical and biological systems. Together, these requirements pose a significant challenge for conventional sensor network methods.

We describe the development and applications of a new general purpose method for actuated sensing - Iterative experiment Design for Environmental Applications (IDEA). IDEA introduces a new in-field adaptation of the sensing systems including static and actuated sensors. IDEA addresses the limitations of previous sampling approaches, for example conventional adaptive sampling, by guiding adaptive sampling with an iteratively developed phenomenon model. This paper presents applications of IDEA to: (1) Three-dimensional characterization of contaminant concentration and flow at the confluence of two major rivers; and (2) Characterization of phytoplankton dynamics in a lake system. These applications provide ideal tests by presenting complex structures associated with each phenomenon and enabling a comprehensive evaluation of the general applicability of IDEA methodology. Improved performance using guided adaptive sampling is demonstrated for two existing methodologies, stratified adaptive sampling and hierarchical non-stationary Gaussian Processes. The IDEA experimental results both validate the general applicability of this method and also have advanced the understanding of interrelated physical, chemical and biological processes in these sampled environments.

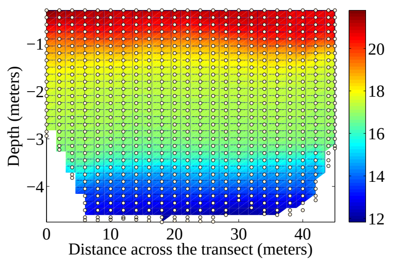
## 1. INTRODUCTION

A broad class of environmental sampling applications, both terrestrial and aquatic, requires sampling of an environment that displays significant heterogeneity in both space and time. As an example, the growth of planktonic microorganisms, that form much of the base of the food webs in freshwater and marine ecosystems, is dependent on the availability of dissolved nutrients and light. These parameters display dynamic spatiotemporal distributions [1] often due to interdependence with other physical and chemical parameters such as temperature and pH. As another example, river observations are useful for answering the questions pertaining to the hydraulics and multi-dimensional river modeling [2], geomorphology, sediment transport and riparian habitat restoration [3]. These require high granularity measurements of coupled hydraulic and water quality parameters in stream cross-sections. Stationary, turbulent flow of water in the streams result in heterogeneity in time. On the other hand, mixing of the two streams in a confluence zone results in spatial heterogeneity. Figure 1 displays the temperature distribution sampled in a lake and the contaminant distribution (in space and time) in a river, showing the high degree of heterogeneity in such an environment.

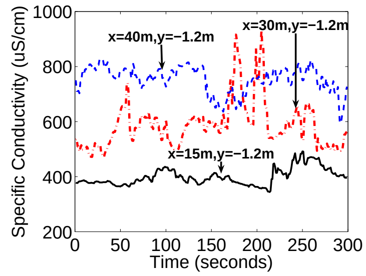
Environmental monitoring for detailed characterization of such phenomena require high resolution (in both space and time) sampling. Traditionally aquatic environments are sampled sparsely in both space and time (horizontally and vertically) because of the difficulties and the associated cost of continuously accessing these environments [4, 5]. These constraints led to the development of models that were often simplified and calibrated using sparse measurements from a small network of static sensors or manually deployed measurements. Over the last few years, with the development of various robotic platforms, several prototype systems have been used for sensing the aquatic environments at different scales. Buoyed or moored deployment platforms now exist which can provide vertical profiling capabilities with high temporal resolution over long time periods at key locations [6, 7]. Mobile robotic platforms such as boats [8], Autonomous Underwater Vehicles [9, 10] and cabled robotic systems [11], 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 to perform complete, three-dimensional characterization of aquatic systems on this scale.

In spite of the recent advancements in technology, currently available systems have many constraints that preclude long-term, remote, autonomous, high-resolution monitoring in the real environment. These constraints include availability of energy, precise autonomous actuation, precise localization, data communication and others. Thus a large team effort for any such field campaign results in only a limited time in the field for characterizing the phenomena with high fidelity. Limited campaign time along with the high associated resource cost (team effort and the cost of robotic and sensing systems) demand for optimized utilization of the available time in the field.

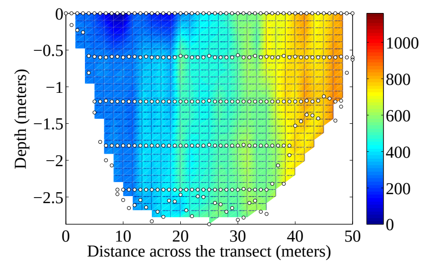
In this paper, we present IDEA - Iterative experiment Design for Environmental Applications; an approach for characterizing an un-




(a) Spatial distribution of temperature ($^oC$) at Lake Fulmor




(b) Temporal variation at San Joaquin




(c) Spatial distribution of specific conductivity ($\mu S/cm$) at at San Joaquin

Figure 1: Heterogeneity in space and time for environmental phenomena. Points in the figure represent sampling locations. x axis is across the transect and y axis is along the depth with negative values below the water surface.

known environment under constraints of rapidly evolving phenomena. This involves an in-field adaptation in the experiment design, to capture phenomena dynamics exploiting observations from prior, 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 sampled phenomena with high fidelity. Analysis of this data can then be used to validate the underlying physical, chemical and biological characteristics and to observe the deviations, if any, from the current understanding. Additionally, based on observations in the field, experiments can be designed to characterize the constraints of the sampling system. Once the phenomena and the system characteristics are known, this information can then be used to create an efficient model for adaptively sampling the phenomenon - enabling sampling with larger spatial coverage without discounting the dynamics in the phenomenon. Thus, by incorporating the spatiotemporal variability of the observed phenomena distribution, IDEA will result in improved performance in sampling fidelity over several existing adaptive sampling approaches that perform sampling without any such prior, experimentally validated information [12, 13].

To illustrate IDEA and its validation in real world applications, we present case studies of two environmental sensing application campaigns where such an approach was successfully followed and validated. The first campaign was at the confluence of San Joaquin and Merced rivers in California from August 21-25, 2006 (hereafter referred to as San Joaquin deployment). Figure 2a displays the satellite view of this deployment site. The scientific objective for this investigation was to characterize the transport and mixing phenomena at the confluence of two distinct rivers: the Merced River and the San Joaquin River. Three-dimensional characterization of the river basin was enabled by performing actuated sampling at multiple river cross-sections. The second campaign was at a small (1.2 hectare) lake located in a sub alpine (1,600 m) coniferous forest within the San Jacinto mountains of Southern California [14] from August 28-31, 2006. Figure 2b displays the satellite view of Lake Fulmor. The goal of this study was to employ a suite of static and mobile sensor technologies enabling biological studies of the dynamic phytoplankton community in a complete three dimensional environment.

As illustrated in Figure 1a and Figure 1c, specific spatial structure is observed in these seemingly similar environments. Spatial distribution at Lake Fulmor displays horizontal stratification while at San Joaquin deployment, vertical stratification is observed. Such structures, when learned in the IDEA iterative cycle, can be used to build models for effectively sampling the phenomena. We provide an illustration of the effectiveness of observing and learning such structures, by demonstrating improvements in two of the existing model based approaches: (1) Stratified adaptive sampling [12]; (2) Hierarchical non-stationary Gaussian Process based modeling approach [15]. More specifically, our main contributions are:

- In-field adaptation of static and actuated sensor systems based on observations made from iteratively designed experiments.
- Validation of the known understanding about the interdependence of several physical, chemical and biological processes in the sampled environment.
- Illustration of improvements in existing adaptive sampling approaches, if adapted based on in-field observations.
- First three dimensional characterization of lake and river aquatic systems, typical of several environmental sampling applications.

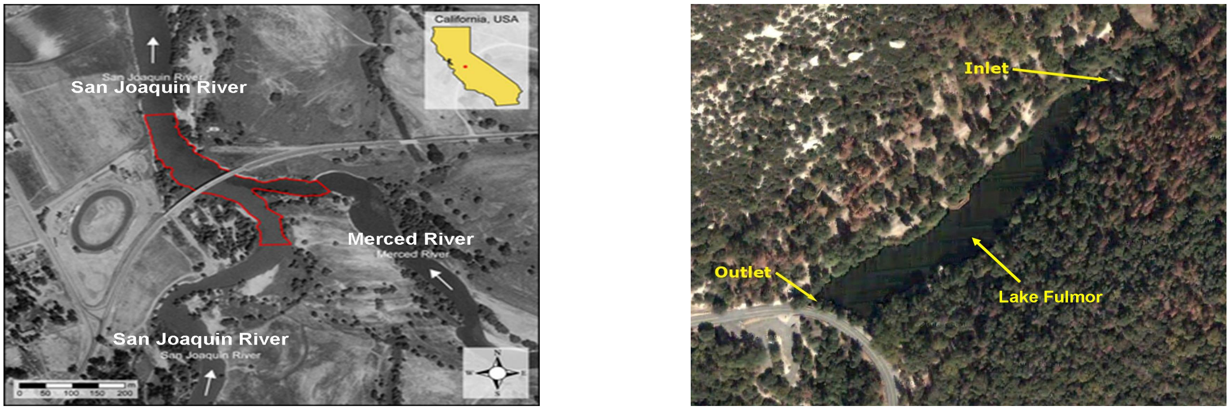
Thus IDEA presents a methodology for characterizing the spatiotemporal distribution of the sampled phenomena by iteratively adapting the experiment design in the field based on observed phenomena characteristics. Details of this methodology are described in Section 4

## 2. SYSTEM DESCRIPTION

An actuated sensor system, Rapidly Deployable Networked Info Mechanical System [11] (hereafter called NIMS), was used during each of these two campaigns to navigate a sensor payload anywhere within the two dimensional cross-section of the aquatic environments. This enabled high resolution sampling of several important environmental phenomena along the complete cross-section. Figure 3a and 3b show the schematic diagram of the system. It is comprised of three main cables, the mounting hardware, actuation module and a shuttle. The three cables are: 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. The static cable can be attached to the ground anchors and routed over the ladders or anchored to any other natural or man made structures (e.g. trees, railing, poles, etc.). The mounting hardware is supported by the static cable and is used to attach the actuation module at one transect end and horizontal and vertical cables at the other end. The shuttle and the vertical node platform are actuated using two motors located on one side of the transect, through the horizontal and the vertical cables.

The following sensors were used during each of the two campaigns for monitoring different physical, chemical and biological phenomena:

1. Hydrolab Minisonde 4a [16]: Sensing temperature, pH and specific conductivity.
2. Hydrolab DS5 [16]: Sensing Photosynthetically Active Radiation (PAR), depth, oxidation reduction potential, turbidity and Luminescent Dissolved Oxygen (LDO).




(a) Confluence of San Joaquin and Merced River

(b) Lake Fulmor, Southern California

Figure 2: Satellite view of the deployment sites where iterative experiment design approach was pursued.

3. ISUS sensor [17]: Sensing nitrate concentration.
4. Argonaut-ADV [18]: Used for flow monitoring.
5. Fluorometer: Measuring the concentration of chlorophyll-a which is indicative of the density of phytoplankton in the environment
6. Physical Sampling Device: Designed and developed at UCLA specifically for these campaigns to collect physical water samples for detailed lab analysis.

Figure 3c shows a picture of all sensors and the physical sampling device. Sensing multiple parameters simultaneously enabled the correlation and effective characterization of the sampled phenomena dynamics. For example, sensing dissolved oxygen, temperature, pH and PAR at Lake Fulmor helped characterize the interrelated parameters causing the variability in the lake environment. Similarly, sensing temperature, specific conductivity, nitrate and pH helped characterize salt concentration in the river basins.

The use of cable infrastructure supporting the mobile system introduces high precision in transport. However, to exploit this capability, several deployment design limits must be incorporated that reduce the effective experiment time during the campaigns. These include installation of the required infrastructure and anchors for supporting large tension required to support the sensor payload over long transects ($\approx 70m$), in-situ localization calibration and robustness against in-field failures. Additionally, success and timely completion of such campaigns heavily depends on participation of a multi-disciplinary team of researchers resulting in a huge team effort. Such factors demand for optimized usage of the limited campaign time to characterize the sampling phenomena with high fidelity.

Additionally, the speed of the mobile robot is limited both by the system electromechanical constraints and also the sensor system requirements. Due to the response time of the above water quality sensors that may be characterized with the time constants of up to 10 seconds, sensor devices must dwell at sampling locations to permit accurate measurements. Limited speed of the mobile robot along with dense spatial sampling and corresponding dwelling time at each location increases the sampling time for each experiment. This necessitates requirement for adaptively sampling the phenomenon under investigation to provide large spatial coverage within the limited sampling time. Such approaches are inherently preferable to deterministic motion profile (for example simple "raster" scan patterns). Spatial structures such as those presented in Figure 1a and 1c, if learned in the IDEA cycle , will further improve the effectiveness and accuracy of adaptive sampling approaches.

Software architecture included system for monitoring the real time data stream from the system to provide immediate fault detection as well as to monitor the variability in the sampled phenomena distribution in real time. Additionally, adaptive algorithms designed based on observed phenomena distribution were integrated in the field with the existing architecture to provide autonomous system operation.

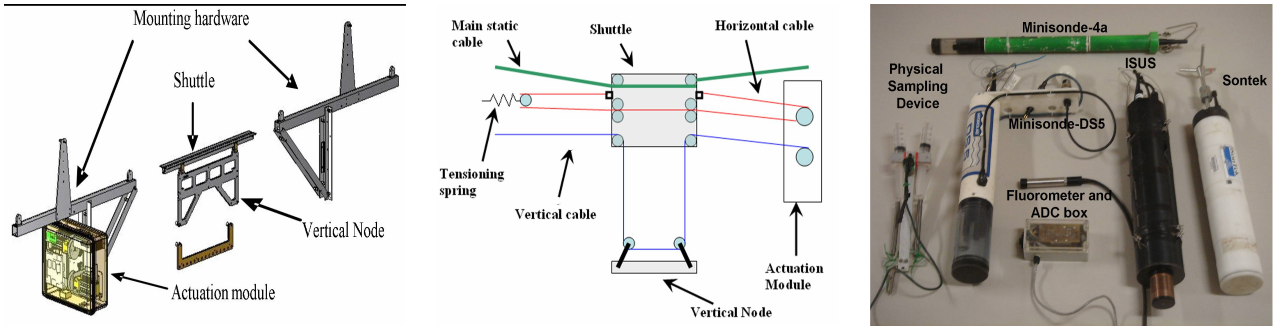
## 3. AQUATIC SAMPLING MOTIVATION

Water quality monitoring is increasingly important from many perspectives including public health, protection of agricultural irrigation sources, and environmental protection. Measurement objectives include the study of ecological behavior, hydraulics and multidimensional river modeling [2], geomorphology, sediment transport and others. High resolution spatiotemporal observations enable distributed model parameterization and provide policy-makers with the information enabling them to make more informed decisions about issues extending beyond water supply and flood control to the restoration of river habitats.

### 3.1 Sampling in San Joaquin river

The first IDEA campaign occurred from August 21-25, 2006 at the confluence of San Joaquin and Merced rivers in Central California.Figure 2a displays the satellite view of the deployment site. San Joaquin is a heavily managed river basin with a robust network of 14 river gauging stations with a spatial granularity of tens of kilometers being operated and maintained by government agencies. This network provides access to the current gross river conditions via the Internet [19], allowing engineers, scientists, and water resource managers to consider real-time adaptive management of these large and complex systems [20]. Deterministic river flow and transport models can play a role in determining optimal distributions. Though, for regulatory work, these are often simplified (one-dimensional) routing models that assume complete mixing of chemical species throughout the river cross-section [21]. An attempt can then be made to calibrate these routing models using time series data from a network of river gauging stations which record gross flow and, in some cases, temperature and bulk salinity in terms of the water's electrical conductivity [21].

In spite of impressive progress towards the automated management of the San Joaquin basin, water resources, population influx, land use changes and climate change impacts create a demand for the distributed three-dimensional high resolution observation of the flow and the water quality parameters. The primary scientific objective for this investigation was to characterize the transport and mixing phenomena at the confluence of two distinct rivers: the Merced River (relatively low salinity) and the agricultural drainage-impacted San Joaquin River (relatively high salinity) by sampling




(a) Major components of the system. (b) Schematic view with the basic cable setup. (c) Sensor suite used during the campaigns

Figure 3: System architecture and sensor payload used during the campaigns. Part a. also shows the relative layout and orientation of the system, although they would be much more spread out in reality.

many parameters that may control the mixing behavior of the two streams.

### 3.2 Sampling in Lake Fulmor

The second campaign occurred at Lake Fulmor, James Reserve [14] from August 28-31, 2006. The goal of this study was to establish, test and employ a suite of static and mobile sensor technologies in Lake Fulmor in order to enable biological studies of the phytoplankton community. Figure 2b shows the satellite view of Lake Fulmor. It is a small (1.2 hectare) lake located in a sub alpine (1,600 m) coniferous forest within the San Jacinto mountains of Southern California. The lake has a single source of freshwater at its northern end and a single outlet at its southern end. River flow in and out of the lake ceases during much of the summer/fall period, making it an excellent model system for studying phytoplankton dynamics in the absence of large advective processes. Understanding the phytoplankton dynamics is dependent on characterizing several chemical and physical properties of the water column such as availability of nutrients, PAR, temperature, pH etc., and the interplay between these factors and the microorganisms community.

## 4. IDEA METHODOLOGY

IDEA provides a methodology for in-field adaptation of experiment design to perform detailed, three-dimensional characterization of the spatiotemporal distribution of the sampled phenomena. Following are the different experiment designs proposed under IDEA methodology with an illustration of their applicability to different sensing systems.

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 distribution of sensor network. On the other hand, with an actuated sensing system, 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 known characteristics of the sampled phenomena. This can then be iteratively improved till it is sufficient to completely characterize the spatial variability in the phenomena distribution. For each of our aquatic sensing applications, dense raster scans were performed by actuating NIMS to navigate the sensor payload in the two-dimensional cross-section of the sampled environment. The density of these raster scans was decided in consultation with environmental scientists and aquatic biologists and was influenced by prior observations in the same environment. This was found to be sufficient for phenomena characterization based on the observed phenomena distribution.

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 but these measurements will be restricted to only a few locations in the sampled environment (due to high resource cost associated with the deployment of dense static sensor network). In the case of actuated systems, diurnal variation should be characterized by repeating the dense raster scans at different times of the day. In addition to providing temporal variation in phenomena characteristics, these 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) that enabled us to understand the mixing effects across different stratas of sampled aquatic system.

Dwelling time at each sensing location, while performing raster scans, should be selected so as to satisfy both the requirements for sampling short term temporal variations across the complete sampling environment and the response time of the sensing system. Further, a subset of locations (representative of the sampled environment based on spatial distribution of the phenomena) can be selected and the dwelling time at these locations can be increased further to perform detailed characterization of the short term temporal variation using actuated sensing systems. 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 confirms 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. Additionally, during the San Joaquin deployment we selected a subset of locations along the cross-section of the river and increased the dwelling time to 5 minutes at each of these locations.

In addition to characterizing the sampled phenomena distribution, experiments must be designed to characterize the constraints of the sampling system such as localization drift, interference in the sampled environment and others, such that the integrity of the sampled data can be established. For example, for sampling in a physically static Lake Fulmor environment, we designed experiments to characterize the mixing of water across different stratas due to the motion of NIMS inside the water. On the other hand, for sampling in the San Joaquin environment with significant flow, we designed experiments to characterize the static deflection of the suspended, pendulous sensor system while sampling inside the water.

Finally, interspersed with these experiments, adaptive experiments should be designed to enable sampling with larger spatial coverage without discounting the 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 sampled parameters, hypothesize the deviations and develop effective adaptive sampling approaches. We present experimental validation of related physical, chemical and biological processes in the sampled environment. Further investigations for the theoretical treatment of several environmental phenomena and for the development of improved adaptive sampling approaches are currently underway.

# 5. SAMPLING USING IDEA

At each of the two campaigns, experiments were designed iteratively based on: (1) Observations from previous iterative experiments; (2) Phenomena behavior dependence on observed parameters; and (3) Environment specific system constraints. These were also designed to optimize the limited field time available for the campaign and to perform detailed three-dimensional characterization of the sampled phenomena distribution. Several dense raster scans were performed using NIMS, navigating the sensor payload in the two-dimensional cross-section. This cross-section was selected based on prior observations from statically distributed sensors used for sampling in the same environment. Sampling density and dwelling time at each of the sampling locations was decided based on prior understanding of the sampled environment. Graphical inspection suggested that this simple algorithm did not produce any misleading structures in the data (in any sensing exercise, we need to be concerned with "outlying" observations, suggesting more robust aggregation techniques). Response time of the used sensors was also accounted for in selection of dwelling time.

Whenever multiple observations were taken at a single location (a "dwell point") by the robotic unit, we report a mean and standard deviation. In what follows, our surface displays are constructed from simple bilinear interpolation. These are observed, through all experiments, to not introduce artifacts. Other reconstruction methods, including those introducing smoothing and providing estimation rather than interpolation may be applicable as well. Points in the figures, wherever applicable represent the sampling locations. Due to the space constraints here, only a subset of the sampled phenomena are shown for each characteristic. However, it is important to note that these are representative of the similar trends shown by all other sampled phenomena in each campaign.

## 5.1 San Joaquin Campaign

Several experiments were conducted to characterize the transport and mixing phenomena at the confluence of the Merced River and the San Joaquin River. High spatial resolution sensing experiments were performed at two different cross-section locations defined by NIMS transects. Location of the two transects was chosen based on prior observations [22]. The first transect was selected at approximately 290 meters downstream from the confluence location of the two rivers. Its width was approximately 50 meters. The second transect was selected at a point closer to the confluence ($\approx$ 130 meters from the confluence location) with the total length of approximately 70 meters. The data is displayed such that the horizontal coordinate, x, origin is at the Merced river while the San Joaquin river side lies at the maximum x values.

### *5.1.1 Spatial Trends*

High spatial resolution raster scans were performed at each of the two selected locations to characterize the spatial trends in the sampled phenomena. Sampling density at each of these transects was selected to be 1 meter across the river and 0.6 meters along the depth resulting in a total of 250 sampling locations at the downstream transect and 293 sampling locations at the upstream transect. Figure 1c and 4a show the spatial distribution of specific conductivity and nitrate concentration respectively. As can be clearly observed there exists a significant variation in concentration across the transect. 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 Figure 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 Figure 1c, one can easily observe the effects of water mixing. Increased mixing of water downstream at the first transect results in reduced gradient in specific conductivity between the two end of the river. Integrating the total salt concentration at each of these locations can further help characterize the total salt burden added from the river basin between the two transects. Higher value of specific conductivity toward San Joaquin can be attributed to the agricultural drainage impacting the river.

### *5.1.2 Temporal trends*

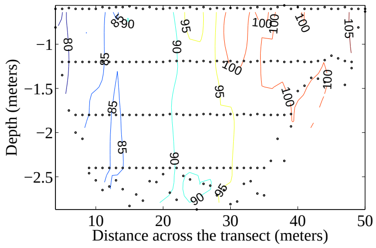
Dense raster scans were repeated at different times of each day and night (wherever possible) to characterize the diurnal variations in the phenomena distribution. Figure 4c shows the distribution of specific conductivity at a depth of 1.2 meters at different times. As the day progresses, specific conductivity on the San Joaquin side reduces while the mixing effect causes it to increase on the Merced river side (reducing the gradient between the two flows).

For each of the raster scans, the dwelling time at each location was selected to be 30 seconds to satisfy both demand for sampling short term temporal distribution and the response time of the sensors. These sensors have been previously characterized and shown to have response characteristics that permit stable measurements well within the 30 second dwell period. Figure 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.

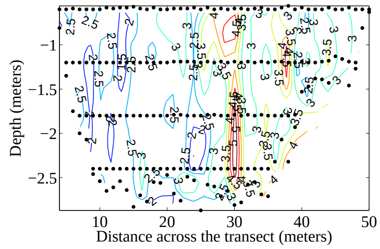
Distributing static sensors for sampling short term temporal variations, at a specific depth, will be difficult in this environment with significant water flow. Thus, to characterize such temporal variations at point locations, as would have been captured by a static sensing device, the NIMS scanning pattern included a dwell for 5 minutes at several locations along the cross-section at a depth of 1.2 meters. Figure 1b shows this distribution at three such locations while Figure 4e shows the mean specific conductivity observed at each of these locations during the sampling 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. Based on this observation, an adaptive experiment was designed where dwelling time was decided based on standard deviation at each location instead of uniform dwelling time (as was the case in raster scans). Details of this adaptive experiment are discussed next.

### *5.1.3 Adaptive Experiments*

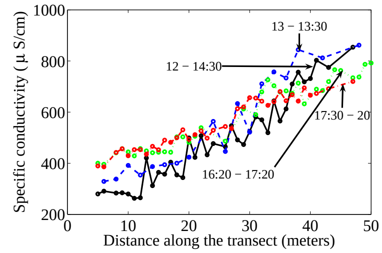
The presence of a high rate of flow in the river system may induce a static deflection of the suspended, pendulous sensor system. Since this may induce a localization error along the vertical, y, axis, means must be provided to ensure that this is precisely characterized to be accounted for in any actuation algorithm. Thus, a depth sensor was used, and integrated with the sensor payload, to directly determine depth. Figure 4f presents a comparison of depth reported



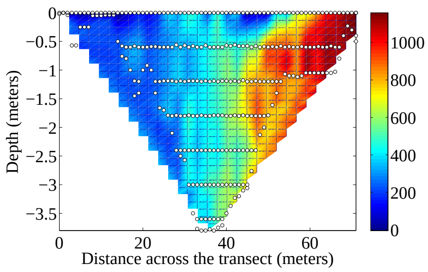

(a) Nitrate concentration ($\mu Mol/liter$) at the downstream transect during a raster scan on August 23 evening



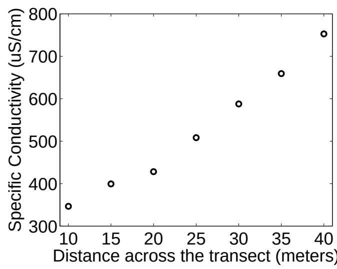

(b) Standard deviation at each location while sensing Nitrate at the downstream transect during the raster scan on August 23 evening



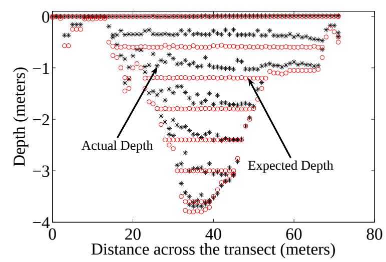

(c) Diurnal variations in specific conductivity at a depth of 1.2 meters




(d) Spatial distribution of specific conductivity ($\mu S/cm$) at the upstream transect as sampled on August 25.




(e) Mean for the Specific Conductivity during 5 minute sampling at each location




(f) Characterization of localization error along y axis.

Figure 4: Iterative experiments performed at the confluence of San Joaquin and Merced rivers, Central California.

by the actuation system (containing possible deflection error), and the actual system localization determined by the integrated depth sensor. Based on this observation, our system can account for the deviation from the expected depth to perform precise sampling, when using any adaptive sampling experiment. Such observations, when incorporated into the design of adaptive approach further improve the performance.

Since the phenomena demonstrated vertical stratification, a sparse sampling experiment was designed by reducing the sampling density more along the vertical, y, axis. The total time for the scan was approximately 35 minutes, compared to approximately 150 minutes required for the complete raster scan. The results from this uniform sparse scan, where sampling density was determined based on observations in the field, were similar to dense raster scan with significantly reduced sampling time, demonstrating a large performance advance.

Based on the observed uniformity in the sampled phenomena and higher temporal variability in the mixing region, another 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 observed standard deviation, instead of keeping it uniform. This has the result of further improvement in performance since the impulse response of the sensor system does not increase dwell time at all locations. Adapting 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 Figure 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. Again, the adaptive method provides a performance advance.

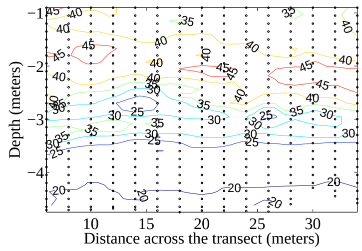
### *5.1.4 Mixing effects in rivers*

Validation of known understanding about the mixing effects in rivers using experiment observations at San Joaquin deployment are discussed next. Shear velocity in an ideal case of uniform, straight, infinitely wide channel of constant depth can be defined as $u^* = \sqrt{gdS}$, where g is the gravitational constant, d is the depth of the channel and S is the slope of the channel. With depth as the relevant length scale, turbulent mixing in an idealized flow will be proportional to the product $du^*$. For a wide range of flows, vertical mixing has been approximated as, $\epsilon_v = 0.05du^*$ and transverse mixing as, $\epsilon_t = 0.15du^*$ [23].

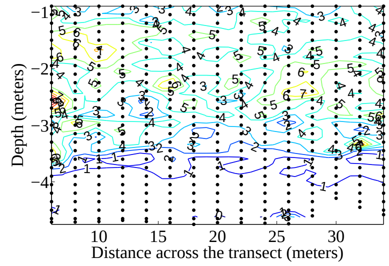
Natural channels differ from such ideal channels in several ways such as irregularity in depth, curvature in the channel and others. None of these factors will influence the vertical mixing rate significantly as it only depends on the local depth. In contrast, they do influence the transverse mixing, resulting in a modified approximation $\epsilon_t = 0.6du^*$ [23]. With these approximations, the transverse mixing coefficient will be an order of magnitude more than vertical mixing coefficient ($\epsilon_t \approx 10\epsilon_v$). Mixing time for a stream is proportional to the square of the length divided by the mixing coefficient. Thus for a river such as San Joaquin ($\approx$ 50 meters wide), transverse mixing time will be $50^2/10$ times the mixing time for vertical mixing. Due to this order of magnitude difference, which is typical of similar river systems, one can assume that the vertical mixing is effectively instantaneous compared to transverse mixing. This conforms with the vertical stratification observed in the sampled phenomena at San Joaquin. Furthermore, this can be used to calculate the mixing effect at a given distance from the confluence, with increased mixing for points further downstream. Figure 1c displays the specific conductivity at the first transect that was further downstream ($\approx$ 290 meters) compared to the second transect ($\approx$ 130 meters). Figure 4d displays the specific conductivity at the upstream transect. Comparing the two figures, increased mixing can be easily observed at the downstream transect with lower gradient of specific conductivity across the two ends of the river.

## 5.2 Lake Fulmor Campaign

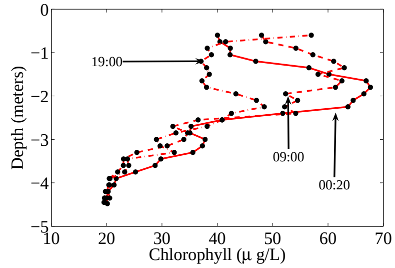
The goal of this campaign was to study the behavior of phytoplankton in the lake and its relationship with the underlying phys-



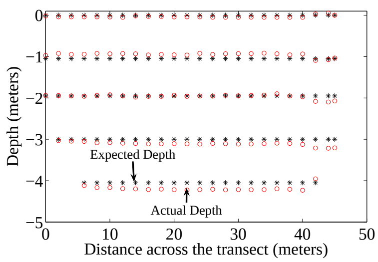

(a) Chlorophyll concentration sampled during one of the raster scans on August 29.



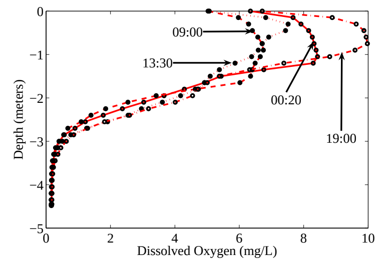

(b) Standard deviation for the chlorophyll data collected on August 29, 2006



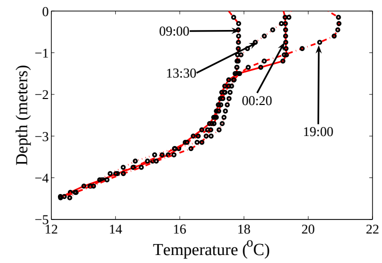

(c) Diurnal variations in chlorophyll-a ($\mu g/L$) at a distance of 32 meters from near shore side




(d) Characterization of localization along the depth




(e) Diurnal variations in dissolved oxygen (mg/L) at a distance of 32 meters from near shore side




(f) Diurnal variations in temperature ($^oC$) at a distance of 32 meters from near shore side

Figure 5: Iterative experiments performed at Lake Fulmor, James San Jacinto Mountain Reserve.

ical, chemical and biological parameters. Location of the transect was chosen based on the observations from previous campaigns in the same lake using statically distributed sensing systems [7, 8]. Total length of the transect was 45 meters.

### 5.2.1 *Spatial Trends*

Dense raster scans were performed to characterize the spatial distribution of the sampled phenomena in the two dimensional cross-section at the selected location. The density of the raster scan was selected to be 2 meters along the cross-section and 0.15 meters along the depth resulting in a total of 732 sampling locations. Phytoplankton concentration is measured by detecting the presence of chlorophyll-a using a compact, immersible, optical fluorescence sensor. Figure 1a and 5a display the spatial distribution of temperature and chlorophyll-a as sampled during one of the raster scans. Amount of chlorophyll-a is calculated from the observations made by the fluorometer attached as part of the sensor payload. Similar to the San Joaquin deployment, there exists a spatial structure in the observed phenomena but in this environment the stratification is horizontal instead of vertical as observed at the San Joaquin deployment. Chlorophyll distribution is uniform close to the top and the bottom while showing considerable dynamic variation in the middle layer. The opposing trend is observed for temperature resulting in horizontal stratification. Detailed analysis of such distribution of chlorophyll and temperature is provided in Section 5.2.4. This confirms that there exists considerable heterogeneity in both scales and type of spatial structure when sampling different phenomena and further confirms the requirement of IDEA methodology with iterative experiment design to characterize the phenomena distribution effectively.

### 5.2.2 *Temporal Trends*

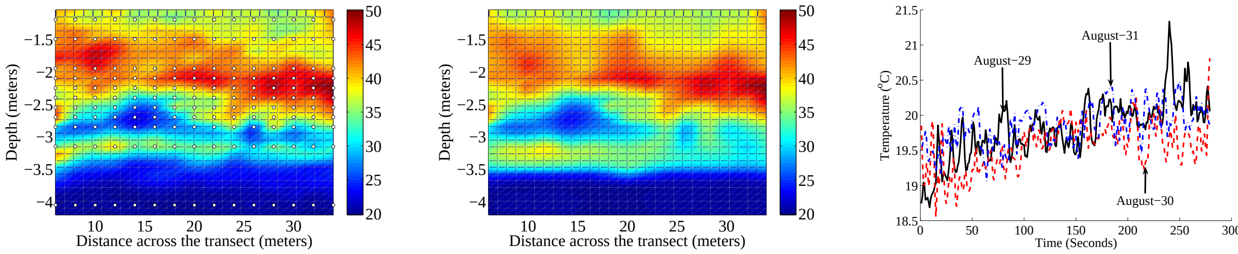
Dense raster scans were repeated at different times of the day to characterize the diurnal variations in the sampled phenomena. Figure 5c, 5e and 5f show the distribution of chlorophyll-a, dissolved oxygen and temperature respectively at a distance of 32 meters from the near shore side (x=32), as sampled during the four dense raster scans performed during the campaign. Considerable dynamics for the chlorophyll-a are observed in the middle layer where the temperature is uniform. Detailed discussion about such distribution is provided in Section 5.2.4.

To characterize the short term temporal dynamics and the response time of each sensor, NIMS scanning pattern included dwell periods at each sensing location for 10 seconds. Figure 5b displays the standard deviation at each position for the chlorophyll data collected during one of the raster scans. Lower values of standard deviation, even in the middle region with considerable dynamic variation, confirm that the dwelling time at each location can be reduced without compromising the quality of the collected data.

### 5.2.3 *Adaptive Experiments*

For such a lake environment with very low input flow (as confirmed by our placement of static sensors) and observed horizontal stratification, it is important to characterize the mixing induced in the water column by NIMS actuation while sampling at several locations along the depth. We performed two experiments to characterize this interference that the sampling system (NIMS) might be causing to the phenomena distribution. Firstly, we placed a static sensor close to the cross-section sampled by the NIMS and observed the sampled data when NIMS was sampling in the vicinity, comparing it with the sampled data during a different day when there was no sampling performed by NIMS in the vicinity.Figure 6c shows the temperature data from the static sensor, placed close to the NIMS transect for the three days during the August deployment from 12:50 - 13:36. Only on August 31, NIMS node was sampling in the vicinity (x=24 to x=34 during this period, when the static sensor was located at x=30). Secondly, instead of sampling at the two nearby locations consecutively along the cross-section, we alternatively sampled at two locations far away i.e. instead of sampling at 12m, 14m, 16m, 18m... along the cross-section as in a regular scan, we sampled at 12m, 24m, 14m, 26m.... Observations from both the experiments confirmed that there was no observable mixing induced by vertical motion of NIMS node. In addition to characterizing the phenomena distribution in the sampled environment, such experiments are required to characterize the effect of sampling system on the sampled phenomena distribution to ensure data integrity.

Based on our discussion with the biologists, we found that ther-




(a) Subset of surface distribution of chlorophyll as sampled by the fluorometer

(b) Surface distribution after local polynomial fit with sampling based on curvature of thermal gradient

(c) Experiment to characterize the churning of water caused by NIMS

Figure 6: Subset of the adaptive experiments performed during at Lake Fulmor. Points in part-a represent sampling locations selected based on curvature of thermal gradient.

mal stratification, resulting from temperature gradient across different stratas of the lake, prevents mixing of water across the gradient resulting in the biomass distribution getting locked in the middle layer of uniform temperature. To characterize the amount of biomass distribution in this middle strata of uniform temperature, we designed an adaptive experiment deciding the sampling density based on curvature of thermal gradient. Resulting sampling locations along the depth are sampled at three different locations along the cross-section (x = 16m, 24m, 32m). We further performed simulations to characterize the effectiveness of sampling based on such an approach. Figure 6a represents the subset of the distribution of fluorometry data from x=6 to x=34. Sampled data was observed in simulation with sampling density of 2 meters along the cross-section (x-axis), and sampling density along the depth (y-axis) based on curvature of the thermal gradient. This resulted in only 165 sampling locations out of a total of 330 sampling location in the complete subset. Points in Figure 6a represent the sampling locations selected using this approach. Local polynomial fit was performed over the data from these locations to create a complete surface distribution. Figure 6b represents a surface of predicted values using the local polynomial fit. As can be observed the predicted surface distribution is similar to the actual surface distribution shown in Figure 6a. The total root mean square error between the predicted and the actual surface was 4.89.

Additionally, in this example of a low net water flow aquatic system, we had again characterized the vertical localization using the depth sensor. Figure 5d displays the comparison of expected depth, reported by the system with the actual depth, reported by the sensor. As for the river environment, the depth sensor provides accurate localization and the NIMS system is shown to provide accurate control for navigating the sensor payload anywhere within the two dimensional cross-section of the sampled environment.

### 5.2.4 *Phytoplankton in Lakes*

Growth patterns of phytoplankton in a lake environment depends on several physical and chemical parameters. Light, required for photosynthesis is abundant near the surface of the water and its availability reduces with depth. Nutrient concentration is abundant near the bottom of the lake due to the sinking of decaying biomass and other material. Thus, in order to maximize the availability of both light and nutrients, phytoplankton tend to show preferred growth patterns at a level near the surface. This conforms with our observation in the field (please refer to Figure 5c), where we also observed peak levels of chlorophyll-a, representing the availability of phytoplankton, at a depth of approximately 1.5 meters with reduced levels both close to the surface (top) and bottom.

Wind effects and differences in density (lower density of the warmer water close to the surface due to increasing air temperature and solar radiation) causes mixing of water in the top layer. This results in significant dynamics in several parameters and conforms with our observations for temperature, pH and dissolved oxygen (please refer to Figure 5c, 5f). A middle layer of uniform temperature (from approximately 1.5 meters depth to 3.5 meters depth, Figure 5f), creates a thermal gradient that prevents waters from the top layer to mix with waters from the bottom layer. This prevents the nutrient rich water from the bottom from mixing with the water at the top, forcing the phytoplankton to display maximum concentration at a subsurface layer. This gradual mixing of water in the vertical direction even within the top layer for such an environment (with negligible total lake flow) is completely opposite to the instantaneous mixing (discussed in Section 5.1.4) in the case of water streams. This reinforces the requirement to first characterize the sampling phenomena distribution, with IDEA methodology, to understand the structure in spatiotemporal distribution. This structure can then be used to create efficient models to sample the phenomena distribution adaptively with high fidelity.

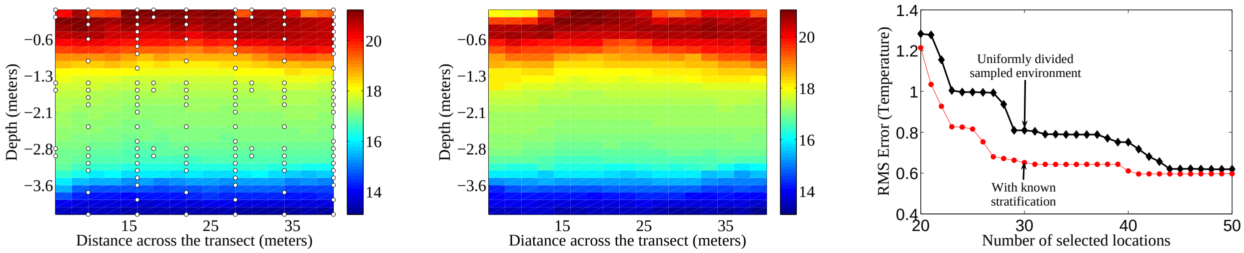
## 6. IMPROVED MODELS USING IDEA

The successful application of IDEA methodology for both the San Joaquin River investigation and the Lake Fulmor campaign resulted in detailed characterization of spatiotemporal variation of different phenomena in the sampled environment. Specific structures related to the spatiotemporal distribution of the phenomena can be extracted using IDEA methodology from any such environmental application. Since similar distributions may not exist in different environments, a single sampling approach suitable for one environment may perform poorly in another environment. However if some information regarding the sampled phenomena distribution is known, then a sampling approach can be biased accordingly to perform high fidelity sensing in these environments.

Next, we discuss how two such approaches, proposed earlier in the literature: (1) Stratified Adaptive Sampling [12]; and (2) Hierarchical Gaussian Process Based Modeling [15]; can perform effectively when prior information regarding the spatial distribution of phenomena is known as an input.

### 6.1 Stratified Adaptive Sampling

Several adaptive sampling approaches have been proposed in literature that hierarchically divide the sampling environment, thus performing dense sampling only in the regions of interest (regions of high phenomena variability) [12, 13]. 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 sampling is performed only in these regions to create high fidelity reconstruction of the sampled environment. If the phenomena shows considerable variability, at a rate faster than the rate at which the first stage scan can be completed, then this approach will result in poor performance. However, if some information about the structure of




(a) Temperature distribution at Lake Fulmor used for simulation for Adaptive Sampling.

(b) Estimation using stratified adaptive sampling

(c) Comparison of GP based approach with known stratification.

Figure 7: Analysis of improvements in Adaptive algorithms using the phenomena distribution information as characterized using IDEA. Points in (a) represent the locations selected by stratified adaptive sampling algorithm.

the phenomena distribution is given as input, then such hierarchical approach can bias the first stage sampling accordingly, resulting in improvement in the results.

Figure 7a displays a typical example of temperature distribution, sampled using NIMS at Lake Fulmor. This field was used as an input distribution to be sampled by the stratified adaptive algorithm, discussed in [12]. When no information was known about the phenomena distribution distribution, the algorithm sampled a total of 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 input (reducing the sampling density along x-axis for the first stage scan and splitting the regions only vertically at each stage instead of 4 equal regions as in the uniform case), the algorithm sampled 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 Figure 7a represent the sampling locations selected by the algorithm when vertical stratification information was known as an input. Figure 7b represent the field distribution estimated by sampling at this subset of locations using local polynomial fitting of second order over the sampled locations.

## 6.2 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) [15]. 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 [24, 25]. 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:

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

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 sensed locations $\mathcal{P}$. Hence mutual information measures the reduction in uncertainty at the unobserved locations.

In particular, we first learned a non-stationary Gaussian Process model, (using an extension of [25]), by maximizing the marginal likelihood [15] using a subset of temperature data at Lake Fulmor. This non-stationary process was learned by dividing the complete region into smaller sub-regions, assuming stationarity within these regions, and learning the required parameters from a set of parameters provided as input. Instead of starting from a random/uniform initial set of parameters to chose from, this was decided based on the spatial distribution of temperature. Then a smooth non-stationary GP is learned from a combination of locally-stationary GPs. The two algorithms compared are when: (1) The sampled environment was divided into smaller sub-spaces uniformly; and (2) The sub-spaces were selected based on horizontal stratification. For each of these algorithms the input set of parameters are still decided based on the observed phenomena distribution. We did not perform comparison with an algorithm where even this input set of parameters is unknown as that is obviously expected to perform poorly compared with these algorithms.

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). Figure 7c compares the root mean square error in prediction after sampling 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 sampled environment. Additionally, any active learning approach, where the GP is learned online based on sampled data will benefit from known information regarding stratification and known information regarding input set of parameters both in terms of accuracy as well as speed of convergence.

# 7. CONCLUSIONS AND FUTURE WORK

Traditional approaches for environmental sampling are not capable of capturing the spatiotemporal dynamics of a complex phenomena 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 sampling 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 sampled phenomena distribution. We demonstrated the effectiveness of the IDEA methodology by providing a detailed characterization of the two markedly different aquatic phenomena. Literature survey reveals that this is the first time an actuated sensing system has been used to perform complete, three-dimensional characterization of aquatic systems on this scale.

The in-depth, three-dimensional field characterization performed using IDEA, can then be used to validate the existing models that now provide a representation of the environment. We also discuss how observations from our experiments validate several underlying physical, chemical and biological attributes. These include: (1) Thermal Stratification in a lake environment; (2) Preferred maximum concentration of phytoplankton in a sub-surface layer; and (3) Mixing effects at the confluence of two rivers showing effectively instantaneous vertical mixing and slow transverse mixing.

We also demonstrated that different environments require characterization of the constraints of the sampling system such that the integrity of the sampled data can be relied upon. We illustrated this characterization through experiments based on in-field observations. These include: (1) Characterizing the static deflection of the suspended pendulous sensor system in an environment with significant flow; and (2) Characterizing the interference, churning of the water while moving along the vertical, y, axis in the water, caused by the sampling system. Observations from such experiments, when included in creating new models for experiment design, result in increased fidelity and reliability in the sampled data.

Finally, we also illustrated how observations from such experiment analysis can be used to improve the performance of existing model-based or adaptive sampling approaches. This is illustrated on two existing approaches: (1) Stratified Adaptive Sampling; and (2) Non-stationary Gaussian Process based approach. We performed simulation using the real temperature data collected by NIMS and presented a comparative analysis of improvements in the performance of these two approaches, given structural information regarding the sampled phenomena distribution. This illustrates that when the detailed characterization of the sampled phenomena distribution is known, using IDEA methodology, then it can be used to improve several existing adaptive sampling approaches.

In the future we plan to investigate model based adaptive approaches, that can be biased towards specific spatial and temporal structure existing in the real environment. 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 can 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 sampled phenomena.

## Acknowledgment

This material is based upon work supported in part by the US National Science Foundation (NSF) under Grants ANI-00331481. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. The authors also acknowledge the support of the team involved in the development of NIMS and the two campaigns discussed in this paper. We will also like to thank Andreas Krause for providing the code for generating non-stationary GPs.

## 8. ADDITIONAL AUTHORS

Additional authors: Victor Chen (Electrical Engineering (University of California, Los Angeles), Mark H Hansen (Department of Statistics, University of California, Los Angeles), Thomas C. Harmon (School of Engineering, University of California, Merced), Gaurav S. Sukhatme (Computer Science, University of Southern California) and William J. Kaiser (Electrical Engineering, University of California, Los Angeles).

## 9. REFERENCES

[1] T. Kratz, L. Deegan, M. Harmon, and W. Lauenroth, “Ecological variability in space and time: Insights gained from the us lter program,” *BioScience*, vol. 53, pp. 57–67(11), 1 January 2003.

[2] 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, March 1999.

[3] 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.

[4] 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.

[5] B. De Serres, A. Roy, P. Biron, and J. Best, “Three-dimensional structure of flow at a confluence of river channels with discordant beds,” *Geomorphology*, vol. 26, pp. 316–335, 1999.

[6] J. V. Reynolds-Fleming, J. G. Fleming, and J. Richard A. Luettich, “Portable autonomous vertical profiler for estuarine applications,” *Estuaries*, vol. 25, pp. 142–147, 2004.

[7] 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.

[8] G. S. Sukhatme, A. Dhariwal, B. Zhang, C. Oberg, B. Stauffer, and D. A. Caron, “The design and development of a wireless robotic networked aquatic microbial observing system,” *Environmental Engineering Science*, 2006.

[9] D. R. Blidberg, “The development of autonomous underwater vehicles (auvs); a brief summary,,” in *IEEE ICRA*, 2001.

[10] 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.

[11] B. Jordan, M. Batalin, and W. Kaiser, “Nims rd: A rapidly deployable cable based robot,” in *IEEE ICRA (Submitted)*, Rome, Italy, May 2007.

[12] M. Rahimi, R. Pon, W. J. Kaiser, G. S. Sukhatmes, D. Estrin, and M. Srivastavan, “Adaptive sampling for environmental robotics,” in *IEEE/RSJ IROS*, New Orleans, LA, April, 2004, pp. 3537–3544.

[13] A. Singh, R. Nowak, and P. Ramanathan, “Active learning for adaptive mobile sensing networks,” in *IPSN,*, Nashville, USA, 2006.

[14] 2006, http://www.jamesreserve.edu.

[15] M. Seeger, “Gaussian processes for machine learning,” *Int. Jour. of Neural Systems*, vol. 14, no. 2, pp. 69–106, 2004.

[16] 2006, http://www.hachenvironmental.com.

[17] 2006, http://www.satlantic.com.

[18] 2006, http://www.sontek.com.

[19] 2006, http://cdec.water.ca.gov/.

[20] N. Quinn, K. Jacobs, C. Chen, and S. W.T., “Elements of a decision support system for real-time management of dissolved oxygen in the san joaquin river deep water ship channel,” *Environmental Modelling and Software*, vol. 20, pp. 1495–1504, 2005.

[21] “Calsim ii simulation of historical swp-cvp operations,” *California Department of Water Resources Technical Memorandum Report*, 2003.

[22] 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 Engineering Science*, 2006.

[23] H. B. Fischer, E. J. List, R. C. Koh, J. Imberger, and N. H. Brooks, *Mixing in Inland and Coastal Waters*. Academic Press, New York, 1979.

[24] C. Guestrin, A. Krause, and A. P. Singh, “Near-optimal sensor placements in gaussian processes,” in *ICML*, 2005, pp. 265–272.

[25] 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)*.