Dynamic Contaminant Identification in Water

Craig C. Douglas
University of Kentucky and Yale University
douglas-craig@cs.yale.edu

In cooperation with
Martin J. Cole, Paul Dostert, Yalchin Efendiev, Richard Ewing, Gundolf Haase, J. Clay Harris, Jay Hatcher, Mohamed Iskandarani, Chris R. Johnson, Raytcho Lazarov, Robert J. Lodder, Steven G. Parker, and Guan Qin.

This work has been supported in part by NSF grants
EIA-0219627, ACI-0324876, CNS-0540178 (U. of Kentucky),
CNS-0540155 (U. of Miami),
EIA-0218229, CNS-0540136 (Texas A&M),
EIA-0218721, CNS-0540153 (U. of Utah).

Goals of Original 3 Year Project

- Develop a variable light wave sensor array to integrate into an ocean observational system.
- Most coastal ocean models are wind driven, but not contamination transport driven: our new model is both.
- We would have done the following:
  - Dynamically inject observed ocean data into multiscale mathematical models and computer simulations.
  - Create research topics in multiscale mathematics, statistics, and software application integration with a flexible, distributed computing database and problem solving environment.

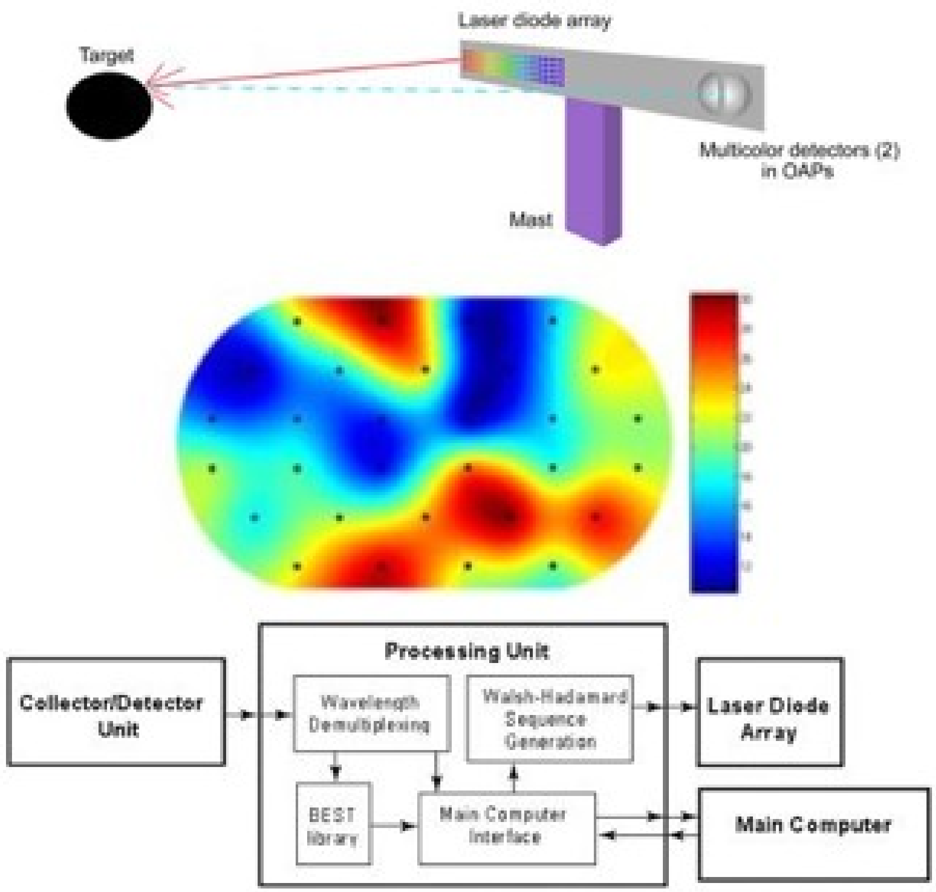
Outline of 1 Year Project

- Develop dynamic component for a sensor prototype so that real sensor will be made later.
- Use intelligence in sensor as model to drive different computing models.
- Develop framework so that we can in the future identify and track contaminants in water (preferably without a person in the loop) near coasts, in bays or rivers, or in reservoirs or lakes.
- Upgrade SCIRun, AggieFEM, and SEOM software components.

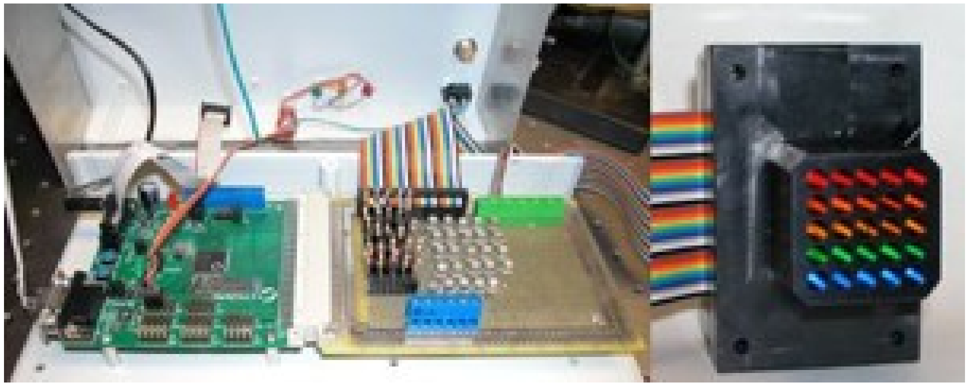
Solid-State Spectral Imager (SSSI)

- A new instrument to gather hydrological and geological data and perform chemical analyses.
- Small, lightweight instrument. Suitable for drones, rovers, etc.
- Uses a laser-diode array, photodetectors, on-board processing.
- Combines near-infrared, visible, and ultraviolet spectroscopy and processing with hyperspace data analysis algorithms. Can add radiation detection, too.
- 25 lasers, discrete wavelengths from 25 nm to 2400 nm, 5 rows per wavelength. Draws 4 watts, weighs 600 g.

SSSI Emiter and Collected Spectrum



SSSI Prototype



What Can the SSSI Identify?

- Virtually every organic compound (e.g., polycyclic aromatic hydrocarbons, paraffins, carboxylic acids, and sulfonic acids) has a near-IR spectrum that can be measured, including two classes of terrestrial biomarkers, lipids, and amino acids.
- Near-infrared spectra consist of overtones and combinations of fundamental mid-infrared bands, giving near-infrared spectra a powerful ability to identify organic compounds while still permitting some penetration of light into samples.
- *Hence, we can build libraries of compound markers that we can download quickly to the SSSI.*

Signal to Noise Enhancements 1: Walsh-Hadamard

- Walsh-Hadamard encoding sequences of light pulses.
  - Multiple laser diodes illuminate a target simultaneously, increasing the number of photons received by the detector.
  - Sequence can be demultiplexed into individual wavelengths with a matrix-vector multiply.
  - Advantages include:
    * Equivalent number of on-off states for each sequence.
    * A constant number of diodes in the on state at each resolution point of a data acquisition period.

Signal to Noise Enhancements 2: CRISP

- Complementary Randomized Integrated Sensing and Processing (CRISP) encoding sequences of light pulses.
  - Uses orthogonal pseudorandom codes with unequal numbers of on-off states.
  - The duty cycle of each code is different, and the codes are selected to deliver the highest duty cycles at the wavelengths where the most light is needed and lowest duty cycle where the least light is needed to make the sum of all of the transmitted (or reflected) light from the samples proportional to the analyte concentration of interest.

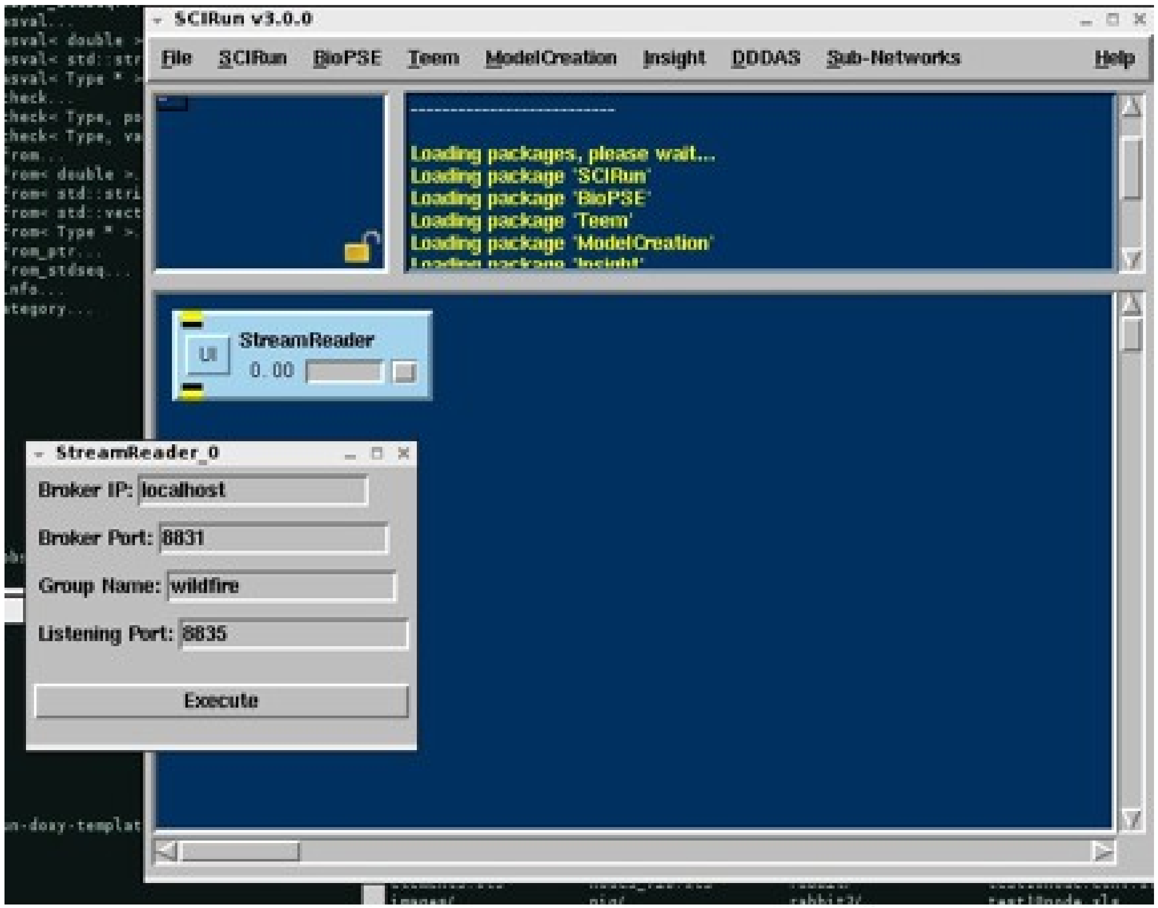
In Field Reprogramming of SSSI

- SSSI observes an interesting set of chemical traces.
- Application monitoring SSSI uploads a different chemical library to SSSI.
- SSSI looks for different chemical traces.
- Application spawns new computational model to work with new SSSI data.

SCIRun Enhancements

- Version 3.0 released by University of Utah in late 2006.
  - Added a telemetry module based on work of Wei Li (Kentucky). Robust, secure Java tools based on a known broker to coordinate sensor data collection and application use. Both GUI and command line interfaces and designed for disaster management in mind and are very easy to use under stressful conditions.
  - Added a Matlab module. Can now pipe data to and from the SSSI allowing easy integrations of sensor networks.
  - Added libgeotiff support for imaging data onto geographic maps.
  - Added higher order and mixed finite element bases.

SCIRun Telemetry Module



Data Handling and Improving

- We can use the data to improve our prediction of the contaminant transport by updating the concentration distribution at some previous time step (considered a time windowed initial condition).

- Update reduces the computational errors associated with incorrect initial data and improves the predictions using the shallow water model.

- Initial data is sought in a finite dimensional space. Using the first set of measurements, the approximation of the initial data is recovered. As new data are incorporated into the simulator, the initial data is updated using an objective function.

Dynamic Data Algorithms

- Initial data is updated using an objective function as new data is incorporated into the simulation.
- Formulated problem is ill-posed: fewer sensors than size of the finite dimensional space describing the initial data.
- Objective function is based on both a measurement error and a penalization term that depends on the prior knowledge about the solution at previous time steps.
- The prior information is refreshed using the updated initial data.

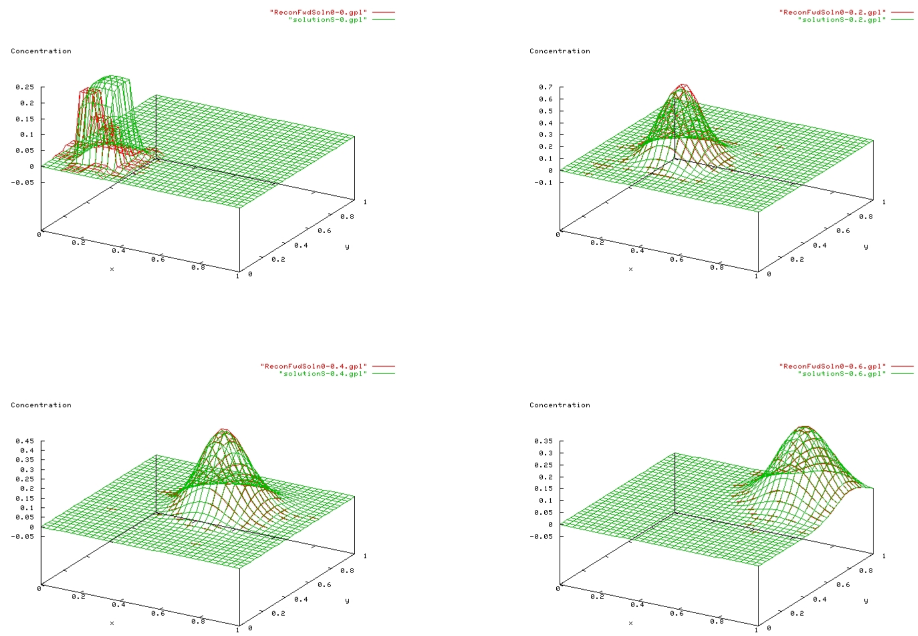
- The penalization constants depend on time of update and can be associated with the relative difference between simulated and measured values. In the simulations, both the prior and penalization constants change in time.

- To account for the errors (uncertainties) associated with sensor measurements, we consider an initial data update within a Bayesian framework.
  - Metropolis-Hasting Markov chain Monte Carlo (MCMC) method to generate samples from the posterior distributions is too expensive.
  - We developed an approach that combines least squares with a Bayesian approach that gives a high acceptance rate. We can prove that rigorous sampling can be achieved by sampling the sensor data from the known distribution.
  - Our approach has similarities with the Ensemble Kalman Filter approach.

Numerical Example

- Consider contaminant transport on a flat surface, a unit dimensionless square, with convective velocity in the direction $(1, 1)$.
- The source term is taken to be 0.25 in $[0.1, 0.3] \times [0.1, 0.3]$ for the time interval from $t = 0$ to $t = 0.05$.
- Initial condition is assumed to have the support over the entire domain.
- The initial condition (solution at previous time step) is derived by solving the original contaminant transport problem with some source terms assuming some prior contaminant history.

Numerical Examle Result



Comparison between reconstructed (red) solution and exact solution at $t = 0.1$ (upper left), $t = 0.2$ (upper right), $t = 0.4$ (lower left), and $t = 0.6$ (lower right).

Results to Date

- Journal articles: 16 published, 12 submitted or accepted.
- Conference proceeding papers: 12 published, 3 submitted or accepted.
- Software packages updated or produced: 4.
- DDDAS Courses: 2 (Kentucky, Texas A&M).
- Technology Transfer: possibly 2 or 3.

Conclusions

- The sensor is ready to move from a prototype toy to a serious LIDAR by adding an expensive laser diode array. Shrinking the overall sensor to the size of a card deck or smaller is feasible.

- Moving from the lab to a real water body is a scaling, security, and privacy research project. We can identify pollutants from a leaking tank on a boat or a sunken vehicle and pollutants in an aquifer and then predict where the interesting chemicals are going or went.

- The software and user interface for using a network of intelligent sensors is in place. Integrating the DDDAS analysis into the ocean modeling code is ready to be done.