1.5 USING ARTIFICIAL INTELLIGENCE TO OPTIMIZE WIRELESS SENSOR NETWORK DEPLOYMENTS FOR SUB-ALPINE BIOGEOCHEMICAL PROCESS STUDIES

Lynette L. Laffea*, Russ Monson and Richard Han
University of Colorado, Boulder, Colorado

John K. Williams
National Center for Atmospheric Research, Boulder, Colorado

1. INTRODUCTION

Researchers in the discipline of biogeochemistry face an enormous challenge as they perform carbon cycle studies related to global climate change. These include quantifying energy and element flows through the earth system and coupling these flows to the dynamic climate system with the goal of devising models that can be used to predict how these flows might change in the future. In facing this challenge, researchers must accommodate spatial and temporal heterogeneity at unprecedented scales and confront non-linearities and intermittency of gas transport that renders many earth system processes intractable for existing approaches. As researchers have embraced these challenges, one reality has emerged clearly: satisfactory sampling of complex biogeochemical systems lies beyond the research community's current observational capabilities (Levin 1992).

For example, current surface measurement systems, such as the Integrated Surface Flux system available for NSF researchers through the National Center for Atmospheric Research Earth Observatory Laboratory (www.eol.ucar.edu), address some aspects of these heterogeneity and scaling questions. Using guyed and freestanding towers and high-quality (in come cases, custom) sensors, EOL surface systems allow investigators to address limited spatial and temporal heterogeneity of energy, mass and momentum fluxes in polar, tropical, and desert environments. However, current observation systems are expensive ($100k per tower) and take considerable time and effort to deploy and maintain. As a consequence, EOL supports only ten stand-alone ground systems and a few instrumented tower levels at any one time, most often for only a single project. Power requirements further constrain systems to locations supplied with line power, which may not be ideally suited to capture the measurements of greatest interest.

To address these constraints, investigators have begun to explore the capabilities of an emerging technology: low cost, battery-powered wireless arrays of environmental and meteorological sensors. Wireless sensors promise researchers a flexible tool for studying biogeochemical processes. Multiple sensors can be combined in an extendable networked array, andmultiple arrays used simultaneously to provide interwoven and cross-ecosystem sensing. Self organization of sensor network communications provides flexibility for ad-hoc deployments, which allow researchers to better ensure that the measurement of interest is adequately captured. Areas where the absence of line power or complex terrain previously made deployments impossible can now be explored.

However, while early explorations—especially using small, inexpensive and uniform sensors in controlled environments—have demonstrated the promise of wireless arrays, only a few groups have confronted the complexity of operations in real environments at the land-atmosphere interface. Not surprisingly, sensor array systems take on a complexity not unlike natural environmental systems, with associated scalability questions. Researchers use the term "embedded'", or even "deeply embedded", to describe transducer interactions with the observed environment. Scalability issues within the network may conflict with the desired ability to capture the measurement of interest.

2. APPROACH

We propose to explore how artificial intelligence techniques can be used to help wireless sensors adequately capture a measurement of interest within the constraints of the sensor network. Our specific application is to investigate carbon fluxes in complex sub-alpine terrain at the Niwot Ridge site near Boulder, CO, and in particular to relate them to environmental and ecosystem measurements (Laffea et al. 2006). We intend to employ a novel strategy based on random forest regression and reinforcement learning for placing sensors and organizing an optimal network topology. Because the sensors' battery power is limited and signal strength varies, algorithms for adaptive measurement and communication that optimize power usage while ensuring that events or measurements of interest are adequately captured are essential to deployments like ours. Additionally, a method that allows researchers to detect areas where additional sensors would be useful or where existing sensors were redundant, would allow the distribution of sensors to be optimized. The techniques we propose to use for these purposes can also be used to discover relationships between sensor data as they evolve in time, aiding in the development or enhancement of models used to describe the physical

* *Corresponding author address:* Lynette L. Laffea, Department of Ecology and Evolutionary Biology (EBIO), University of Colorado, Campus Box 334, Boulder, CO 80309. Email: Lynette.Laffea@colorado.edu.

system being studied. The proposed techniques are being developed using simulated data, but will be utilized in a planned wireless sensor array deployment in the summer of 2007 if they prove practicable.

Our approach consists of two facets: (1) Use a machine learning algorithm, random forest regression, to learn statistical relationships between data from the various sensors. Applied to data from sensors of the same type, these relationships can be used to diagnose regions of poor predictability where additional sensors should be deployed or current sensors should report more frequently to better measure the processes of interest. Used with data from sensors of different types, this approach can be used to discover scientifically interesting nonlinear relationships between different physical processes. (2) Use reinforcement learning to optimize elements of the sensor reporting and network routing strategy by periodically retraining network control parameters based on a history of communications and battery usage. Learning would occur at a base station or central server having adequate processing resources and access to network performance data. The parameters would then sent be to the network nodes to improve future network performance.

Sensor Placement

When monitoring areas of interest, it is important to distribute sensors in such a way that they capture the phenomena of interest without being redundant. In addition, sensors must be placed so that the network is capable of reporting measurements from regions of interest without prematurely exhausting the battery power of sensor or intermediate nodes. We propose to address the question of sensor redundancy by using random forests to predict a sensor's time series based on the timeseries data from other sensors of the same type. If a sensor's measurements can be accurately predicted by the other sensors, it may be judged as potentially redundant. If they are poorly predicted, another sensor may need to be placed nearby to accurately capture the spatial inhomogeneity of the field being measured (see Guestrin et al. 2005). In addition, sensors that are triggered to report or relay data more frequently may require the addition of other sensors or network nodes in the same region to ensure that the data can be reliably communicated without exhausting any sensor node's battery power.

Adaptive Sensor Reporting

Real environments, however, evolve in time both in terms of the observation system and the process being measured. Learning relationships between data from various sensors "on the fly" will allow the identification of significant events or changes in the dominant regime as they occur. These events may require that additional measurements be taken to adequately capture the transitions. We propose again using random forests for this purpose, training new trees in the forest as new data come in and aging off old ones to maintain a robust but adaptive predictive model. If the ability of the random forest to predict or relate the incoming sensor measurement values suddenly falls off, the base station would then signal the sensors to increase their reporting accuracy.

We say reporting "accuracy" instead of reporting "rate" because we envision that the sensor nodes will report in a novel way. Instead of reporting at fixed temporal rates, the sensors will be supplied with a prescribed reporting accuracy, or "tolerance". Recent past measurements will be used to fit a linear or quadratic "trend" to the sensed data, and if a measurement falls outside of the prescribed tolerance from the trend's prediction, a new report will be made. That report will include not the measurement itself, but the time and the parameters of the observed trend. The base station will then be able to provide measurements and error bars for all times based on the reported trends and error tolerances, and will be able to request that a smaller tolerance be used if the situation mandates greater accuracy. The transmission of polynomial fit parameters rather than the data itself have been proposed by Guestrin et al. (2004), and the idea of using tolerances from a trend are akin to standard methods in data compression.

Network Routing

Network routing will be optimized by applying reinforcement learning techniques, with network parameters being optimized periodically (e.g., nightly) based on the network's recent performance. We envision that the network's routing strategy will be stochastic at each node, an appropriate probability distribution over parent nodes being selected at each timestep based on the sensor's state and its knowledge of the state of the network. A candidate method for learning optimal stochastic policies in the context of partially-observable Markov decision processes is described in Williams and Singh (1998).

Relationship Discovery

Finally, another use of random forests is in analyzing the multi-sensor, multi-scale data collected by the sensor array deployment to discover relationships that may be of scientific value. For instance, the purpose of the Niwot Ridge deployment is to determine how various environmental factors are related to carbon flux in a complex alpine ecosystem. Using the random forest by training on the environmental state data to predict the observed CO_2 flux may produce a model which will provide insight into the governing processes and phenomenology. In addition, the random forests are capable of providing lists of the most important variables to the learned relationships; these may prove helpful in determining what physical phenomena are related. An example of using random forests to better understand a complex process, atmospheric turbulence, is described in Williams et al. (2007).

3. SUMMARY

Techniques for placing wireless sensors to adequately capture measurements of interest and dynamically managing their reporting accuracy and network topology to capture significant events while maximizing battery life will become more important as wireless sensor networks continue to enter complex new areas of application, such as the Niwot ridge deployment described in the present paper. Previous sensor array research in this area has focused primarily on theoretical analysis independent of actual network operation and physical process evolution. We believe the artificial intelligence techniques we have described will offer insight into whole system management and process system discovery, while paving the way for a complex sensor deployment that we hope will cast new light on the biogeochemical processes related to global warming.

Note: The latest version of this paper may be obtained from www.rap.ucar.edu/staff/williams/papers/ or by contacting the first author.

4. REFERENCES

Guestrin, C., P. Bodik, R. Thibaux, M. Paskin, and S. Madden, 2004: Distributed regression: an efficient framework for modeling sensor network data. *Information Processing in Sensor Networks, IPSN 2004*, Berkeley, California.

Guestrin, C., A. Krause, and A. Singh, 2005: Near-optimal sensor placements in Gaussian processes, *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany.

Laffea, L. L., R. Monson, R. Manning, R. Han, A. Glasser, S. Oncley, J. Sun, S. Burns, S. Semmer and J. Militzer, 2006: Comprehensive monitoring of CO_2 sequestration in subalpine forest ecosystems and its relation to gobal warming. *4th ACM Conference on Embedded Networked Sensor Systems, SenSys '06*, Boulder, CO.

Levin, S. A., 1992: The problem of pattern and scale in ecology. *Ecology*, **73**, 1943-1967.

Wessman, C. A., 1992: Spatial scales and global change: bridging the gap from plots to GCM grid cells. *Ann. Rev. Ecol. Syst.*, **23**, 175-200.

Williams, J. K. and S. Singh, 1998: Experimental results on learning stochastic memoryless policies for partially observable Markov decision processes. *Advances in Neural Information Processing Systems*, **11**, 1073-1079.

Williams, J. K., J. Craig, A. Cotter, and J. K. Wolff, 2007: A hybrid machine learning and fuzzy logic approach to CIT diagnostic development. *AMS 5th Conference on Artificial Intelligence Applications to Environmental Science*, 1.2.