A new era in the environmental sciences is dawning, in which technological developments and the emergence of scientists with increased levels of training in, and familiarity with, cyberspace promise to revolutionize the way we collect and interpret data. Sensor-based observing systems will form a new and critical resource for environmental sciences, by characterizing such diverse subjects as plankton colonies, endangered species, and soil and air contaminants. Across a wide range of applications, sensor networks promise to reveal new insights by expanding and integrating the scales at which scientists can perceive the natural world, in both space and time (Arzberger et al. 2005; Porter et al. 2005; Hamilton et al. 2007). From studies of nesting success to ocean currents to ecosystem metabolism, previously unobserved phenomena are being detected and ecological theory extended, based on information derived from sensor systems (Porter et al. 2009). The many emerging environmental observatories (eg National Ecological Observatory Network [NEON]; WATer and Environmental Research Systems [WATERS]; Ocean Observatories Initiative) clearly testify to the importance of this new era of sensor observations within the environmental sciences.

Innovation in sensors and sensing networks is advancing as a result of collaborations among engineers, computer scientists, and environmental scientists. The application of these technologies to answer specific scientific questions can uncover unexpected limitations that can motivate further novel technological solutions. Here, we highlight examples from such partnerships to illustrate emerging advances in sensor design and operation, sensor platforms, networking and communications, information management systems, and training programs.

New sensors: changing the way ecologists conduct research

Measurement tools often place severe restrictions on the power and reliability of environmental research. To measure chemical, isotopic, or genetic properties of organisms or the environment, ecologists typically need to physically remove samples from the environment and transfer them to the laboratory – a procedure that often risks altering the very properties they want to measure. Field-based measurements are frequently limited to specific pulses in time or locations where field crews can conduct the measurements. Thanks to partnerships between engineers and ecologists, new sensors are enabling ecologists to conduct more research in situ and to greatly expand...
the temporal and spatial domain of their research. The following are only a few examples.

Moving wet chemistry from the lab bench to the field

Complex “lab bench” chemical analyses are being automated and packaged into field-durable sensors, allowing real-time detection of chemical and genetic properties in the field. For example, the autonomous microbial genosensor (AMG; Figure 1) combines systems for sample collection, filtration, cell lysis, RNA extraction/purification/concentration, isothermal gene amplification by nucleic acid sequence-based amplification (NASBA; Compton 1991; Smith et al. 2007), and data transmission. The prototype AMG under development is configured to detect the “red tide”-causing dinoflagellate Karenia brevis, but its modular design can be adapted to detect any organism for which gene sequence information is known. A conceptually similar environmental sample processor is under development at the Monterey Bay Aquarium Research Institute and has been successfully used to detect marine microbial eukaryotes (Greenfield et al. 2006; Jones et al. 2008).

**Advances in species identification**

A research team composed of scientists from Oregon State University and the University of Washington is developing technology to automate invertebrate biodiversity surveys, by integrating recent advances in computer vision and robotics (Larios et al. 2007). A prototype laboratory system is capable of identifying stonefly larvae in freshwater streams. The research team is now expanding the system to work with soil arthropods and freshwater zooplankton. Also in the developmental phase is a system capable of using wingbeat frequency to differentiate between insect species (eg Belton and Costello 1979; Caprio et al. 2001). Flying insects power their wings using highly coordinated rhythmic contractions, with species-specific wingbeat frequencies (Moore et al. 1986; Moore and Miller 2002). Scientists at the Oregon Health Sciences University, the University of Guam, and APTIV Inc of Portland, Oregon, are developing and field testing a system that enables remote, automated, real-time identification of flying insects, called FAST-ID (Flight Activity Signature Technology for IDentification). The FAST-ID system compares “flight signatures” (wingbeat plus harmonics) for insects flying between a light source and photosensor to a library of signatures, using a statistical classifier or an artificial neural network. The FAST-ID system is being tested at the HJ Andrews Experimental Forest in Oregon.

**Expanding the “sensed domain”: digital eyes, moving platforms, and fiber-optic cables**

The use of cameras for surface-based remote sensing has led to new ways of detecting and quantifying biotic activity. Engineers and biologists at the University of California, Los Angeles’ (UCLA) Center for Embedded Networked Sensing (CENS) are deploying dense networks of imagers within range of prototype wireless platforms (Allen et al. 2008). One system, called “Cyclops”, is a low-power wireless device designed to detect animals in...
pitfall traps and to study avian reproductive behavior at nest sites.

High-resolution digital cameras with pan–tilt controls, mounted on towers and mobile robots, are being used to monitor plant phenology (Graham et al. 2009). The timing of leaf flushes, photosynthesis, and flowering events can be readily quantified through a combination of visible light spectroscopy and pattern recognition (Richardson et al. 2007). The next major challenge will be to develop efficient algorithms that will run on the microprocessor of the digital camera to detect trends and events at the moment that they occur, and even adapt the processing to improve the imager’s ability to classify a pattern (Allen et al. 2008).

Placing sensors on a mobile platform extends the spatial domain of sensor measurements. CENS collaborators have developed fixed, as well as rapidly deployable, cable-based sensor platforms, and unmanned aerial vehicles, which are suitable for a wide range of applications in both terrestrial and aquatic environments (Kaiser and Sukatme 2007). As another example, at the University of Colorado and the National Center for Atmospheric Research in Boulder, Colorado, researchers have developed and deployed a 150-m mobile tram system, embedded within an array of flux towers, with a sensor trolley traveling along the track at a speed of 5 m s$^{-1}$ and measuring a suite of environmental factors, as well as location and altitude (Figure 2). Data signals are transmitted wirelessly to a base station. Using the tram system, researchers have collected relatively high-frequency data, permitting the characterization of horizontal CO$_2$ concentration gradients and fluxes (Oncley et al. 2009). Another approach to extending the spatial domain of measurements uses fiber-optic cables to measure temperature (Selker 2008). With careful calibration, precision approaching 0.01°C can be achieved at a spatial resolution of 1 m, along distances of up to 4 km; new-generation systems have even greater capacity, although there is a trade-off between frequency and precision of measurements. Among many other applications, Distributed Temperature Sensor systems are being used to investigate the dynamics of water flow, impacts on fish habitat in watersheds, and the processes and ecological consequences of air turbulence in airsheds (Selker et al. 2006a, b; Lowry et al. 2007; Westhoff et al. 2007).

**Transcending power limitations**

Although ecology is increasingly being studied in urban and suburban settings, it remains largely a science of natural phenomena in undeveloped ecosystems. The placement of instruments and sensors in remote locations is often restricted by the limited capacity for power storage and/or access; batteries have a limited life before requiring a recharge, and there is often no access to power lines.

**Harvesting remote power**

The introduction of passively, or remotely, powered devices has progressed to the prototype stage. These devices do not require an internal power source, but rather extract their power from propagated radio waves, sunlight, or mechanical vibration. Researchers at Oregon State University are developing an integrated system that allows sensors with unique antennae to scavenge energy from radio frequency (RF) signals transmitted from a central hub (Figure 3). The hub itself is powered by traditional energy sources (eg line power or solar panels). Ultimately, up to 200 sensors will be placed in an array around the hub, using RF waves both for power and communication. The ability for sensors to directly scavenge RF power was hampered for many years by the high power requirements of the scavenging apparatus itself. Recent changes in circuit design, however, have greatly reduced internal power requirements, leading to the development of circuits capable of a net energy gain during the extraction process (Le et al. 2006).

**Power conservation through intelligent sensor aggregation**

In addition to new strategies for generating electrical power, limitations can be overcome through more efficient network design, thereby reducing rates of power consumption. Efficient network designs can be achieved,
by balancing power consumption evenly across the network through the introduction of “network intelligence”. Intelligent networks have the potential to assess and respond to the energy status of individual sensor nodes and evenly distribute the energy required for data collection, processing, and transmission. Intelligence is designed into networks through triggered changes in sensor duty cycles (e.g., sleep/work modes) and the ability to self-organize the network into sensor clusters with shared power burdens. Self-organization means that individual sensor nodes can “ally” themselves either with other nodes with similar power consumption rates or around a central node with better access to power, for the purpose of data transmission. These central nodes can carry more of the transmission energy burden and thus reduce the need for parallel transmission paths from nodes with less access to power. The organization of nodes into common networks can be self-induced, based on the power status of each node, or induced from outside the network via instructions sent by a computer base station (Intanagonwiwat et al. 2003; Cha et al. 2007). The power consumed during data transmission is exponentially proportional to transmission distance (Rajaraman 2002). The use of self-organizing transmission designs can therefore minimize transmission distances and power consumption.

Achieving robust, efficient, and scalable wireless networking

Wireless collection of data from a network of sensor nodes provides great flexibility by eliminating the need to string wire from node to node; however, this can cause a number of problems. Unreliable communication links can drop data at any point in the wireless network. Error protection techniques that improve reliability must be achieved efficiently, given the limited memory and energy lifetime of resource-constrained sensor nodes. A hierarchical wireless networking architecture is emerging that will eventually allow sensor networks to consist of thousands of nodes (Mainwaring et al. 2002; Werner-Allen et al. 2006). For example, researchers at the Universities of Colorado and Montana jointly deployed a hierarchical wireless sensor network to study weather factors surrounding forest fires in the Bitterroot National Forest of Idaho (Hartung et al. 2006). The long-range wireless network consisted of a series of WiFi links with directional antennae, each with a range of up to 50 km. Short-range wireless sensor networks were attached to the long-range wireless network at strategic points of scientific interest (e.g., a chosen mountainside). The short-range sensor networks collected local sensor data and relayed them to the long-range network, which then relayed them to a base camp with a satellite uplink to the internet. Using this hierarchical relay, the wireless network was able to cover large areas in an efficient, targeted manner.

Management and analysis of real-time data

Sensor-based observing systems present new challenges for managing and analyzing data. The large volumes of data produced, and the need to process data in near-real-time, require new information-processing solutions. An instrumented lake buoy – sampling once per minute – generates several megabytes of data per day, whereas video sensors and flux towers produce much greater volumes of data. As sensor networks are scaled to handle hundreds to thousands of sensors across a wide-area network, the need for automation intensifies. Both science and system operation
can require real-time data acquisition and processing. Rapid detection of important patterns within data streams can provide opportunities for efficient system diagnostics and repair, minimizing any loss of measurements. Real-time data-stream processing can also allow adaptive control of measurements in response to environmental conditions (e.g., adjusting sampling rates, instrument activation, or orientation).

A partnership (the “Autoscaling Group”) of computer scientists (from the University of California, San Diego [UCSD], State University of New York–Binghamton, and Indiana University) and ecologists (from the North Temperate Lakes Long Term Ecological Research site, affiliated with the University of Wisconsin–Madison) developed prototypes of several open-source (defined in Panel 1) tools (Figure 4) to automate data processing and accommodate growth and changes in large sensor networks. These tools include (1) a real-time data acquisition system, in which data flow directly from sensors to publicly accessible databases; (2) a common data model for the integration of diverse measurements; and (3) tools for data retrieval, system administration, and status monitoring. These components (described below) are engineered as modules that can be customized for specific applications and integrated with other system components as needed.

**Scalability**

Scalability is a desirable property of a system, indicating that it accommodates well to growth in demands on the system (such as increases in the number of sensor nodes). To address the scalability of real-time data acquisition and processing, the Autoscaling Group adopted a streaming data middleware (Panel 1) approach, in which data from sensors are handled as continuous, real-time streams. Because observing systems incorporate various instruments from several vendors and developers, efficient system management requires integration of these diverse instruments into a cyberinfrastructure (Panel 1) framework. Based on the experiences of the Autoscaling Group, UCSD computer science and engineering researchers collaborated with Creare Inc, the original developers of the DataTurbine streaming data middleware system (www.dataturbine.org). They published the DataTurbine system as open-source software and developed extensions to adapt the system to the requirements of environmental observing systems (Tilak et al. 2007). DataTurbine was enhanced with device interfaces to assorted sensors, dataloggers, and video cameras. It has been used for a wide variety of research purposes, including NASA (National Aeronautics and Space Administration) airborne sensing projects, earthquake simulation experiments, and underwater cameras focused on coral reefs. It has also been tested in lake monitoring applications at several sites.

**Storing and retrieving data**

How sensor data are stored in a database critically affects both the automation of system configuration and the flexibility of data retrieval. The Autoscaling Group has developed a new data model (Panel 1), Vega (Winslow et al. 2008), and the ability to integrate it with streaming data protocols. This data model accommodates reconfigurations of the sensor network, such as the addition of new sensors, without changes to the database schema. The Vega data model was inspired by the Consortium for the Advancement of Hydrologic
Sciences Incorporated (CUAHSI) Data Model (Observation Data Model, ODM; Horsburgh et al. 2008). The data model has been used in an information management system package currently deployed at several sites in the Global Lakes Ecological Observatory Network (GLEON; www.gleon.org) – a grassroots network of limnologists, information technology experts, and engineers who have a common goal of building a scalable, persistent network of lake ecology observatories (Kratz et al. 2006).

**Integration with models in near-real-time**

A software tool developed for GLEON (dbBadger) has been used to run a hydrodynamic model that predicts lake circulation and water temperature distribution (Yuan and Wu 2006; Kimura 2007). In the so-called “nowcasting” mode, real-time wind data drive the hydrodynamic model (Figure 4) that predicts lake circulation. In the prototype nowcasting system, dbBadger handles all database interactions and data handling to retrieve and prepare data for modeling, thereby freeing the modeler from basic data-management tasks.

**Data visualization interface**

A group of collaborators at the James Reserve in California confronted the management and data access challenges for an extensive embedded sensor system by developing an integrated data visualization interface. The James Reserve Google Earth layer (Figure 5) ties together more than 300 in situ sensor nodes (fixed and mobile platforms managing environmental, physiology, chemical, and imaging sensors) into a single visualization interface (Hamilton et al. 2007). Google Earth allows for straightforward integration of a diverse array of sensing systems with web-based data visualization tools in use at the James Reserve (Askay 2006). The interface provides realistic three-dimensional (3D) representations of sensor equipment, current snapshots from the reserve’s various imaging networks, quick access to current and time-series sensor measurement graphs, and the integration of various Geographic Information Systems thematic layers.

**New approaches to training researchers in next-generation technology**

In order for sensor networks to be used effectively to address critical questions in the environmental sciences, we need to train scientists who can work well across disciplinary boundaries and to form new, interdisciplinary collaborations. This need is leading to the introduction of several formal and informal training programs, but we believe the technology is currently expanding more rapidly than is the preparation of the workforce to use this technology.

<table>
<thead>
<tr>
<th>Panel 1. Definitions of selected terms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Open-source software:</strong> software whose license allows unrestricted modification and redistribution, rights that are usually reserved for the author.</td>
</tr>
<tr>
<td><strong>Cyberinfrastructure:</strong> infrastructure based on distributed computer, information, and communication technologies, consisting of enabling hardware, software, communications, institutions, and personnel. The goal is to provide an effective platform for communities of researchers to innovate and eventually revolutionize what they do, how they do it, and who participates (Atkins et al. 2003).</td>
</tr>
<tr>
<td><strong>Streaming data middleware:</strong> a layer of software that moves data from sensors to a publicly accessible database and provides a range of services. For example, DataTurbine, a streaming data middleware system, provides reliable data transport, a framework for integrating heterogeneous instruments, and a comprehensive suite of services for data management, routing, synchronization, monitoring, and visualization.</td>
</tr>
<tr>
<td><strong>Data model or data schema:</strong> a description of the organization of a database.</td>
</tr>
</tbody>
</table>
Innovative programs are emerging in higher education to train environmental scientists and engineers who can then collaborate effectively. For example, at Oregon State University, a National Science Foundation (NSF)-funded IGERT (Integrative Graduate Education, Research, and Training) in Ecosystem Informatics program trains graduate students to apply computer science, mathematics, and sensor networking technology to ecological problems (http://ecoinformatics.oregonstate.edu/). The graduate students also serve as instructors and mentors for undergraduates in a new EcolInformatics Summer Institute – a 10-week program of team-based research that recruits math, computer science, engineering, and environmental science majors from around the US. At UCLA, CENS offers summer training for graduate students and professionals working in sensing technology for the soil environment, in addition to its undergraduate internships for engineering and biology students, geared specifically toward underrepresented groups.

All of the emerging “Environmental Observatory Networks” include an education component (e.g. D’Avanzo et al. 2008). Education and training are also a focus of smaller groups, such as GLEON. Through meetings that address both emerging science and technology issues, web interfaces for sharing expertise, assistance with deployment of sensors and information systems, and opportunities for students to learn skills at the aquatics/information technology interface, GLEON promotes a shared vision of a global network of hundreds of lakes with instrumented buoys collecting real-time sensor data, selected strategically around the globe.

There is also a growing number of training workshops that bring together teams of experts and groups of students for intensive, hands-on learning. These workshops typically provide course credit for students, but they may also include scientists from academia and industry, both as instructors and as students. For example, in the summer of 2008, a two-week course, Flux measurements and advanced modeling, was organized at the University of Colorado’s Mountain Research Station, and is intended to continue in following summers. The course included 22 graduate students from nine countries and invited lecturers from several disciplines. Of particular importance was the participation of Campbell Scientific Inc and LiCor Inc, leading companies in the development of monitoring instrumentation. Representatives of both companies gave lectures and provided the students with opportunities for learning more about advanced sensor and datalogger technologies through explanations of underlying theory and hands-on training in the use of sensor and datalogger instruments.

A further example involves a series of workshops co-sponsored by the NSF and CUAHSI, which featured strong industry participation (Selker 2008). Graduate students, industry representatives, and university faculty representing several disciplines worked in teams to design and implement innovative field research activities involving state-of-the-art sensor technology. This type of workshop provides an exciting and efficient venue for training students and introducing new technology to seasoned scientists. The collaborations inspire the formation of partnerships that provide the basis for ongoing technological developments.

**Conclusions**

The use of automated sensing systems in environmental science is rapidly expanding (Porter et al. 2009). However, realizing the potential of sensor-based observing systems to transform advances in environmental science involves some major challenges. Partnerships of ecologists with computer scientists and engineers are leading to the development of next-generation technology for sensor networks designed to meet these challenges.

Although barriers to communication still exist between environmental scientists, computer scientists, and engineers, these communities are demonstrating that they can work together effectively. Currently, the biggest limitation to these partnerships is funding. Innovative funding mechanisms for research and education are urgently needed, as most traditional funding does not accommodate these multidisciplinary activities. Moreover, it is not sufficient to fund prototype projects at the cutting-edge of technology; the science based on these new sensor networks requires the development of fully mature technology products. Maximum benefit will come through funding the development of widely usable tools; local collaborations in system design have an important role to play, but we need to move beyond “one-off” solutions.

Often throughout the history of science, new tools advance more rapidly than the science that uses them. The resulting scales and volumes of data will generate novel scientific questions and models. We need to train the next generation of scientists to conduct research in this expanded information environment.

**Acknowledgments**

T Fiez, T Dietterich, T Fountain, S Tilak, T McMahon, M Smith, S Oneley, J Jones, D Tullow, J Selker, P Kirsch, and L Winslow provided insights and information. We thank our collaborators in the projects described in this paper, as well as B Feeny and A Kennedy for assistance in developing figures. We gratefully acknowledge our sponsors, including NSF, the Natural Reserve System, industrial partners, the Gordon and Betty Moore Foundation, and the authors’ institutions. In particular, we want to acknowledge our intrepid field engineers, programmers, and hardware developers. This research is supported by grants from the NSF, including CENS STC (#0120778), NIMS ITR (#0331481), AMARSS biocomplexity (#0410408), DBI-05292230, CNS/SENSORS-0330466, Andrews LTER (#0904115), North Temperate Lakes...
LTER DEB-0217533, DBI-0843037, NEON 0446802, NEON 0446017, DBI-0639229, and DBI/SST-0528793.

References


