New opportunities in ecological sensing using wireless sensor networks

Scott L Collins[1], Luís MA Bettencourt[2], Aric Hagberg[2], Renee F Brown[1], DI Moore[1], Greg Bonito[1,4], Kevin A Delin[3,5], Shannon P Jackson[3], David W Johnson[3], Scott C Burleigh[3], Richard R Woodrow[3], and J Michael McAuley[3]

Measuring environmental variables at appropriate temporal and spatial scales remains an important challenge in ecological research. New developments in wireless sensors and sensor networks will free ecologists from a wired world and revolutionize our ability to study ecological systems at relevant scales. In addition, sensor networks can analyze and manipulate the data they collect, thereby moving data processing from the end user to the sensor network itself. Such embedded processing will allow sensor networks to perform data analysis procedures, identify outlier data, alter sampling regimes, and ultimately control experimental infrastructure. We illustrate this capability using a wireless sensor network, the Sensor Web, in a study of microclimate variation under shrubs in the Chihuahuan Desert. Using Sensor Web data, we propose simple analytical protocols for assessing data quality "on-the-fly" that can be programmed into sensor networks. The ecological community can influence the evolution of environmental sensor networks by working across disciplines to infuse new ideas into sensor network development.

Front Ecol Environ 2006; 4(8): 402–407

Ecologists struggle to measure complex environmental variables that change rapidly in space and time. To date, environmental monitoring and measurement have been limited by methodology, particularly the types of field-based sensors available to ecologists, their costs, and the constraints imposed by the need to physically wire sensors to stationary data loggers. These limitations usually lead to suboptimal placement of a few sensors within reach of data loggers, rather than in locations that optimize measurement of the variable of interest. Small, inexpensive wireless sensors (eg Johnston *et al.* 2004) coupled with the widespread availability of low-cost wireless data transmission infrastructure (eg Peterson *et al.* 1995; Atkins *et al.* 2003) will free us from a wired world and revolutionize our ability to measure environmental variables at appropriate spatial and temporal scales (Porter *et al.* 2005; Hart and Martinez in press).

Although ecologists are increasingly aware of the power of sensor networks and are involved in the development of new environmental sensors (Palmer *et al.* 2004), discussion thus far has focused on how such networks will increase our ability to gather data at spatial and temporal scales appropriate for understanding regional ecological phenomena (Porter *et al.* 2005). Although these sensor networks will revolutionize our ability to measure and monitor the environment, they do pose huge data analysis and management challenges that will be new to most ecologists. To date, in the ecological literature, there has been little explanation of the capability of sensor networks to not only collect vast amounts of data, but also to process, analyze, summarize, and interpret such data in both passive ecological monitoring contexts and active experimental settings.

One approach to this data richness problem is to conceptually reduce a sensor network and associated cyberinfrastructure to three simplified components that are common to many experimental setups: (1) the sensor, which is measurement specific; (2) a sensor network that gathers and transmits sensor data; and (3) the end user who analyzes and interprets these data with a particular question in mind. All subcomponents are linked by cyberinfrastructure, including hardwire transmission networks (eg the Internet), computers, data archives, and analytical and graphical software. In this stylized deconstruction, the sensor and the user are problem specific, whereas the sensor network can be generalized across different applications. However, sensor networks have the potential to integrate these components in useful and novel ways. Although we traditionally think of data processing as occurring at the user end of the transmission sequence, many sensor networks have the capacity for embedded computing, an important capability that should be exploited by ecologists (Delin and Jackson 2000; Estrin *et al.* 2003). Taking advantage of this technology may therefore require shifts in experimental design towards distributed and more real-time data screening and analysis, and

[1]*Department of Biology, University of New Mexico, Albuquerque, NM 87131;* [2]*Mathematical Modeling and Analysis, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545;* [3]*NASA/Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109-8099;* [4]*Present address: Department of Biology, Duke University, Durham, NC 27708;* [5]*Present address: SensorWare Systems Inc, Arcadia, CA 91007*

Figure 1. *(a) Desert grassland vegetation with scattered creosotebush (*Larrea tridentate*), showing the distribution of vegetation, shrub islands, and bare soils characteristic of aridland ecosystems. (b) Sensor Web v3.1 pod underneath a juniper tree (see http://sev.lternet.edu/research/SWEETS/index.html for a site map and access to Sensor Web data).*

ultimately adaptive experimentation (Cook *et al.* 2005), hypothesis formulation, and testing. In short, the network itself produces information in a usable form from data (Delin 2002; Delin *et al.* 2005).

In this article, we illustrate the potential ability of sensor networks to function beyond the acquisition of large, complex data streams. To do so, we briefly present results from an ongoing experiment at the Sevilleta Long-term Ecological Research (LTER) site that uses a wireless sensor network, the Sensor Web, to measure microclimate beneath different species of native desert shrubs. We then highlight the fundamentals of Sensor Web technology and describe how the sensor network itself can be used for data quality assurance and quality control, data manipulation, and eventually actuation – the explicit coordinated and distributed control of experimental infrastructure based on in-situ data processing.

Islands of fertility

Aridland ecosystems worldwide are undergoing dramatic changes in response to a variety of environmental drivers (Archer *et al.* 1995; Fenn *et al.* 2003). One consequence of these pressures in many semiarid regions worldwide is desertification and degradation, including conversion of C_4-dominated grassland to C_3-dominated shrub- and woodland environments (Geist and Lambin 2004). Desertification has substantial ecological consequences, including altered surface and subsurface hydrology, reduced biodiversity, diminished capacity to retain nutrients, altered carbon storage capacity, and altered soil resource heterogeneity (Jackson *et al.* 2002; Briggs *et al.* 2005), as resources are increasingly concentrated in "islands of fertility" beneath shrub canopies.

Aridland plant communities are characterized by relatively distinct patches of vegetation with intervening bare areas of soil (Peters *et al.* 2006; Figure 1a). The original island of fertility model focused on how the local distribution of soil resources changed, from relatively uniform to increasingly concentrated beneath plant canopies (Schlesinger *et al.* 1990). In fact, soil resources are much higher beneath grass and shrub canopies, as compared to bare soil patches at the Sevilleta (Keift *et al.* 1998). Shrub encroachment not only alters the distribution of soil resources but may also affect local microclimate. At the Sevilleta we asked the question, "Are all islands of fertility equal?" We were particularly interested in determining how microclimate differed beneath three common native shrub species that have increased in abundance locally and regionally: the semi-evergreen creosotebush (*Larrea tridentate*); honey mesquite (*Prosopis glandulosa var torreyana*), a small deciduous tree; and one-seeded juniper (*Juniperus monosperma*), an evergreen shrub. Although this is a relatively straightforward experiment, we chose this design to learn more about how different species modify their local environments, and to assess the longevity and durability of an environmental sensor network in a relatively harsh environment. In the process, we gathered extensive data streams that can be used as test beds for embedded data harvesting algorithms and estimation of data error rates within a long-running sensor network, in this case a technology called the Sensor Web (Delin 2002), developed at NASA's Jet Propulsion Laboratory.

Sensor Web

The Sensor Web (www.sensorwaresystems.com) is a flexible network of spatially distributed sensor platforms (pods) that wirelessly communicate with each other (Figure 1b). Advantages include synchronous measurements of environmental variables and a unique data transmission protocol that allows every pod to share data with every other pod throughout the network at each measurement cycle. This data sharing protocol provides an ideal opportunity for embedded data processing within

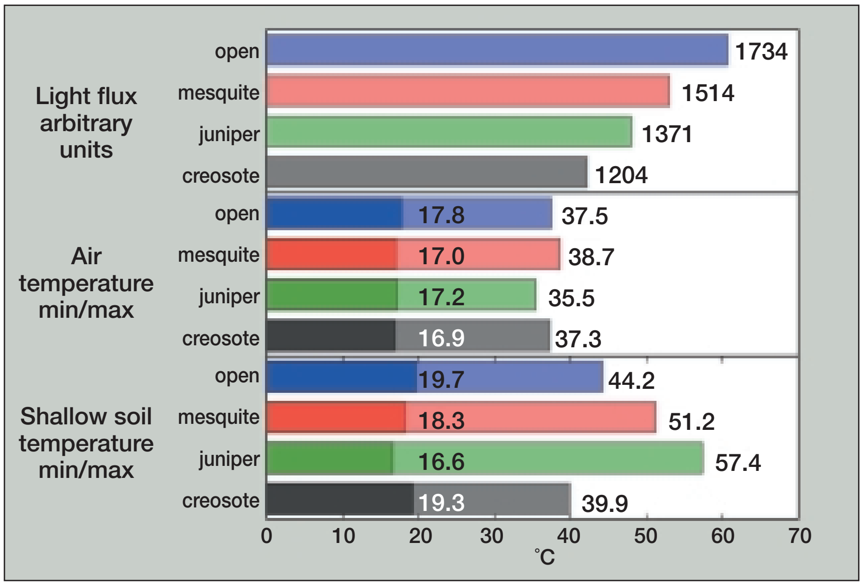

***Figure 2.** Average daily ranges (minimum, maximum) of light flux, air temperature, and shallow soil temperature under creosotebush (*Larrea tridentate*), honey mesquite (*Prosopis glandulosa *var* torreyana*), and juniper (*Juniperus monosperma*), and bare soil in July 2004, at the Sevilleta LTER site in New Mexico.*

the sensor network itself. Because pods can be located as far as 100 m or more apart, sensor networks can greatly expand the spatial extent of any experimental context, allowing a more flexible statistical design than if sensor placement is limited by wired links to data loggers. In the present case, we were able to measure several microenvironmental variables under individual shrubs that were widely scattered along a 300 m transect. Cost-effective sensor networks, such as the Sensor Web, thus provide new avenues for research design, data collection, and analysis.

To measure microclimate variation at the Sevilleta, in October 2003 we placed three Sensor Web pods (v3.1) each in randomly selected open areas, and under the east side canopy of three individuals of each shrub species (12 pods in total), arrayed along a 300 m transect. An additional pod served as a data relay, while a 14th pod served as the mother pod, which is connected to a laptop that contains the database and serves as a portal into the system via the Internet. Sensors on each pod measure soil temperature at 1 and 10 cm depths, soil moisture at 10 cm depth, relative humidity, air temperature, and light at 5 minute intervals. The pods are powered by solar-assisted batteries, providing virtually unlimited field life. Sensor Web data from the Sevilleta are available at http://sev.lternet.edu/research/SWEETS/index.html.Example data streams are shown in WebFigure 1a and b.

Microclimate in islands of fertility

A complete analysis of microclimate differences between shrub species and open areas is not possible or appropriate here. Rather, we provide an analysis of three mid-summer microclimate variables (Figure 2) and the potential for using sensor networks in ecological research. Clear differences in soil temperature at two depths and light availability occur between shrub canopies and open areas. In both winter and summer, daily temperature oscillations in bare areas are greater than under creosotebush. Surprisingly, maximum soil temperatures beneath some shrubs are actually higher than in open areas, a result of differences in soil albedo within different microenvironments; beneath shrubs, the soil is covered by organic matter, which darkens the surface and increases heat absorption, particularly during mid-summer. As a result, during July the average daily maximum temperature in the shallow soils under juniper was significantly higher than in open areas or under the canopy of creosotebush, and to a lesser extent under mesquite (Figure 3). Nocturnal temperature minima were slightly lower under shrubs than in open areas, and maximum daily light levels were lowest under creosotebush. Clearly, not all resource islands are equal, which probably has implications for the distribution and abundance of plant and animal species associated with resource islands in aridland environments.

Embedded processing

Post hoc analyses such as our examination of shrub microclimates depend on the quality of the data generated by sensors and sensor networks. Indeed, the deployment of larger and more complex sensor networks will yield huge and ever-growing datasets, which will increase the need to automatically screen, check, and process data quickly and efficiently. These challenges suggest the need for analytical solutions that automatically perform statistical analyses concurrently. This, in turn, will provide the opportunity for validation of ecological hypotheses in real time by the instrument itself, shifting the burden of data analysis and its logistical costs and delays away from the researcher and onto the sensor network. For example, because it is a temporally synchronized, spatially distributed network, the Sensor Web allows for analytical procedures to be easily programmed into the network, so that data quality can be assessed during every measurement interval and data summaries can be generated at any desired measurement interval.

In the future, large sensor networks will measure multiple environmental variables at short time intervals and operate over large areas for years. Although the Sensor Web cluster deployed at the Sevilleta LTER has only a modest number of pods, research platforms, such as the

National Ecological Observatory Network, envision sensor networks with hundreds or even thousands of sensors. As sensor networks continue to offer better spatial coverage and include remote areas, problems with data quality assessment, storage, retrieval, and manipulation will increase quickly, outstripping traditional human resources dedicated to offline analysis. Shifting portions of data analysis from the user to the network itself will therefore no longer be simply a matter of convenience, but a very practical necessity (Delin and Jackson 2000; Delin 2002; Larkey *et al.* unpublished).

Figure 3. *Photograph of a warming apparatus in a new nighttime warming, winter rainfall, N-deposition experiment at the Sevilleta LTER site in New Mexico.*

A simple first step in data processing is to identify and eliminate erroneous sensor readings; these may occur for many reasons, including the occasional corrupted sensor measurement or data transmission error. Even if data error rates are very small (eg < 0.001%), large sensor networks with frequent data acquisition protocols will generate potentially hundreds or thousands of error measurements annually. For example, the 12-pod Sensor Web array at the Sevilleta LTER makes six environmental measurements at each node every 5 minutes, yielding more than 12.6 million data points annually. Even if data error rates are exceptionally low, eg 1 in 10 000, then about 1260 data values will be potentially erroneous each year, and the number of errors will increase with the size and complexity of the sensor network. Although small, such error frequencies could potentially affect overall data quality, thereby possibly creating erroneous action in autonomous systems, as well as reducing the reliability of comparative analyses.

Data analysis design choices must be balanced between the need for data quality assurance and constraints imposed by limited memory and processing capabilities within the network. Automatic procedures to assess environmental data quality, for example, pose particular challenges for standard statistical treatments because most environmental data, such as air temperature or soil water content, are non-stationary, changing in quasi-periodic but unpredictable ways over daily and seasonal cycles. Nevertheless, simple and practical ways to deal with such variations are being developed (Panel 1) that exploit the spatially distributed nature of sensor networks. These procedures rely on explicit features of data collection protocols, including effectively synchronous measurements and local data sharing over sets of nearby sensors within the network. If these criteria are met, then the large data streams generated by a sensor network become an asset, for use in statistically based data quality assessment protocols. Moreover, these properties permit the use of a data quality assurance scheme that is simple enough to be embedded within the sensor network itself (see Panel 1). That is, based on real-time data streams, the network itself can assess whether data values fall within some predetermined level of statistical confidence and, if not, these values can be excluded and identified as errors. In addition, missing values can be inferred using similar algorithms (Larkey *et al.* unpublished). In practice, quality assurance algorithms can be employed for each measured quantity within a subset of neighboring sensors, or for averages of multiple sensors in comparable habitats. In this way, sensor networks can be configured to produce knowledge from the raw data, rather than just providing a passive stream of data to the end user.

Strategies for data quality assurance, as illustrated in Panel 1, represent a vital first step in data processing that yields a constant statistical summary, which can be reported and stored in a permanent database along with the raw data. Because most natural data quality assurance strategies rely on the statistical comparison of measurements at different sensors, they also produce immediate statistical comparisons of data among contrasting environments or time intervals, or between treatment and control areas in ecological experiments. In this way, statistical syntheses of important ecological variables are naturally generated in tandem with ensuring data quality. Shifting data analysis to the network allows the sensor network not only to detect naturally anomalous events corresponding to errors, but also to identify other infrequent but important environmental dynamics, such as rainfall pulses in arid environments, which lead to rapid, spatially coherent, and time-correlated changes in multi-

Panel 1. In situ error detection in distributed sensor networks

In any experiment, errors in sensor measurement or data transmission occur occasionally. In many cases erroneous readings fall within the normal range of daily or seasonal variation and may prove difficult to identify. However, these rare error values may be large enough to affect statistical inferences. Figure 4a shows a large, rare data transmission error from a Sensor Web pod at the Sevilleta LTER. It is possible to embed algorithms in a pod's processing unit that compare data values among sensors, giving a basis for error detection and for inferring missing readings. As a consequence, outlier values can be detected and flagged when they occur and, in the process, Sensor Web data are analyzed and summarized "on-the-fly".

To do this efficiently, and to accommodate the fact that average values change throughout the day and across seasons, we estimate the probability distribution of differences between a given quantity (eg air temperature) measured at adjacent pods P(ΔT; see Figure 4c). Measurement errors are identified as point failures that occur with a small probability and typically correspond to large and sudden temperature differences of tens of degrees or more between adjacent sensors. These can be identified and eliminated at a chosen level of confidence, *C*, (eg 99%) by standard statistical tests. By adjusting *C*, it is possible to fine tune the level at which data values are considered to be potential errors. Using inferred probability distributions from data, we determine the probability of observing a difference in measurements between the pod in question and its neighbors that is larger (in absolute value) than the value observed. If this total probability is less than *C* the measurement is classified as anomalous; otherwise, the datum is accepted and stored in the database. As new data values are accepted they can be used concurrently, to update the probability distribution of valid values. Missing readings at a sensor might also be inferred through knowledge of those of its neighboring pods and of the statistical distribution of their differences (Larkey *et al.* unpublished).

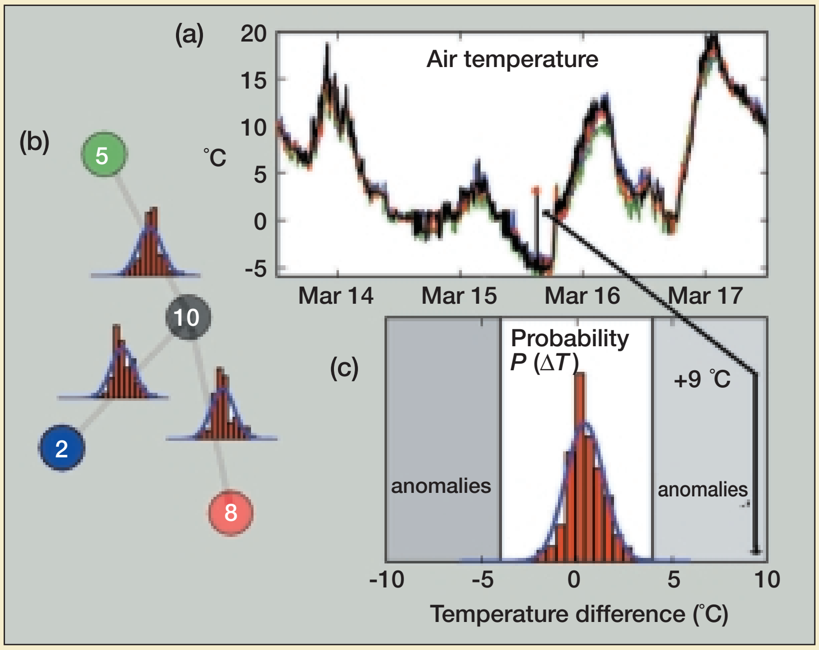

Figure 4. *A data transmission anomaly (a; red circle) in the air temperature measurements at one of the Sensor Web pods can be identified and eliminated via comparison to the measurements transmitted by nearest-neighbor pods (b). The anomalous reading is identified as an outlier (c) in the distribution of differences between the readings at a pod and those of its neighbors at a predetermined level of statistical confidence (Larkey* et al. *unpublished).*

ple environmental variables (Potts *et al.* 2006). Additional algorithms based on strategies for anomaly detection, similar to those used here for quality assurance, can then be implemented to increase or decrease sampling frequency in response to environmental triggers. Such adaptive sampling algorithms can be used to reduce data collection rates at times when little change is occurring and then to rapidly increase them whenever it becomes necessary to capture changes in environmental conditions at high spatial and temporal resolution. It will also be possible to autonomously trigger sampling whenever conditions are judged to be sufficiently interesting by the network (Delin 2002).

One of the least discussed aspects of sensor networks in the ecological literature is the potential for coordinated, distributed control and actuation – the use of autonomous determination of physical changes in the environment to control experimental infrastructure. For example, at the Sevilleta LTER we recently established a multifactor environmental change experiment in which we manipulate nighttime temperatures, winter rainfall, and nitrogen deposition to determine their individual and combined effects on creosotebush encroachment into grassland. Our nocturnal warming treatment is applied by using lightweight aluminum fabric shelters mounted on rollers; as with a window shade, these are drawn across the plots each night to reduce heat loss and elevate nighttime air temperatures. The shelters roll up again each morning. We plan to use the error detection and data summary algorithms described above in this experiment, to summarize nightly treatment effects and generate statistical summaries of outlier values that we can use to detect and quickly repair deployment failures. Ultimately, we plan to use embedded processing to calculate statistical differences among treatments and eventually to develop in situ algorithms to deploy the warming apparatus and to automate the winter rainfall treatments.

Outlook

Wireless sensor networks have tremendous potential in environmental research. Ecologists are becoming increasingly aware of the capability of these networks to collect multiple point measures of ecological variables at high temporal frequencies across vast spatial scales. The use of inexpensive, long-lasting sensor networks will increase our ability to conduct research at scales relevant to some environmental grand challenges (NRC 2003). Yet environmental sensor networks offer a far greater potential than simply switching from a wired to a wireless world. Wireless sensor networks can be programmed to assess data quality, modify sampling regimes, and ultimately activate ecological infrastructure. The optimal use of such sensor networks will require a multi-disciplinary effort, including ecologists, engineers, computer scientists, and statisticians, to take full advantage of a technology that is likely to revolutionize not only data collection, but also data processing, analysis, and manipulation of experimental infrastructure. Because sensor network technology is still maturing, the ecological community is in a unique position to influence the growth of this technology by working across disciplines to infuse new ideas into wireless sensor network development.

Acknowledgements

We thank J Johnson and K Eack for helpful comments and discussions. This research was supported by NSF grant DEB-0217774 for Long-term Ecological Research. The Sensor Web was constructed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA, by support from the Technical Support Working Group with funding from the Department of Defense, Department of Homeland Security, and Department of State.

References

Archer SR, Schimel DS, and Holland EH. 1995. Mechanisms of shrubland expansion: land use, climate or CO_2? *Climatic Change* **29**: 91–99.

Atkins DE, Droegemeier KK, Feldman S, *et al.* 2003. Revolutionizing science and engineering through cyberinfrastructure: report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. www.nsf.gov/cise/sci/reports/atkins.pdf. Viewed 28 August 2006.

Briggs JM, Knapp AK, Blair JM, *et al.* 2005. An ecosystem in transition: causes and consequences of the conversion of mesic grassland to shrubland. *BioScience* **55**: 243–54.

Cook WM, Casagrande DG, Hope D, *et al.* 2004. Learning to roll with the punches: adaptive experimentation in human-dominated systems. *Front Ecol Environ* **2**: 467–74.

Delin KA. 2002. The Sensor Web: a macro-instrument for coordinated sensing. *Sensors* **2**: 270–85.

Delin KA and Jackson SP. 2000. Sensor Web for in situ exploration of gaseous biosignatures. *IEEE Aerospace Conf Proc* **7**: 465–72.

Delin KA, Jackson SP, Johnson DW, *et al.* 2005. Environmental studies with the Sensor Web: principles and practice. *Sensors* **5**: 103–17.

Estrin D, Michener W, and Bonito G. 2003. Environmental cyberinfrastructure needs for distributed sensor networks: a report from a National Science Foundation Sponsored Workshop, Scripps Institute of Oceanography. www.lternet.edu/sensor_report. Viewed 28 August 2006.

Fenn ME, Baron JS, Allen EB, *et al.* 2003. Ecological effects of nitrogen deposition in the western United States. *BioScience* **53**: 404–20.

Geist HJ and Lambin EF. 2004. Dynamic causal patterns of desertification. *BioScience* **54**: 817–29.

Hart JK and Martinez K. Environmental sensor networks: a revolution in the Earth system science. *Earth-Sci Rev* **77**. In press.

Jackson RB, Banner JL, Jobbagy EG, *et al.* 2002. Ecosystem carbon loss with woody plant invasion of grasslands. *Nature* **418**: 593–94.

Johnston CA, Groffman P, Breshears DD, *et al.* 2004. Carbon cycling in soil. *Front Ecol Environ* **2**: 522–28.

Kieft TL, White CS, Loftin SR, *et al.* 1998. Temporal dynamics in soil carbon and nitrogen resources at a grassland–shrubland ecotone. *Ecology* **79**: 671–83.

Larkey L, Bettencourt LMA, and Hagberg A. In situ data quality assurance for environmental applications of wireless sensor networks. ACM Transactions on Sensor Networks. Unpublished.

National Research Council. 2003. NEON: addressing the nation's environmental challenges. Washington, DC: National Academy Press.

Palmer MA, Bernhardt ES, Chornesky EA, *et al.* 2005. Ecology for the 21st century: an action plan. *Front Ecol Environ* **3**: 4–11.

Peters DPC, Gosz JR, Pockman WT, *et al.* 2006. Integrating patch and boundary dynamics to understand and predict biotic transitions at multiple spatial scales. *Landscape Ecol* **21**: 19–33.

Peterson RL, Ziemer RE, and Borth DE. 1995. Introduction to spread-spectrum communications. Englewood Cliffs, NJ: Prentice Hall.

Porter J, Arzberger P, Braun H-W, *et al.* 2005. Wireless sensor networks for ecology. *BioScience* **55**: 561–72.

Potts DL, Huxman TE, Enquist BJ, *et al.* 2006. Resilience and resistance of ecosystem functional response to a precipitation pulse in a semi-arid grassland. *J Ecol* **94**: 23–30.

Schlesinger WH, Reynolds JF, Cunningham GL, *et al.* 1990. Biological feedbacks in global desertification. *Science* **247**: 1043–48.