Life Under your Feet: A Wireless Soil Ecology Sensor Network

Răzvan Musăloiu-E., Andreas Terzis*, Katalin Szlavecz†, Alex Szalay, Joshua Cogan‡, Jim Gray+
Department of Earth and Planetary Sciences †
Department of Physics ‡ , Computer Science Department *, The Johns Hopkins University,
Microsoft Research+

Abstract

Wireless sensor networks can revolutionize soil ecology by providing measurements at temporal and spatial granularities previously impossible. This paper presents an experimental soil monitoring network that we developed and deployed in a Baltimore urban forest as a first step towards realizing this vision. Each network node measures soil moisture and temperature every minute and stores the measurements in local memory. Raw measurements are periodically retrieved by a sensor gateway and inserted to a database in which a calibrated version is also derived and stored.

At a high level, this was a successful test deployment exposing high level variations of soil factors. However, we encountered a number of challenging technical problems: need for low-level programming at multiple levels, calibration across space and time, and cross-reference of measurements with external sources. These problems must be addressed before sensor networks can fulfill their potential as experimental instruments that can be deployed by scientists without major effort or cost.

1 Introduction

Lack of field measurements, collected over long periods and at biologically significant spatial granularity, hinders scientific understanding of the effects of environmental conditions to the soil ecosystem. The recent emergence of Wireless Sensor Networks (WSNs) promises to address the ecologists' predicament through a fountain of measurements from low-cost wireless sensors deployed with minimal disturbance to the monitored site.

During the Fall of 2005 we set out to evaluate the validity of this claim through a proof-of-concept WSN we built and deployed at an urban forest. The end-to-end system we developed includes motes that collect environmental parameters such as soil moisture and temperature, static and mobile gateways that monitor the health of the network and periodically download collected measurements through a reliable transfer protocol, a database that stores collected measurements, access tools that allow us to analyze the data, a web site that serves the data and tools to the Internet so that others can use the data, and tools to monitor the network.

The notable aspects of our system are: **(1)** Unlike previous WSNs, *all* the measurements are temporarily stored on each mote's local flash and are periodically retrieved with a reliable transfer protocol [2]. **(2)** We implemented sophisticated calibration techniques that translate raw sensor measurements to high quality scientific data. **(3)** The database and WSN are accessible via the Internet, providing access to the collected data through graphical and Web Services interfaces.

We acknowledge that this system is only one link in the long chain of steps from collecting raw measurements to scientifically important results, but it shows great promise in improving ecology science and also ecologist's productivity. However, today the project has one ecologist and several supporting computer scientists, a ratio we are working to reverse.

2 The Need for Monitoring in Soil Ecology

Soil is the most spatially complex stratum of a terrestrial ecosystem, harboring an enormous variety of plants, microorganisms, invertebrates, and vertebrates. These organisms are not passive inhabitants of the soil; their movement and feeding activities significantly influence the soil's physical and chemical properties. Because soil is an important water reservoir in terrestrial ecosystems and, thus, an important component for surface and groundwater hydrology models [3], interest in the behavior of soil biota spans multiple scientific disciplines.

It has been observed that soil organisms are patchily distributed in all three dimensions. Such variations can be either due to biological mechanisms or they are the result of differences in the physical environment, because many soil invertebrates are sensitive to such abiotic factors as soil moisture, temperature and light. For this reason, any field study on soil biota requires soil temperature, soil moisture, and other physical measurements.

These data are usually collected by a technician visiting the field site once a week, month, or season, and taking a few spatial measurements that would be subsequently averaged. These techniques are labor-intensive, and do not capture the underlying spatial and temporal variations at a biologically meaningful scale. Moreover, frequent visits to a site disturb the habitat and may distort the results.

2.1 Requirements

WSNs promise inexpensive, hands-free, low-cost and low-impact data collection – an attractive alternative to manual data logging, in addition to providing considerably richer data. However, to be of scientific value, the data collection system design should be driven by the experiment's requirements, rather than by technology limitations. Following this principle, we present a list of key requirements that soil ecology sensor networks must satisfy:

Measurement Fidelity: *All* the raw measurements should be collected and persistently stored. Should the scientist later decide to analyze the data in a different way, to compare it to another dataset, or to look for discrepancies and outliers, the original data must be available. Furthermore, given the communal nature of field measurement locations, other scientists might use the data in the future in ways unforeseen at the time when the original measurements were

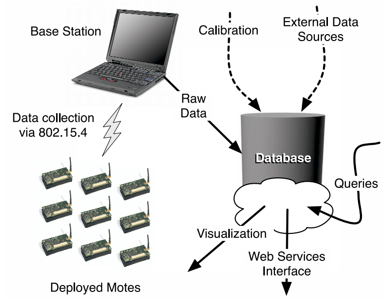

Figure 1: Architecture of the end-to-end data collection system.

taken. Generally speaking, *techniques that distill measurements for a specific purpose, potentially discard data that are important for future studies.*

Measurement Accuracy and Precision: To support ecological research, temperature data should have accuracy of at least 0.5°C, and volumetric moisture data should be given within 1%. While temperature variation of half a degree does not directly affect soil animal activity, soil respiration exponentially increases with temperature, so half a degree makes a big difference. Therefore, *raw measurements need to be precisely calibrated, to give scientists high confidence that measured variations reflect changes in the underlying processes rather than random noise, systematic errors or drift.*

Fusion with External Sources: *Comparing measurements with external data sources is crucial*. For instance, soil moisture and temperature measurements must be correlated with air temperature, humidity, and precipitation data. Animal activity is determined by these factors as much as by soil temperature and moisture. In the case of hydrology models, one can only make sense of soil moisture if precipitation data is available. In addition to "traditional" external data sources, data from other WSNs can be integrated with the results collected from the local WSN. For this reason, collected data should be exported using a controlled vocabulary and well defined schemas and formats.

3 System Architecture

3.1 Data Collection Subsystem

The data collection subsystem includes the motes and the base station illustrated in Figure 1. We use MicaZ motes from Crossbow Inc. [6]. Each mote is connected to an MTS101 data acquisition board providing ambient light and temperature sensors in addition to ports for up to five external sensors [9]. We attach two sensors to these ports: a Watermark soil moisture sensor and a soil thermistor, both available from Irrometer [7]. We chose the Watermark soil moisture sensor, because it responds well to rain events, closely follows the soil wetting-drying cycle [14], and because it is inexpensive – an important issue for large WSNs. Each mote along with the data acquisition board is enclosed in a waterproof case and is powered by two AA batteries.

Motes sample data every minute and store them on a circular buffer in their local flash. We use the on-board flash memory so we can retrieve all observed data even over lossy wireless links – in contrast to *sample-and-collect* schemes such as TinyDB which can lose up to 50% of the collected measurements [17]. Because a mote collects 23KB per day, the MicaZ 512KB flash measurements will be overwritten if data is not retrieved after 22 days. In practice, the sensor measurements were downloaded from the motes weekly or at least once every two weeks. To ensure reliable delivery, the base station requests the mote's stored data using a simple, NACK-based sliding window ARQ protocol, and stores the retrieved measurements in the database. We cannot provide more details about the transfer protocol due to space restrictions. A detailed description, including an analysis of its performance is provided in the associated technical report [16].

In order to keep track of the status of the motes, each of them broadcasts its status, containing the number of samples collected and the current battery voltage, every two minutes. To ensure that that the basestation get the message it is sent multiple times during a one second interval. During this period the basestation can initiate a download.

3.2 Database

Database Design The database design, visualized in Figure 2, follows naturally from the experimental design and the WSN. The experimental layout is broken into Patches which contain Nodes (motes). There are types of Nodes and types of Sensors that are described in the Type tables in Figure 2. Each Node has a descriptive record in the Nodes table. Each Node has one or more Sensors. Each sensor has a table entry describing the details of that object. The Event table records state changes of the experiment such as battery changes, maintenance, site visits, replacement of a sensor, etc. Global events are represented by pointing to a NULL patch or a NULL node. Measurements are recorded in the Raw and Derived (calibrated) tables. External weather data is recorded in the WeatherInfo table. Various support tables contain lookup values used in sensor calibration.

The database, implemented in Microsoft SQL Server 2005, benefits from the `skyserver.sdss.org` database we built for Astronomy applications [12]. It inherited a self-documenting framework that uses embedded markup tags in the comments of the DDL scripts to characterize the metadata (units, descriptions, enumerations, etc.) for the database objects, tables, views, stored procedures, and columns. A set of stored procedures generate an HTML rendering of the hyperlinked documentation (see *Schema Browser* on our website [1]).

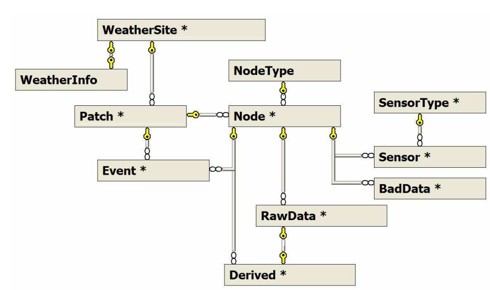

Figure 2: Sensor Network Database Schema.

Data Loading The hardware configuration (Patch, Node, Sensor) and sensor calibrations are preloaded before data collection begins. When new motes or sensors are added, new records are added to those tables. When new types of motes or sensors are added, those types are added to the database type tables.

To date we have loaded 1.6M readings of 3-5 sensors per node, for a total of 6M data points. Raw measurements arrive from the base station as ASCII files in commaseparated-list format. The data are first loaded into a temporary table, where duplicates are removed. Next, the data are copied into the RawData table. Converting the raw data to scientifically meaningful values requires a multi-step pipeline shown in Figure 3 and performed automatically for all the sensors within the database as a stored procedure. The conversions apply to all new RawData values and produce entries in the Derived table.

The interface boards on some sensors had a loose connection for a while. As a result, some RawData measurements were invalid. These intervals are represented in a BadData table, and the corresponding rows in the Derived table are marked as "bad".

Background weather data from the BWI airport is harvested monthly in CSV format from `wunderground.com` and loaded into the WeatherInfo table. This data includes temperature, precipitation, humidity, pressure and weather events (rain, snow, thunderstorms, etc.).

Calibration Knowing and decreasing the sensor uncertainty requires a thorough calibration process. To alleviate errors due to sensor variation we test them for both precision and accuracy. Moisture sensor precision is tested with eight sensors in buckets of wet sand measuring their resistance every ten minutes while varying the temperature from 0°C to 35°C over 24 hours. We found that six sensors gave similar readings, but two did not. This process indicates that such outliers need to be identified and replaced before deployment.

We also performed a preliminary check with the soil thermistors and found they are relatively precise (±0.5°C), yet consistently returned values 1.5°C below a NIST approved thermocouple. The 1.5°C bias does not present a large problem because we convert resistance to temperature using the manufacturer's regression technique. Furthermore, there is a 10 kΩ reference resistance connected in series with the moisture sensors on each mote. Because the resistance's value directly factors into the estimation of the sensor resistance, the bias is individually measured, recorded in the database, and used during the conversion from raw to derived temperature.

The temperature sensors can be calibrated relatively easily as their output is only a function of temperature. On the other hand, moisture sensors require a two-dimensional function that relates resistance to both soil moisture and temperature. We calibrate each moisture sensor individually by taking resistance values at nine points (three moisture contents each at three temperatures), and using these values to calculate individual coefficients to an already published regression form [13].

Data Access and Analysis We use several stored procedures and user defined functions to access the data in various aggregated forms. These functions are accessible through Web-form interfaces that present tabulated values for all the sensors on a given node or for one sensor type across all nodes. We also provide a Web Services interface to display the node locations on a map, where the values of a particular sensor displayed in color. The time series data can also be displayed in a graphical format, using a .NET Web Service. This is an area that needs considerably more work – soil scientists do not want to learn SQL and they often want to see graphical and spatial displays rather than tables of numbers. Beyond better reporting, we plan to use database OLAP and data mining tools.

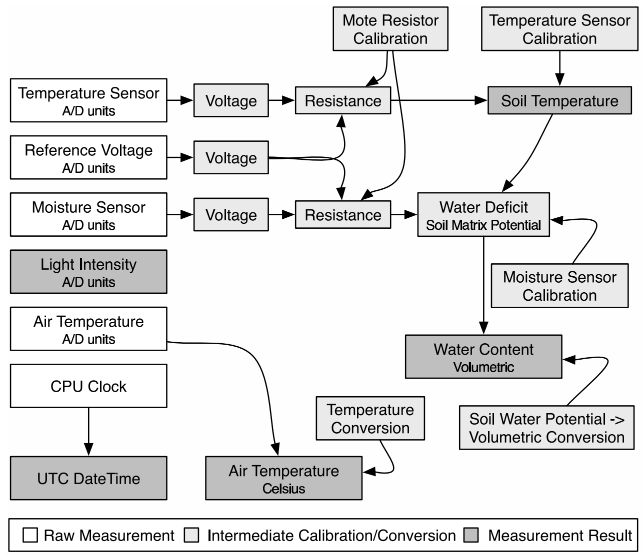

Figure 3: Calibration workflow converting raw to derived science data.

In addition to analyzing the low level readings looking for unusual cases, ecologists analyze aggregations and averages of the sensor data. We are implementing a datacube of the measurements so that spatial and temporal averaging and correlation functions can be displayed more easily and quickly. These aggregates pivot on several dimensions: position on the hillside, depth in the soil, shade vs. in the open, etc.

4 Results

On September 19, 2005, we deployed 10 motes into an urban forest environment nearby an academic building on the edge of the Homewood campus at Johns Hopkins University. The motes are configured as a slanted grid approximately 2m apart from each other. A small stream runs through the middle of the grid; its depth is dependent on recent rain events. The motes are positioned along the landscape gradient and above the stream so that no mote is submerged.

A wireless base station connected to a PC with Internet access resides in an office window facing the deployment. Originally this base station was expected to directly collect samples from the motes. Once the motes were deployed, however, we quickly determined that some motes could not be reliably and consistently reached by our base station (the percentage of received status messages over several months was between 28-33%). Our temporary solution to this problem was to travel to the perimeter of the deployment site and collect the measurements using a laptop connected to a mote acting as base station.

As we mentioned earlier, we were able to download all the data collected by the motes using our reliable transfer protocol. However, due to an unexpected hardware behavior (writing to flash can fail sometimes and should be attempted several times in order to succeed) some motes stopped taking samples. Except the periods between such a failure and the subsequent restart of the failed mote as well as some data lost due to human errors during the download process all data were retrieved successfully.

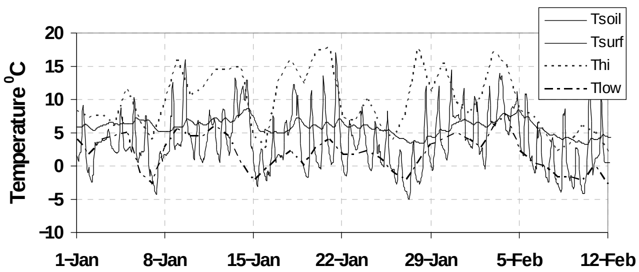

Figure 4: Air and soil temperature over six weeks. Each point represents six hour averages. Tsoil: soil temperature at 10 cm depth; Tsurf: air temperature at soil surface; Thi and Tlow: maximum and minimum temperatures, respectively, for the Baltimore Metropolitan Area.

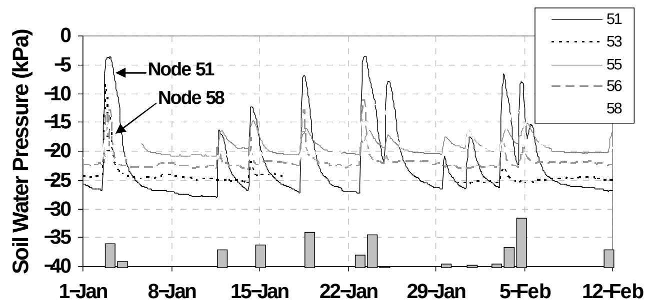

Figure 5: Soil moisture readings over six weeks recorded by five nodes. Each point represents six hour averages. Bars on the bottom indicate precipitation events in the Baltimore Metropolitan Area. Highest column (Feb 5) corresponds to 25.4 mm rain.

4.1 Ecology Results

During the 147 of days of deployment the sensors collected over 6M data points. A subset of the temperature and moisture data are shown on Figures 4 and 5 respectively. Temperature changes in the study site are in good agreement with the regional trend verifying our results. An interesting comparison can be made between air temperature at the soil surface and soil temperature at 10cm depth. While surface temperature dropped below 0°C several times, the soil itself was never frozen. This might be partially due to the vicinity of the stream, the insulating effect of the occasional snow cover, and heat generated by soil metabolic processes. Several soil invertebrate species are still active even a few degrees above 0°C, and, thus, this information is helpful for the soil zoologist in designing a field sampling strategy.

Precipitation events triggered several cycles of quick wetting and slower drying. In the initial installation, saturated Watermark sensors were placed in the soil and the gaps were filled with slurry. We found that about a week was necessary for the sensor to equilibrate with its surrounding. Although the curves on Fig 5 reflect typical wetting and drying cycles, they are unique to our field site. It is because the shape of the soil water characteristic curve depends on soil type, primarily on texture and organic matter content [10].

We deliberately placed the motes on a slope, and our data reflect the existing moisture gradient. For instance mote 51 (Fig. 5) placed high on the slope showed greater fluctuations then mote 58, which was closer to the stream. We occasionally performed synoptic measurements with Dynamax Thetaprobe sensors to verify our results.

Not every sensor worked smoothly, and there were some missing data. However, we are confident, that differences among individual sensors reflect real spatiotemporal heterogeneity. With this information soil ecologists will be able to predict better where and when microbial and invertebrate activity occurs. This activity is tightly coupled with biogeochemical processes, *e.g.* soil respiration, which is an important, but largely unknown component of the global carbon cycle. Continuous in situ monitoring of the soil will improve our estimates on the contribution of the soil biota to these large scale processes.

4.2 Energy Consumption

We power the motes using inexpensive AA Alkaline batteries with capacity of 2100mAh. During a 2 minutes cycle, a mote keeps the radio on for 1 second (drawing 23mA), samples the sensors twice (once per minute with each sample lasting half a second and drawing 0.6mA of current), and stays in low power mode for the rest of the time (0.16mA). Based on this profile, the average current draw is then 0.35mA

After 70 days operation the supplied voltage dropped from 3V to 2.6V. Considering that the cutoff voltage of a single battery is 0.8V and using the linear discharge model to approximate the remaining battery capacity, as suggested in [5], a drop of 0.4V (0.2V per battery) corresponds to consumption of 600mAh. This is very close to the 595mAh computed using the average current drawn by the mote[1]. This calculation holds even for smaller scales: the voltage drop during one week is almost 0.02V (as observed through our online monitoring tools [11]) corresponding to an expense of 60mAh using the linear battery model. For the same period, our average current model estimates energy consumption of 59.47mAh, a 0.8% difference. The high accuracy of of this model indicates that it can be used as a planning tool for estimating the lifetime of a network for a given sampling frequency, or conversely for determining the highest possible sampling frequency given a network lifetime goal.

The MicaZ specification recommends operation voltage range from 2.7V to 3.6V but from our tests we determined that motes can reliably operate down to 2.2V when the flash memory stops responding. We also found that the MCU and the radio work down to 2.17-2.10V. Because the current batteries were installed at the end of November, we expect the nodes will stop recording data in mid April and will stop sending status messages a few weeks later.

5 Discussion

Developing, deploying, and managing the testbed demonstrated several lessons about the state of sensor networks and about barriers to using them as an effective and economical research platform for domain scientists. While some of our observations have been repeated in the literature (*e.g.*, [15]), many of them are new.

We learned, as previously reported, that reprogramming is essential for long-term network deployments. In our case, we discovered two major software faults after the network was initially deployed. The first bug was related to putting the MCU to sleep mode, while the second one was related to occasional errors when writing to the mote's flash memory. In both cases, we had to retrieve the motes and reprogram them in the lab. Had we used a tool such as Deluge, we would be able to reprogram the motes in the field, decreas-

[1] The difference can be explained by the power consumed during data downloads, a factor not included in our analysis.

ing the length of the outage [4]. An unexpected side-effect of reprogramming the motes was that the waterproof cases started leaking after they were opened and close a few times. This unexpected failure argues for designing the hardware in such a way that the case is never opened after the network has been deployed (*e.g.* by using cable connectors so that sensors can be unplugged without opening the case).

Contrary to the promise of cheap WSNs, sensor nodes are still expensive. We estimated the cost per mote including the main unit, sensor board, custom sensors, enclosure, and the time required to implement, debug and maintain the software to be around $1,000, equivalent to the price of a mid-range PC! Calibrating each of the sensors costs more than the sensors themselves – and is not a novice task. While equipment costs will eventually be reduced through economies of scale, there is clearly a need for standardized connectors for external sensors and in general a need to minimize the amount of custom hardware necessary to deploy a sensor network. Unfortunately, sensor and mote vendors seem to want proprietary interfaces to encourage lock-in. We had to manually calibrate each of the sensors prior to deployment to derive the coefficients for the regression form (cf. Section 3.2). These coefficients are subsequently used to automatically translate raw measurements to calibrated values within the database. We could potentially avoid the initial manual step by using higher quality and, thus, more accurate sensors but this would only add to the per-mote cost.

We also found that low-level programming is (still) a necessary and challenging task when building sensor networks. Not only did we have to write low-level device drivers for the soil temperature and humidity sensors, but also for power control, as well as for calibration procedures. Moreover, using acquisitional processors such as TinyDB [8] was not an option in our case given the requirement to collect *all* the data.

Even-though the download strategy we used was sufficient for our purposes, it can be further improved to be more efficient in the face of losses. Furthermore, we will need an efficient, in terms of speed and energy, routing infrastructure that allow us to retrieve the data using multiple hops for larger deployments.

Finally, there is a need for network design and deployment tools that instruct scientists where to place gateways and sensor relay points. These tools will replace the current trial and error, labor-intensive process of manual topology adjustments that disturbs the deployment area.

6 Concluding Remarks

A wireless sensor network is only the first component in an *end-to-end* system that transforms raw measurements to *scientifically significant* data and results. This end-to-end system includes, calibration, interface with external data sources (*e.g.* weather data), databases, web-services interfaces, analysis, and visualization tools.

While the WSN community has focused its attention on routing algorithms, self-organization, and in network processing among other things, environmental monitoring applications[2] require quite different emphasis: reliable delivery of the majority (if not all) of the data and metadata, high quality measurements, and reliable operation over long deployment cycles. We believe that focusing on these problems will lead to interesting new avenues in WSN research.

[2]Sometimes derided as *academically dull applications*, a characterization with which the ecologist in our team does not agree.

Acknowledgments

We would like to thank the Microsoft Corporation, the Seaver Foundation, and the Gordon and Betty Moore Foundation for their support and Sam Small for his help with this paper. Răzvan Musăloiu-E. is supported through a partnership fund from the JHU Applied Physics Lab. Josh Cogan is partially funded through the JHU Provost's Undergraduate Research Fund.

References

[1] Life Under Your Feet. Available at `http://lifeunderyourfeet.org/`.

[2] R. Cardell-Oliver. ROPE: A Reactive, Opportunistic Protocol for Environment Monitoring Sensor Networks. In *Proceedings of the 2^{nd} EmNets Workshop*, May 2005.

[3] R. Cardell-Oliver, K. Smettem, M. Kranz, and K.Mayer. A Reactive Soil Moisture Sensor Network: Design and Field Evaluation. *International Journal of Distributed Sensor Networks*, pages 149–162, 2005.

[4] Deluge: TinyOS Network Programming. Available at `http://www.cs.berkeley.edu/~jwhui/research/projects/deluge/`.

[5] Energizer. Cylindrical Alkaline Batteries - Application Manual. `http://data.energizer.com/PDFs/alkaline_appman.pdf`.

[6] J. Hill and D. Culler. Mica: A Wireless Platform for Deeply Embedded Networks. *IEEE Micro*, 22(6):12–24, Nov. 2002.

[7] Irrometer Company, Inc. `http://www.irrometer.com/`.

[8] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. The Design of an Acquisitional Query Processor for Sensor Networks. In *Proceedings of SIGMOD 2003*, June 2003.

[9] MTS/MDA Sensor and Data Acquisition Board User's Manual. Available at `http://www.xbow.com/Support/Support_pdf_files/MTS-MDA_Series_Users_Manual.pdf`.

[10] R. Munoz-Carpena. Field devices for monitoring soil water content. Bulletin 343, Dept. of Agricultural and Biological Engineering, Univ. of Florida, IFAS Extension. Available at `http://edis.ifas.ufl.edu/current.html`, 2004.

[11] Mote Status of the Olin Deployment. `http://hinrg.cs.jhu.edu/olin/status`.

[12] The Sloan Digital Sky Survey SkyServer. Available at `http://skyserver.sdss.org/`, 2002.

[13] C. Shock, J. Barnum, and M. Seddigh. Calibration of Watermark Soil Moisture Sensors for irrigation management. In *Proceedings of the International Irrigation Show, Irrigation Association*, 1998.

[14] C. Shock, E. Feibert, and S. Jaderholm. A Comparison of Six Soil Moisture Sensors. Malheur Experiment Station, Oregon State University Ontario, OR. Available at `http://www.cropinfo.net/AnnualReports/2001/Popsensortest01.htm`.

[15] R. Szewczyk, A. Mainwaring, J. Anderson, and D. Culler. An Analysis of a Large Scale Habitat Monitoring Application. In *Proceedings of SenSys 2004*, Nov. 2004.

[16] A. Terzis, R. Musaloiu-E., J. Cogan, K. Szlavecz, A. Szalay, J. Gray, R. Burns, and S. Small. Life Under your Feet: A Wireless Sensor Network for Soil Ecology. Technical report, HiNRG, Johns Hopkins University Technical Report, May 2006.

[17] G. Tolle, J. Polastre, R. Szewczyk, N. Turner, K. Tu, P. Buonadonna, S. Burgess, D. Gay, W. Hong, T. Dawson, and D. Culler. A Macroscope in the Redwoods. In *Proceedings of the Third ACM Conference on Embedded Networked Sensor Systems (SenSys)*, Nov. 2005.