

A Macroscope in the Redwoods

Gilman Tolle,
Joseph Polastre,
Robert Szewczyk, and
David Culler
Computer Science Division,
University of California,
Berkeley
Berkeley, CA 94720

Neil Turner, Kevin Tu,
Stephen Burgess, and
Todd Dawson
Department of Integrative
Biology, University of
California, Berkeley
Berkeley, CA 94720

Phil Buonadonna,
David Gay, and Wei Hong
Intel Research Berkeley
Berkeley, CA 94704

ABSTRACT

The wireless sensor network “macroscope” offers the potential to advance science by enabling dense temporal and spatial monitoring of large physical volumes. This paper presents a case study of a wireless sensor network that recorded 44 days in the life of a 70-meter tall redwood tree, at a density of every 5 minutes in time and every 2 meters in space. Each node measured air temperature, relative humidity, and photosynthetically active solar radiation. The network captured a detailed picture of the complex spatial variation and temporal dynamics of the microclimate surrounding a coastal redwood tree. This paper describes the deployed network and then employs a multi-dimensional analysis methodology to reveal trends and gradients in this large and previously-unobtainable dataset. An analysis of system performance data is then performed, suggesting lessons for future deployments.

Categories and Subject Descriptors

C.2.1 [Computer - Communication Networks]: Network Architecture and Design—*Wireless communication*; C.3 [Special-Purpose and Application-Based Systems]: Real-time and embedded systems; J.3 [Life and Medical Sciences]: Biology and genetics

General Terms

Design, Experimentation, Measurement, Performance

Keywords

Wireless Sensor Networks, Microclimate Monitoring, Macroscope, Application Analysis

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SenSys'05, November 2–4, 2005, San Diego, California, USA.
Copyright 2005 ACM 1-59593-054-X/05/0011 ...\$5.00.

1. INTRODUCTION

Wireless sensor networks offer the potential to dramatically advance several scientific fields by providing a new kind of instrument with which to perceive the natural world. As the telescope allowed us to perceive what is far away and the microscope what is very small, some refer to sensor networks as “macroscopes” [5] because the dense temporal and spatial monitoring of large volumes that they provide offers a way to perceive complex interactions. As the technology has progressed, we have gotten ever closer to obtaining such macroscopic views of previously unrecorded phenomena [9, 11, 15]. This paper reports on a case study of microclimatic monitoring of a coastal redwood canopy, a case study that we believe has clearly crossed that threshold. Using a large number of wireless micro-scale weather stations we have obtained an unprecedented picture of environmental dynamics over such a large organism. Here we describe the study, present an overview of the data that has been obtained, and use a multidimensional analysis methodology to more deeply understand the dense and wide-ranging spatiotemporal data obtained from the macroscope.

2. MOTIVATION

In meeting with a collection of local biologists, we began with the question of what would they like to observe that they simply cannot measure today. The responses covered a wide array of interests, including the dispersal patterns of wind-borne seeds, the water profiles experienced by spawning salmon, insect densities across riparian environments, and the microclimate of meadow and woodland transects. In classifying these desires against the requirements they place on the underlying technology and the state of the art in the measurement and analysis techniques, we arrived at an initial choice of studying the ecophysiology of coastal redwood forests.

The microclimate over the volume of an entire redwood tree is known to have substantial variation and to have substantial temporal dynamics. When you walk in the forest it is temperate and moist, despite the wide variation in weather conditions. The top of the tree experiences wide variation in temperature, humidity, and, of course, light, whereas the bottom is typically cool, moist, and shaded. This variation was understood to create non-uniform gradients, essentially weather fronts, that move through the structure of the tree. For example, as the sun rises, the top of the canopy warms quickly. This warm front moves down the tree over time until the entire structure stabilizes or until

cooling at the canopy surface causes the process to reverse. Humidity fronts also move through the canopy, but the process is complicated by the tree moving so much water up from the soil and into the air. At some point, the observed humidity is driven by the transpiration process and decoupled from the prevailing climate conditions. The biologists working in this area had substantial experience with sensors for measuring the relevant climatic factors and had devised methods for instrumenting trees using conventional technology. They would climb the tree to attach a winch near the top, typically 50m to 70m up, and haul a suite of weather monitoring instruments up the vertical transect with a very long serial cable connecting to a battery powered data logger at the base. With this limited apparatus, they were able to validate that there was substantial variation, but they could not get a detailed picture of the entire structure over time. This defined our challenge.

Together we designed a wireless micro-weather station based on the Berkeley Motes manufactured by Crossbow. The requirements and design of the node are described in Section 4.1. We built upon the established system, networking, and data access technology provided by TinyOS [7], MintRoute [18], and TinyDB [8]. This provided a flexible basis for both the ad hoc multihop routing and the range of query processing the biologists might want to explore. However, as this provided much more complex functionality than, for example, the dedicated sense-and-send capability deployed on Great Duck Island [17], we provided a simple internal data logging backup. We also had to develop a GPRS base-station to provide the transit link between the deployment site and the Internet. The requirements and design of the network and software architecture are described in Section 4.2. In addition to the base technology, we found that we needed to create an experimental methodology that described how to calibrate a large collection of instruments en masse, deploy them, and collect the data. This aspect is outlined in Section 4.3.

Having carried out a 44-day study of trees in a study area in Sonoma California, the main thrust of this paper is showing the data we were able to obtain: a month in the life of a redwood tree. After the initial excitement of seeing that the data really showed the kinds of dynamic gradients that the biologists were hoping to see, we spent months figuring out how to analyze and present the data systematically. Here we illustrate one such methodology. We start out by viewing the entire network as a single instrument. This allows us to summarize the overall climatic distribution with a substantial sampling over space and time. Then we examine time series data using the entire network as a distributed instrument. We also examine the spatial characteristics, summarized over the duration of the study. After understanding these overall characteristics, it is much more enlightening to examine the detailed dynamics over space and time. Ideally, this is done by movies, but we present snapshots here.

3. RELATED WORK

Several prototype sensor networks have been used to monitor environmental parameters within a well-defined spatial region.

Researchers at the Center for Embedded Networked Sensing deployed a sensor network called the Extensible Sensing System into the James Reserve Forest [4]. The ESS has been organized into various topologies, with up to 40

nodes. Directed Diffusion and Tiny Diffusion are used as a reliable publish-subscribe middleware system to enable runtime queries for data collection and sensor reconfiguration [14]. Microclimate sensors, and soil temperature, moisture, and nitrate sensors are all present within the ESS. The ESS has been used to monitor temperature and humidity in six bird nesting boxes for a period of several days.

Cardell-Oliver et al. measured waterflow through soil during periods of active rainfall [3]. This sensing task required a network that could react to the beginning of a rainstorm by beginning to sample at a high temporal frequency, then return to an infrequent monitoring mode after the end of the storm. Their work shows readings taken from 3 soil moisture sensors over 15 days.

In a system called GlacsWeb, researchers at the University of Southampton deployed 9 sensor nodes inside a glacier [11]. The nodes monitored pressure, temperature, and tilt, in order to monitor glacier melting behavior. Their work showed the pressure on a single sensor falling over the course of seven days as the temperature and tilt remained the same. Additionally, the system was able to track changes in the tilt of the base station over several months as warm weather caused the shape of the glacier to change.

As an experiment in adaptive sampling methods, a hanging autonomous vehicle was deployed into the James Reserve by researchers at UCLA and USC [1]. The hanging vehicle monitored microclimate over a plane perpendicular to the ground instead of over a flat area. Using this technology, the researchers interpolated and plotted solar radiation intensity over a 70m by 15m transect of the James Reserve forest.

To study the movements of nesting birds, researchers at the University of California, Berkeley deployed 2 networks with a total of 147 nodes onto an island on the coast of Maine, called Great Duck Island [9]. The GDI deployment used a combination of ground-emplaced motes to monitor microclimate, and motes placed in burrows to detect the presence of birds. The size of this network was steadily increased until it covered an 221m by 71m ellipsoidal pattern. The occupancy data collected by this network was verified by a superimposed network of cameras. The researchers showed 16 days worth of temperature readings from a single bird burrow, and correlated the rises in temperature with the presence of a nesting bird [9]. Two years later, the researchers demonstrated an interpolation of surface temperature and burrow temperature at a single instant in time over the complete deployment area [16].

There is, of course, a wealth of biological studies that used small collections of data loggers and remote sensing to study temporal and spatial gradients in forest microclimate [10, 12, 13]. We apply some of their analysis methods to data obtained from our wireless sensor network.

4. DESIGN

Gathering data on the environmental dynamics around 70-meter tall redwood tree for 44 days requires robust system design and a careful deployment methodology. We selected a suite of sensors, integrated them with an existing wireless sensor node platform, and designed a package that both resists the elements and allows the sensors access to the environment. We used the latest TinyOS and TASK software for the node operating system, networking stack, and data collection framework. Finally, we performed two

separate calibration procedures to test the system prior to placing it in the field, and then spent a day in the forest with ropes, harnesses, and a notebook.

We decided on the following envelope for our deployment:

Time: One month during the early summer, sampling all sensors once every 5 minutes. The early summer contains the most dynamic microclimatic variation. We decided that sampling every 5 minutes would be sufficient to capture that variation.

Vertical Distance: 15m from ground level to 70m from ground level, with roughly a 2-meter spacing between nodes. This spatial density ensured that we could capture gradients in enough detail to interpolate accurately. The envelope began at 15m because most of the foliage was in the upper region of the tree.

Angular Location: The west side of the tree. The west side had a thicker canopy and provided the most buffering against direct environmental effects.

Radial Distance: 0.1-1.0m from the trunk. The nodes were placed very close to the trunk to ensure that we were capturing the microclimatic trends that affected the tree directly, and not the broader climate.

Figure 1 shows the final placement of each mote in the tree. We also placed several nodes outside of our angular and radial envelope in order to monitor the microclimate in the immediate vicinity of other biological sensing equipment that had previously been installed.

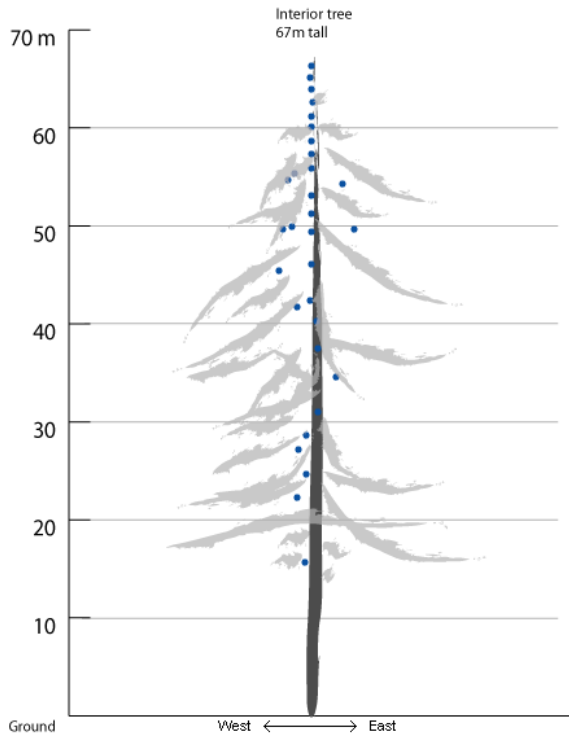


Figure 1: The placement of nodes within the tree

4.1 Hardware and Network Architecture

The sensor node platform was a Mica2Dot, a repackaged Mica2 mote produced by Crossbow, with a 1 inch diameter form factor. The mote used an Atmel ATmega128 microcontroller running at 4 MHz, a 433 MHz radio from Chipcon operating at 40Kbps, and 512KB of flash memory. The mote was connected to digital sensors using I2C and SPI serial protocols and to analog sensors using the on-board ADC.

The choice of measured parameters was driven by the biological requirements. We measured traditional climate variables – temperature, humidity, and light levels. Temperature and relative humidity feed directly into transpiration models for redwood forests. Photosynthetically active radiation (PAR, wavelength from 350 to 700 nm) provides information about energy available for photosynthesis and tells us about drivers for the carbon balance in the forest. We measure both incident (direct) and reflected (ambient) levels of PAR. Incident measurements provide insight into the energy available for photosynthesis, while the ratio of reflected to incident PAR allows for eventual validation of satellite remote sensing measurements of land surface reflectance. The Sensirion SHT11 digital sensor provided temperature ($\pm 0.5^\circ\text{C}$) and humidity ($\pm 3.5\%$) measurements. The incident and reflected PAR measurements were collected by two Hamamatsu S1087 photodiodes interfaced to the 10-bit ADC on Mica2Dot.

The platform also included a TAOS TSL2550 sensor to measure total solar radiation (300nm - 1000nm), and an Intersema MS5534A to measure barometric pressure, but we chose not to use them in our deployment. During calibration, we found that the TSR sensor was overly sensitive, and would not produce useful information in direct sunlight. Because TSR and PAR would have told roughly the same story, and because PAR was more useful from the biology viewpoint, we decided not to gather data on total solar radiation. As for the pressure sensor, barometric pressure is simply too diffuse a phenomenon to show appreciable differences over the height of a single redwood tree. A standard pressure gradient would exist as a direct function of height, but any pressure changes due to weather would affect the entire tree equally. Barometric pressure sensing should be useful in future large-scale climate studies.

The package for such a deployment needs to protect the electronics from the weather while safely exposing the sensors. Our chosen sensing modalities place specific requirements on the package. Standardized temperature and humidity sensing should be performed in a shaded area with adequate airflow, implying that the enclosure must provide such a space while absorbing little radiated heat. The output of the sensors that measure direct radiation is dependent on the sensor orientation, so the enclosure must expose these sensors and level their sensing plane. The sensors measuring ambient levels of PAR must be shaded but need a relatively wide field of view.

The package designed for this deployment is shown in Figure 2. The mote, the battery, and two sensor boards fit inside the sealed cylindrical enclosure. The enclosure is milled from white HDPE, and reflects most of the radiated heat. The endcaps of the cylinder form two sensing surfaces – one captures direct radiation, the other captures all other measurements. The white “skirt” provides extra shade, protec-

tion of the bottom sensing surface from wind and water, and it serves as an attachment point to the tree.

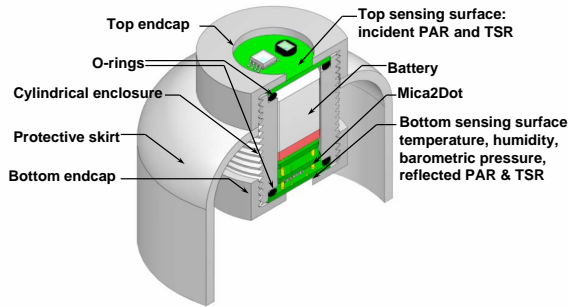


Figure 2: Sensor node and packaging

The gateway between the sensor network and the wider world was the Stargate system developed by Intel Research – a PC-class node intended to run from batteries and solar panels, housed in a sealed package for protection from the elements. Data from the sensor network nodes was collected over an Mica2 node attached to RS232 serial, stored in a local database, and then transmitted over a GPRS cellular modem to an offsite database.

4.2 Software Architecture

Instead of building a custom application, we chose to use the publicly available TASK system developed by UCB and Intel Research Berkeley [2]. It provides a query-based framework linking a network of sensors to a database running on a gateway. TASK supports duty cycling of the nodes for power conservation and robust routing for data collection.

To meet the lifetime requirements of our deployment envelope, we needed to perform power management on the node. TASK uses a duty-cycling approach in which the network wakes up periodically, takes sensor readings, stays awake for a defined period of time to transfer the data back to the base station, and then returns to sleep. Our network was awake for 4 seconds every 5 minutes – a duty cycle of 1.3%. Time synchronization was needed to schedule this network wakeup, and TASK provided this service as well.

Collecting data from each node in the mesh network requires a routing tree between the gateway and the nodes. TASK builds upon MintRoute [18], which uses beacon messages to build an estimate of the number of transmissions required to successfully send a packet to each neighbor. MintRoute then selects the neighbor that minimizes the total number of expected transmissions required to forward a packet to the base station, recursively constructing a routing tree.

Because of the diversity of sensors in our sensing platform, we desired a flexible data collection package. TASK includes the TinyDB data collection framework, which presents the metaphor of a database table in which each column represents a particular sensor and each row represents readings taken at a particular time. Node ID, sample number, and sample reception time are also represented as columns in this virtual table. The data collection process begins with a query over the table, using a SQL-like language called TinySQL. TinyDB does support the computation of aggrega-

tion functions over table columns, but for this deployment, we used a simple selection query:

```
SELECT result_time, epoch, nodeid,
       parent, voltage, depth,
       humidity, humid_temp, hamatop, hamabot
FROM sensors
SAMPLE PERIOD 5 min
```

To provide a backup in case of network failure and to establish a basis for analyzing the performance of the network, we extended the TASK framework to include a local data logging system. The data logger recorded every reading taken by every query before the readings were passed to the multi-hop routing layer, and stopped recording once the 512 kB flash chip was full. After the deployment, we attached each mote to a serial connection, and then installed a new program to transmit the contents of the flash over the serial link. We chose to include a complete data logger because we knew that the capacity of the flash was sufficient for the duration of our deployment. Longer deployments should consider including a storage system that supports multiple resolutions [6], or one that can be partially retrieved over the network.

Every reading received from the multihop network was stored into a DBMS running on the Stargate gateway node, as a staging area before we retrieved the results. We retrieved some results directly over the GPRS cellular modem connection in real time, and we periodically attached a laptop computer to the Stargate in order to download the rest. The GPRS connection was also used to remotely reboot the gateway node in the event of failure.

4.3 Deployment Methodology

Collecting high-quality real-world data with a wireless sensor network requires a comprehensive deployment strategy that carefully tests and calibrates the sensors prior to deployment. Our strategy used two calibration phases: roof and chamber. Each phase provided performance data on different subsets of the sensors installed on the mote. We built several racks that could hold the motes in a known position and orientation, making it easier to move and install the nodes for each phase.

The roof calibration provided a real-world data source for the PAR sensors: direct sunlight. Direct sunlight provides a much wider range of readings than can be achieved in an environmental chamber, and using sunlight made it possible to study the response of the upper PAR sensor to varying incident light angles as the sun moved overhead. We chose to use a roof for this calibration phase because it provided a clear, unobstructed view of the sky. The sensors were left on the roof in their racks for two days as we collected PAR readings every 30 seconds. Reference data was collected by a high-quality well-established PAR sensor. We established that the PAR sensors were producing acceptable readings.

The purpose of the chamber calibration phase was to understand the response of the temperature and humidity sensors to a wide range of phenomena. We placed the racks of motes into a controllable weather chamber, and cycled it between 5 and 30 °C and between 20 and 90 %RH as the temperature and humidity sensors were sampled every 30 seconds. We then used this data to perform a two-point calibration for each humidity and temperature sensor. The results of this chamber calibration are shown in Section 5.7.

After installing batteries, we placed racks of nodes in the chamber, cycled it through several hours for a second calibration, and without turning off the nodes, drove the racks to the Grove of the Old Trees in Sonoma, California. With the nodes still in the calibration racks, we issued a reset command and then issued the query that would run for the duration of the deployment. We started the query before installing the nodes for two reasons: to check that each node could synchronize with the gateway installed at the site and to ensure that all nodes were operating correctly in the event that connectivity was lost during the installation process. Individual nodes were then installed in the tree. A written log was kept of each node’s height in the tree, distance from the trunk and compass orientation (North, South, East, West). As a final step, we verified that the gateway was receiving data from at least some of the nodes before departing.

After the end of our deployment envelope, we brought the nodes down from the tree and returned them to the calibration racks. Because the TASK persistent log could only be accessed over a local serial connection, we had to dismantle each mote before downloading the log. We also used this opportunity to inspect each mote for weather damage, and found even though a few nodes had been infiltrated by water and dirt, the packaging was successful on most of the nodes.

5. ANALYSIS

When sampling for long periods of time at the high spatial and temporal densities described in Section 4, very large quantities of data are acquired. Our first reading was taken on Tuesday, April 27th 2004, at 5:10pm, and our last reading was taken on Thursday, June 10th 2004, at 2:00pm, nearly 44 days later. By taking a reading every five minutes from four different sensors: temperature, humidity, incident photosynthetically active solar radiation (PAR), and reflected PAR, the maximum number of readings we could have acquired is 50,540 real-world data points per mote. With 33 motes deployed into the tree, we could have recorded 1.7 million data points. Extracting meaning from the resulting set of 820,700 data points, a 49% overall yield, requires a combination of expressive visualization and careful analysis.

5.1 Multi-Dimensional Analysis

Each data point collected by this sensor network can be viewed as having a location in a three-dimensional space: a *time* dimension, a *height* dimension, and a dimension corresponding to the sensor *value* itself. Even though the deployment envelope did include a small amount of variation in the radial location and distance from the trunk, the range of possible locations were not covered densely enough to justify considering these to be separate dimensions. Understanding 820,700 data points that lie in four different 3-dimensional spaces, one for each sensor, is a daunting task. However, analyzing the data becomes simpler if we project the points onto a subset of the dimensions. We start by projecting all of the points onto the *value* dimension, which will be referred to as “stage 1”. This transforms each 3-dimensional dataset into a simpler 1-dimensional dataset that we can use to understand the range and distribution of the data independent of space and time. We then begin to add dimensions back. In stage 2, we project onto the 2-dimensional subspace defined by *time* \times *value* and use this projection to study the temporal trends in the data values. In stage 3, we project

Sensor	Min	Median \pm Quartile	Max
Temperature ($^{\circ}$ C)	6.6	14.1 \pm 3.7	32.6
Humidity (%RH)	16.4	65.5 \pm 17.9	100.2
Incident PAR	0	38.4 \pm 84.8	2154
Reflected PAR	0	0 \pm 0	180

Table 1: Range for each sensor

onto the 2-dimensional subspace defined by *height* \times *value* and look for spatial trends in the distribution of data values. The 2-dimensional subspace of *time* \times *height* offered little information, so it is omitted from this analysis. Once we have examined the 1-D and 2-D subspaces, we have a better understanding of the trends in the data. Being aware of these trends allows us to more effectively grasp the complexity in the full 3-D dataset in stage 4.

The trends revealed in each stage of analysis are best shown in a particular graph for that stage. The purpose of stage 1 is to examine the overall range of the measurements, and so we show a histogram. In stage 2, we see that there are many more points in the time dimension than in the height dimension. Thus, we use the day as a natural grouping unit, and then present the distribution of readings on each day using a collection of box plots. In stage 3, we are focusing on spatial trends over many fewer points, so we do not perform any grouping. We show one box plot for each different height. In both stages, related box plots will be shown on the same axis so that their medians and spreads can be compared visually.

As is common to experimental sensor network deployments, missing values are scattered throughout time and height. In response, our methodology works in spite of missing values because it studies distributions. Missing values may shift the distributions, but they will not make the analysis impossible to complete. The analysis of how much the distributions can be shifted by missing values is left for future work. In the fourth stage, when the analysis focuses on individual readings that have been localized in both space and time, careful interpolation will be used to produce cleaner trends.

In the first two stages, we show that even without explicitly considering spatial information, a collection of sensors can produce more detailed information on the range of a temporal phenomenon. When we include the spatial information in the third and fourth stages, we will present new data that can only be attained with a dense network of wireless sensors.

5.2 Range Analysis

The first step of the analysis is the simplest, but provides a grounding for future work and can highlight the presence of bad data. We can study the range of each microclimatic variable by projecting onto *value*, removing the effects of both *time* and *height*. The ranges are shown in Table 1. This first step lets us sanity-check the data by ensuring that the readings are within the normal ranges for the phenomena under study. Both temperature and humidity are normal, and the maximum readings collected by the two PAR sensors are the expected values for direct and reflected sunlight. We also examined the raw ADC values and determined that none of the sensors were being saturated.

Then, we can examine the distribution of the readings. Figure 3(a) shows the distribution of the sensor readings for each of the four sensor types. Temperature shows a unimodal distribution, while humidity shows a bimodal distribution. We attribute the extra peak at 100 %RH to the fog that is so common on the California coast in the early summer. In the incident PAR readings we see a bimodal distribution, because full sun and no sun are present for more time than the transitions in between. The reflected PAR readings do not show the same bimodal distribution, because the direct sunlight is diffused by the reflection process.

5.3 Temporal Trends

By projecting onto the *time* and *value* dimensions, we collapse the readings taken by different nodes at the same timestep. We can then extract the purely temporal trends in the local microclimate. Figure 3(b) shows the distribution of the sensor readings taken on each of the 44 days. Each box shows the median of the data and the interquartile range between the 25th and the 75th percentile. The dashed whisker lines contain all the data values that fall within 3 multiples of the interquartile range, and the outlying points show all data outside of that range. In the incident PAR box plots, the high density of outlying points appears to be a solid line, but these are still outliers. In the reflected PAR box plots, the extremely low sensor readings were quantized in such a way that the outliers appear to be dashed lines, but these are outliers as well. By looking at the daily median readings, we can see weather movement in the large. Week one includes warm, dry days and cold, wet days, but the three following weeks contain predominantly cold, wet days. Weather slowly improves in the fifth week, only to return to cold and wet in the final week.

Each daily distribution contains a collection of readings taken from a large number of sensors spread over many different points. Thus, the range of readings seen on a given day accurately captures the range of microclimatic parameters experienced by the whole tree on that day. On May 7th, for example, we see that the bulk of the relative humidity readings lie above 95 %RH, and at no time during the day is the relative humidity lower than 75 %RH. This suggests that the entire tree is encased in fog, for nearly all of the day. If we examine the range of the incident PAR readings on the same day, we can see that the distribution is skewed much lower than on the neighboring days, confirming our earlier hypothesis. A single weather sensor near the top of the tree would show a very different range of readings than one placed near the bottom, but by combining readings from spatially-distributed sensors, we can better understand the overall temporal trends.

5.4 Spatial Trends

By projecting onto the *height* and *value* dimensions, we collapse the temporal effects and can then look for spatial trends. Figure 3(c) shows the distribution of all the readings taken by each sensor at each height. In the PAR readings, we can see a spatial trend in both the mass of the distributions and in the outliers. Light is absorbed by the canopy leaf mass as it propagates downward, which lessens the amount of light that reaches the lower portions of the tree. By examining the distribution of the readings, we can understand whether this is an absolute relationship or a tendency. In

the mass of the PAR readings, we can see that lower sensors receive less light. However, though the mass of the distribution at each height moves closer to zero as height decreases, we see in the outliers that the incident PAR sensors still see almost the full range of light readings at every height. Even the lowest node still receives full sun occasionally, due to gaps in the leaf mass. In contrast, we see that the reflected PAR sensors at the lower levels of the tree never do receive the same amount of light as the higher sensors. These trends in the falloff of light over height and in the changing ratios of incident light to reflected light can help to ground the theoretical models of canopy density.

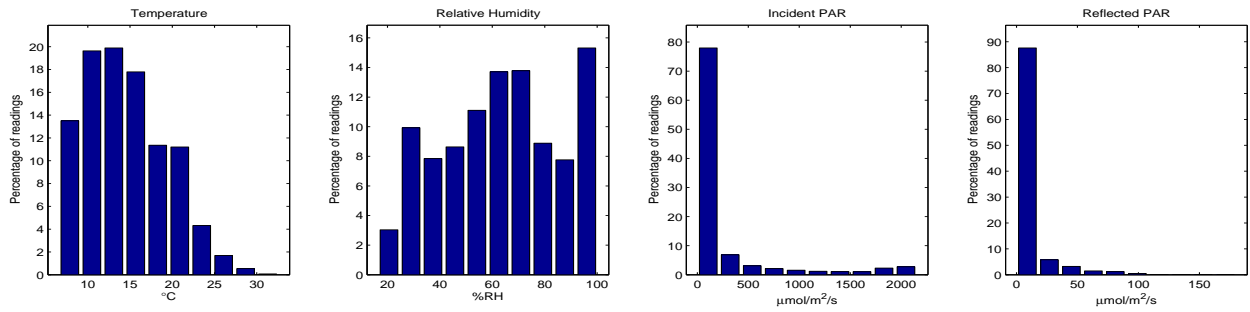
We see a very different spatial trend in the temperature and humidity readings: nothing. Every sensor reached practically every point in the space of possible temperature and humidity readings. This suggests that the amount of variation over time overwhelmed the amount of variation over space. However, instead of just collapsing the temporal variation, we can explicitly remove it and focus more closely on the spatial trends. At every timestep, we take the mean of all the sensor readings. We can subtract the timestep mean from each sensor reading, and examine the distributions of the differences (Figure 3(d)).

We now see that the lowest sensors in the tree are colder than average, in a way that is never seen at the top of the tree. As a side note, the anomalously wide-ranging sensor in the middle of the tree was actually placed 3 meters out from the trunk, outside of the deployment envelope. Thus, it was less insulated by the canopy and saw a wider range of temperatures. In the relative humidity readings, we see the opposite trend: the bottom of the tree can be up to 40 %RH more humid than average, the top of the tree can be up to 30 %RH less humid than average, and these deviations are never reached by the opposing regions. Interestingly enough, we also see that the higher sensors can be up to 10 %RH more humid than average and that the lower sensors can be up to 7 %RH less humid than average. These results suggest that spatial gradients may be present over the height of the tree, but we cannot confirm this without correlating the readings in time.

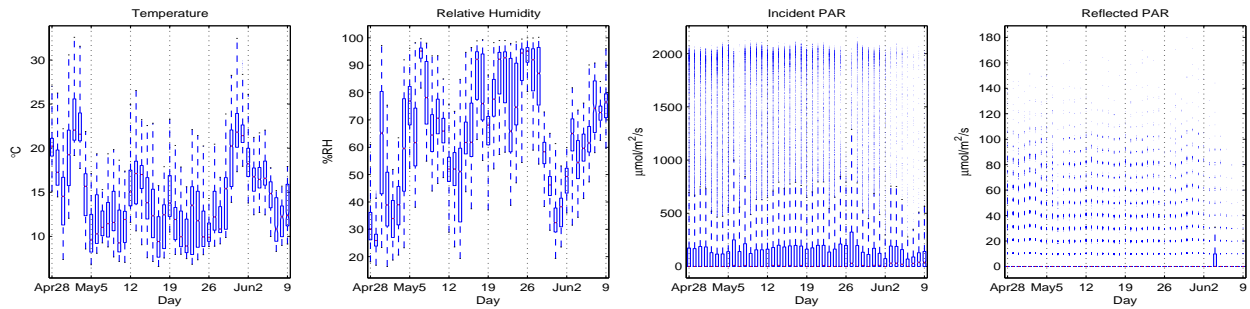
5.5 Combined Analysis

When considering both time and space, we can start to understand just how dynamic the microclimate surrounding a coastal redwood truly is. We focus on a single day, May 1st, because it contains a wide range of temperature and humidity readings throughout the course of the day, and many different amounts of spread over space. At this stage of the analysis, we no longer study distributions. Rather, we now consider the individual data points in order to extract dynamic behavior. Because the dataset is now fully three-dimensional, we must worry about an additional projection: onto two-dimensional paper. First, we use a figure that shows the actual data points in *time* \times *value* and *height* \times *value* on two separate plots. Then, we show a combined plot that uses color to represent the *value* dimension.

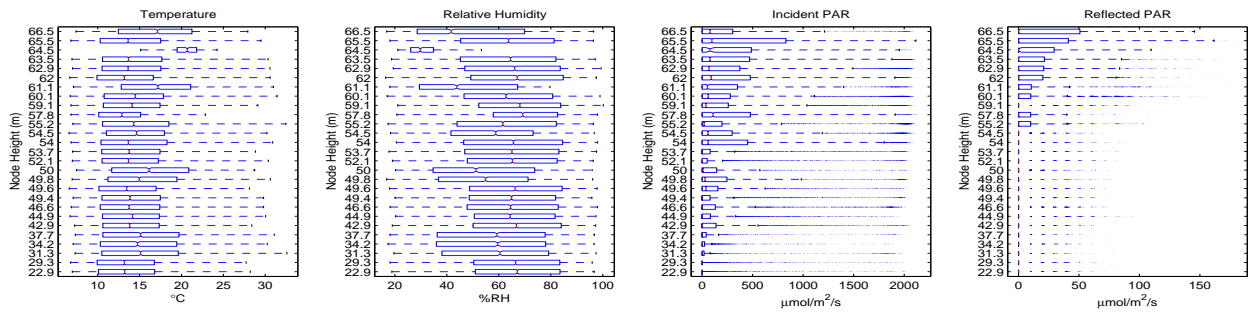
Figure 4 shows a day in the life of a redwood tree, as seen through our network. The charts on the left show the temporal trends of all the sensors, but they discard information about each sensor's location in space. The charts on the right place each sensor in its correct spatial location, and show the spatial gradients at a single moment in time.



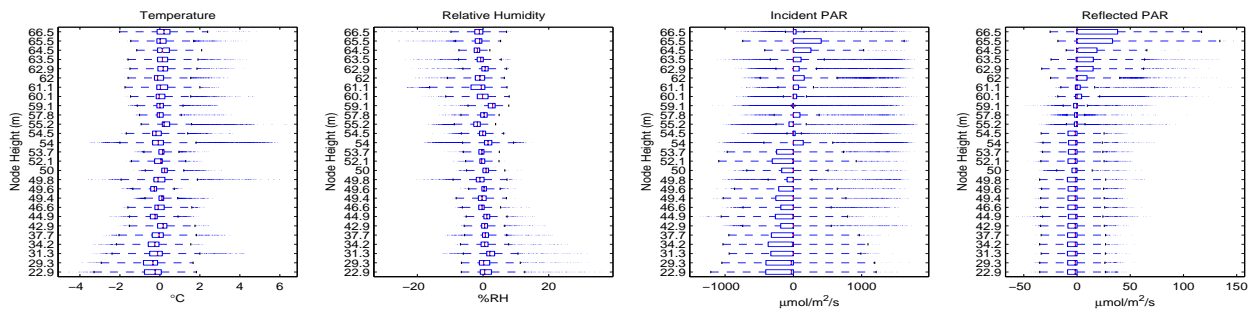
(a) Distributions of sensor readings projected onto the *value* dimension



(b) Distributions of sensor readings projected onto the *time* × *value* dimensions



(c) Distributions of sensor readings projected onto the *height* × *value* dimensions



(d) Distributions of sensor reading differences from the mean

Figure 3: Multidimensional analysis reveals distributions and trends

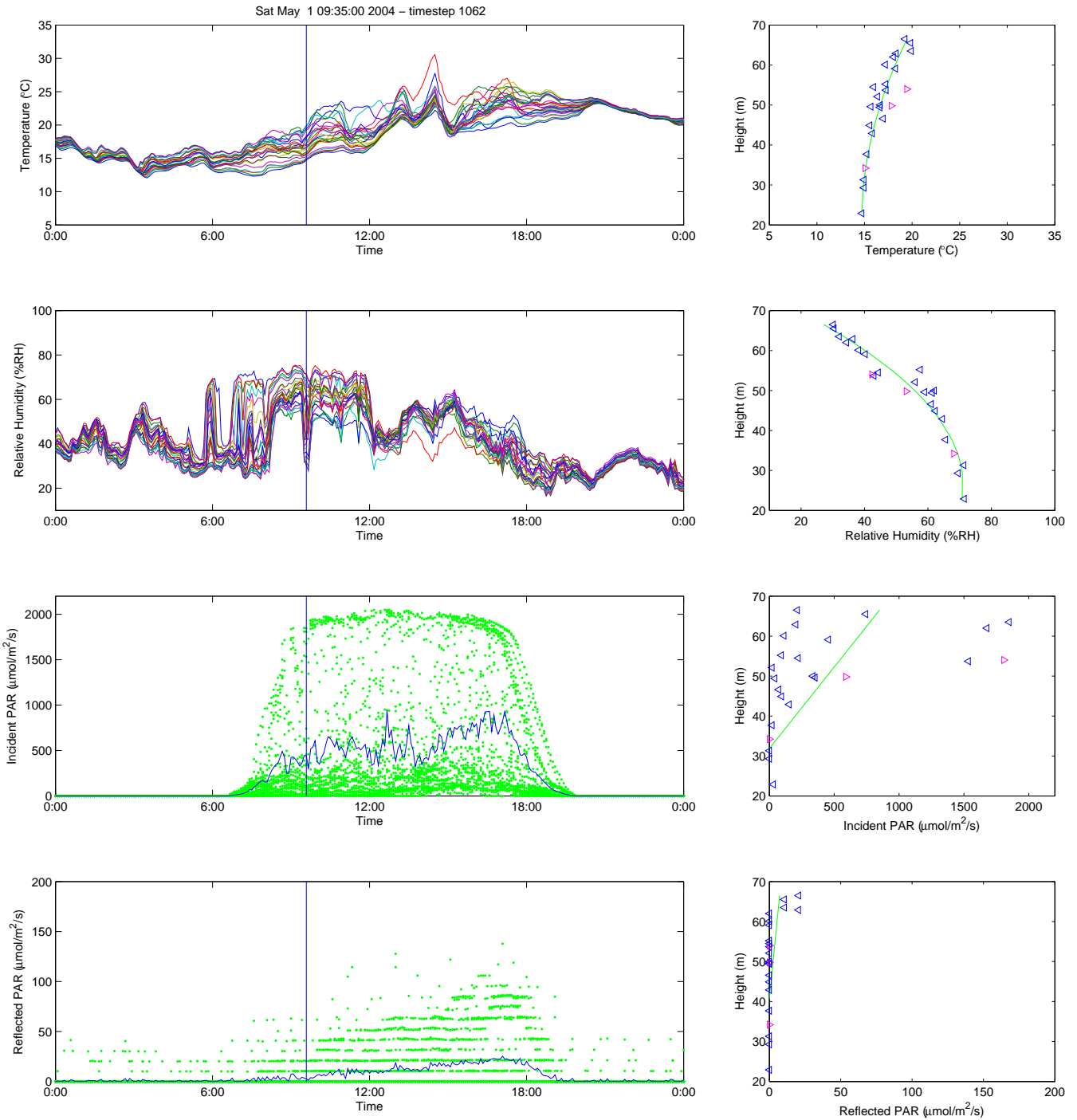


Figure 4: Temporal trends and a snapshot in space

We can see the effects of solar movement in the two bottom-left charts. The dots represent individual light readings, while the line represents the mean of the light readings at each timestep. Incident PAR clearly follows the normal movement of the sun as it rises and then sets, but reflected PAR is more noisy because the absolute level of the readings is much lower, and the sunlight must take a more complicated path to reach the sensor. One would expect the mean illumination in the morning and in the afternoon to be roughly the same, but the increased quantity of light in the afternoon can be accounted for the placement of our sensors on the western side of the trunk. For PAR, we choose not to show each sensor’s individual trend because the individual sensors do not move in unison, even though the general trend reflects the solar movement. We explore this effect more thoroughly in Section 6.

In the two upper-left charts, we can see the movement of temperature and relative humidity throughout the day. Each line represents the readings taken by an individual sensor. We see that both the temperature and the spread of temperatures throughout the tree increase as the sun rises. Using only this graph, we cannot see whether the increased spread represents a particular trend over space or whether it is due to random local variations in air movement and solar access. Temperature peaks in the afternoon, and then descends after the sun has set. The setting of the sun also drastically reduces the differences between the sensors, confirming that the temperature spread is due to solar influence.

The movement of relative humidity tells an entirely different story. Prior to sunrise, the humidity around the tree changes very quickly, and over a notably wide range. In 30 minutes, we see a change of 50 %RH. One would normally expect relative humidity to move inversely with temperature, because cooler air can hold less water and therefore presents a higher %RH for the same absolute amount of water present in the air. We see this effect in the matched peaks at about 2pm. However, the humidity movements in the morning are not matched by corresponding temperature movements. Thus, they must represent changes in the local absolute humidity. Moving forward, we see that the rising sun is correlated with similarly drastic humidity changes, unmatched by temperature changes. Once the sun has risen, we see that temporal variability is reduced, but that spatial variability remains high. However, the highlighted dip in humidity occurred during the day, and suggests that every microclimatic trend has its own exceptions. In the afternoon, humidity decreases overall, and after the sun sets, the spread in humidity reaches its lowest point. The tree’s microclimate has entered the still, calm night.

On the right side of the figure, we can see a new way to visualize the local microclimate – a view that can only be obtained through a sensor network. The stick represents an instantaneous gradient over the height of the tree, taken at the timestep shown by the vertical marker line on the time-series charts. We have chosen to highlight an extremely rapid dip in humidity – a dip uncorrelated with temperature or PAR. Examining the lower-right PAR charts, we see that the top of the tree is receiving more light, both incident and reflected, at this moment. However, the incident light falls off less quickly than the reflected light does, matching the spatial trend seen in Section 5.4. The individual sensors are scattered around the trend line, but the correct trend is present.

In the two upper-right charts, we can see the presence of instantaneous temperature and relative humidity gradients. At this time, the top of the tree is 5 °C warmer and 40 %RH less humid than the bottom of the tree. We can also see that temperature and humidity change in a nonlinear fashion over height, and we have modeled the gradient with a quadratic equation instead of with a linear equation. Unlike in the PAR gradient, the individual sensors are very close to the trend line. The rightward-facing arrows indicate sensors that were placed on the east side of the tree, outside of the deployment envelope. Because this snapshot was taken in the morning, we can see that the eastern sensors are warmer than the gradient line would suggest.

Even with these instantaneous snapshots of spatial gradients, we still cannot see how the gradient changes over time. We know that the top of the tree is less humid than the bottom, but we do not know how much more humid it was in the last timestep. Fortunately, our data allow us to examine the temporal movement of spatial gradients as well. A plot like that of Figure 5 is one way of showing these trends. This figure shows time on the horizontal axis, height on the vertical axis, and the magnitude of the reading with color. The color at each height actually represents the interpolated gradient line, as was performed on the right side of Figure 4. For reference, this figure has a vertical line marking the time shown in Figure 4.

This figure allows us to see the spatial gradients, and how they change over time. At the marked time on the humidity graph, the gradient is shown by the contrast between the dark top of the figure and the bright bottom. Now, we can see that the highly-curved humidity gradient persists for only a short while after the snapshot time and then disappears just as quickly as it arrived. We could see this brief movement in Figure 4, but now we can see that the dip only affects the upper sensors. Even though several other rapid changes in humidity can be seen to affect all heights of the tree equally, the presence of at least one such humidity gradient suggests that the dense canopy can have a buffering effect on the lower regions of the forest.

Looking at the rest of the data, we see that the top of the tree usually leads in any movement of temperature or humidity. However, as predicted by the spatial trend data in Section 5.4, the top of the tree can lag and the expected gradient can reverse. The reasons for these movements, and the effects of the newly-observable microclimate trends on the life of a redwood tree, are left as future work for plant biologists.

5.6 Outlier Rejection

When we first examined the data, we found many anomalous readings. Some sensors never produced readings in the normal range, or produced readings that did not track the other sensors. Other sensors performed normally, but produced erroneous readings prior to dying. After investigating many outlier detection methods, we found that battery failure was correlated with most of the outliers in the data. Figure 6 shows the death of 4 sensors and the effect on their temperature readings. Once the battery voltage falls from a maximum of 3 volts to a minimum of about 2.4 volts, a node’s temperature reading begins to rise far out of the normal range. Other nodes that did not produce any correct data at all were also running on very low batteries.

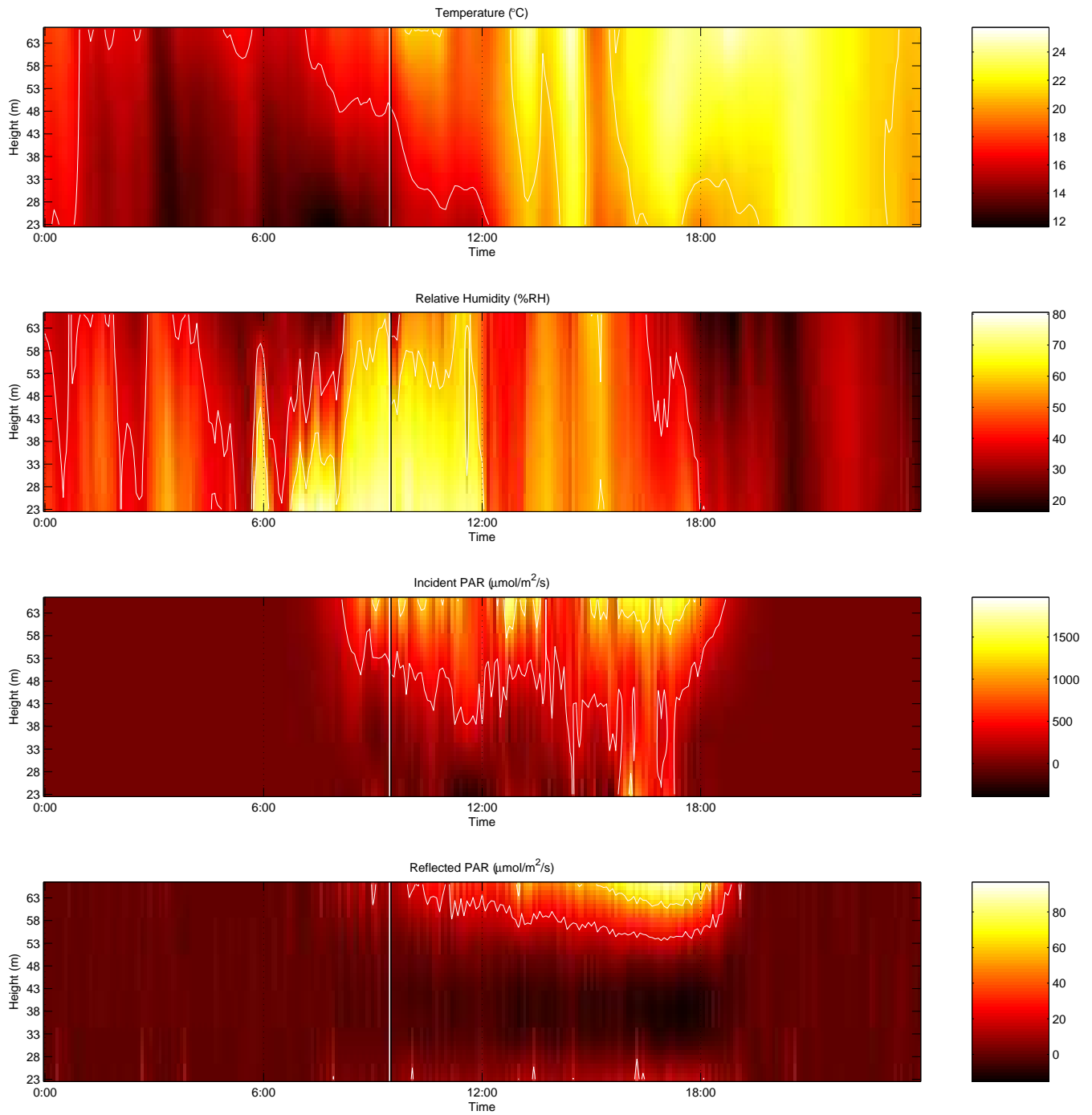


Figure 5: Change in spatial gradient over time

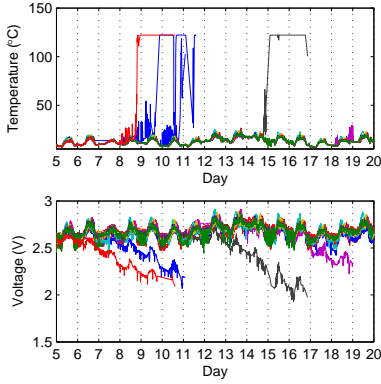


Figure 6: Temperature outliers are correlated with battery failure

The dependence of the reading on the battery voltage provided an excellent opportunity for automatic outlier rejection. Prior to performing the analysis whose results are displayed above, we removed all readings that were taken when the node’s voltage was higher than 3 volts or lower than 2.4 volts. This removed nearly all of the outlying points in our dataset. If the voltage reading is wrong, either the node is sensing incorrectly or the battery is dying. Either one suggests that the node’s data should not be trusted.

We also removed one node by hand after visually determining that its humidity sensor did not track the other nodes, and removed 3 other humidity readings that were above 100 %RH but did not correspond to voltage problems. Finally, we removed all sensors that did not produce any readings at all, which included some sensors whose readings had been entirely removed by the voltage screening.

5.7 Calibration Analysis

In this deployment, we found that calibrating the temperature and humidity sensors on the nodes offered a large relative increase in accuracy, but also found that the absolute increase was small enough to be of minimal benefit. As described in Section 4, we placed the motes in a controlled weather chamber prior to placing them in the tree. For each sensor, we gathered a range of temperature and humidity readings, then used linear regression to derive a two-point calibration equation. The mean of the temperature deviations from the mean of the sensors was $0.35\text{ }^{\circ}\text{C}$ prior to calibration, which suggests that the sensors are already quite accurate “out of the box”. Applying the calibration equations to the chamber data resulted in an improvement in the mean deviation from $0.35\text{ }^{\circ}\text{C}$ to $0.18\text{ }^{\circ}\text{C}$. A similarly small improvement was seen in the relative humidity sensors, as the mean deviation improved from 1.24 \%RH to 0.60 \%RH . These results suggest that the accuracy of the sensors used in a network should be studied, but the calibration itself may not dramatically improve the data quality for certain mature sensor types.

5.8 Data Yield Analysis

We can also use a multi-dimensional analysis technique to study the data delivery performance of the sensor network. For each mote at each timepoint, consider it to report a 1 if it reports any data at all, and to report a 0 if it does not

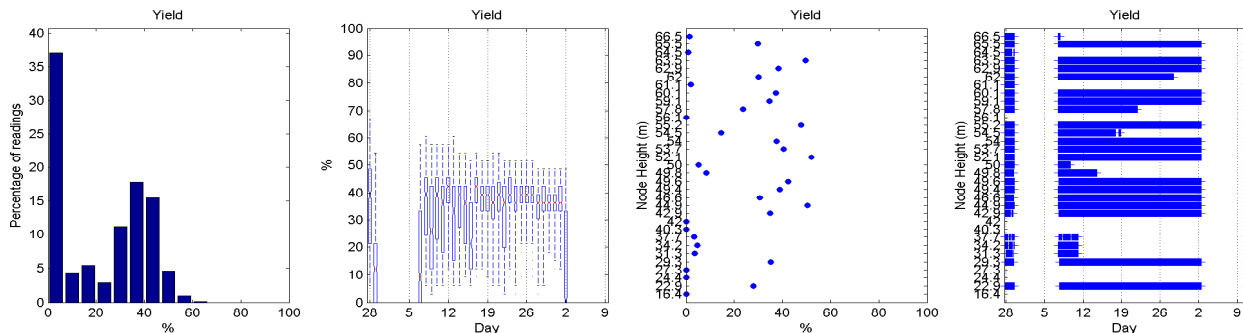
report any data. We can then calculate a yield for each time point by summing and dividing by the number of installed motes. We can also calculate a yield for each node by summing and dividing by the number of time points. It then becomes possible to examine the range of yields, and how the yield varies over time and space. Figure 7(a) shows a multi-dimensional analysis of the data yield.

As seen in many other deployments, data yield over the sensor network is often less than desirable, although the technology is clearly maturing. The most common yield percentage is, unfortunately, zero, though the distribution is bimodal with a second mode at 40% yield. In the second graph, we can see that all the times that have data have a median around 40%. We also see that all of the timepoints without any data are concentrated into 2 week-long periods. The period at the beginning resulted from an outage at the gateway, while the period at the end resulted from our process of periodically downloading the data stored in the gateway. June 2nd is simply the last day on which we took a snapshot of the data, as we had gone past the end of our original deployment envelope. In the third graph, we can determine that per-node yield is not at all correlated with height. In the fourth graph, we examine the lifespan of each individual node. Five nodes stopped functioning in the second week, and four more nodes died thereafter. However, the dying nodes only had the effect of lowering the maximum yield, and the median yield was unaffected.

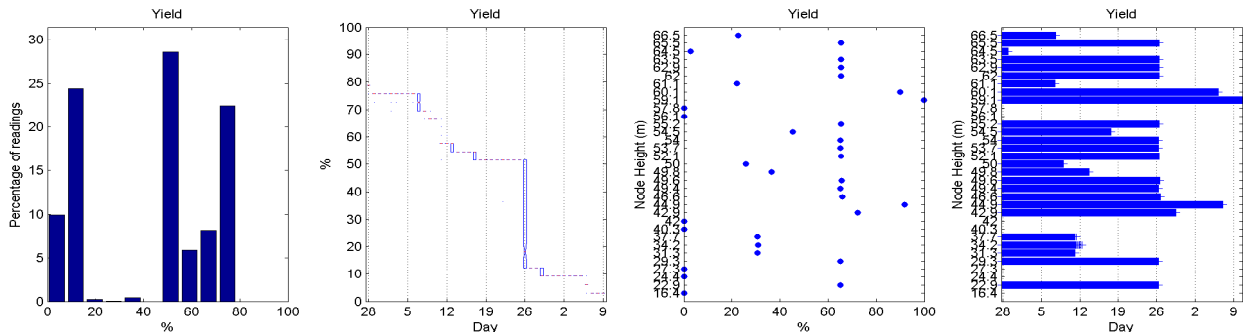
Previous deployments have shown that there is often a large difference between the data yield received over the network and the data yield offered by the nodes. We used the data recorded by the data logger to capture the offered yield so that we could make the comparison, and Figure 7(b) shows the same analysis applied to the data received from the log. In the log data, we see a strongly bimodal distribution, with peaks from 0-20% and 50-80%. The yield does not vary within any given day, indicating that each log was reliably recording data until the day on which the node died. We see that most nodes seem to “die” on May 26th. Actually, the logs filled up, and this explains the bimodality in the earlier distribution. In the third and fourth graphs, however, we can see that four nodes achieved higher than 65% yield because their logs did not fill up on May 26th.

The logs filled up early because each node was also recording data during the testing and calibration phases, lowering the amount of space available for deployment data. In addition, we did not clear the logs for all nodes prior to placing them in the tree, fearing that the effort required to disassemble each node, attach it to a PC, download a log-clearing program, and reassemble it would increase the likelihood of later node failure. However, some nodes received more testing than others, and thus had less available space at the start of the real deployment. This lesson suggests that logfile manipulation and other such node and network management functions should be accessible through the network itself.

It would have been possible to treat this deployment as a storage-free network of sensors in which all data is immediately transferred to the gateway. It would also have been possible to treat it as a collection of independent battery-powered data loggers with no communications capability, but neither design would have provided the best yield. We see that for the week after the log filled up, the nodes continued to return data over the network. However, once the network portion was finished, we obtained still more data



(a) Data received over the network



(b) Data stored in the local log

Figure 7: The network and the log perform differently, but neither one is perfect

from several of the logs. These effects both extended the usable range of the deployment data, suggesting that a well-designed sensor network should not just be a distributed data logger, but should not just be a network either.

6. DISCUSSION

Many lessons were learned from this close collaboration with biologists. We discuss two such lessons, and then outline some future work that can be performed with this data.

6.1 Lessons Learned

When the sensors get small enough and the phenomenon gets directional enough, tiny differences in positioning get magnified into large effects on the resulting data. This is especially noticeable in our PAR data. During an otherwise clear day, each mote’s PAR readings fluctuated between full sun and minimal light, and the pattern of each mote’s fluctuations were different. At first, we attributed this to the wind moving the foliage and blocking solar access to different motes at different times. In Figure 8, we actually see that the fluctuation pattern is nearly the same on two different days. This suggests that our nodes were actually seeing consistent patterns of light and shadow as the sun moved through subsequent days. Slightly different orientations for each light sensor resulted in different fluctuation patterns for each node, resulting in the seemingly “random” appearance of the light data. Our noisy data was actually a deterministic response by a highly focused sensor.

In addition to the challenge of physical installation, the success of a deployment depends crucially on the the management of the network. In the Section 5.8, we saw that

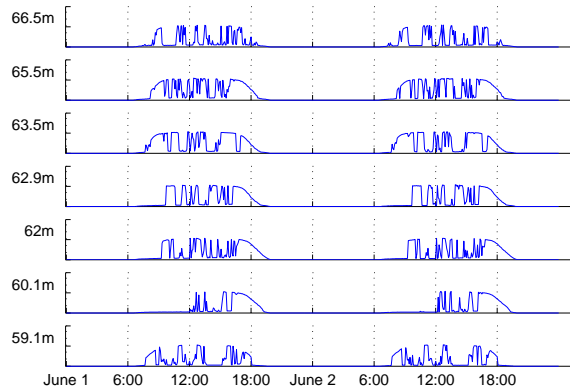


Figure 8: Fluctuations in PAR readings are similar on two different days

local logging compensated for failures in the network. Unfortunately, the local logs on most of the nodes ran out of space during the run, which was not discovered until the nodes were brought in from the tree. The TASK architecture did include the ability to remotely reboot the base station, but did not provide for comparable failure detection and response in the sensor network itself. More broadly, any long-term sensor network deployment should include a network monitoring component that can provide real-time information about the performance of the system, and can alert the researchers when the system begins to behave abnormally. The network can then provide a means to detect and compensate for failures in the logging, while the logging provides a means to compensate for failures in the network.

6.2 Enabling Biological Study

Having verified the existence of spatial gradients in the microclimate around a redwood tree, and having captured enough data to track the changes in these gradients over time, we can begin using this data to validate biological theories. For example, plant biologists know that the rate of sap flow through a tree varies over time, in response to humidity, air temperature, and PAR. This newly-obtained data from the macroscope could be used to build a quantitative model of the effect of microclimatic gradients on the sap flow rate. With a more detailed understanding of sap flow and transpiration, biologists can work toward understanding the large-scale processes of carbon and water exchange within a forest ecosystem.

7. CONCLUSION

The sensor network macroscope offers the potential to advance the state of science by enabling dense temporal and spatial monitoring of large volumes. We have demonstrated a real-world sensor network deployment that delivers on this promise. This particular instance of the macroscope has captured the complex environmental dynamics of the microclimate surrounding a coastal redwood tree. Extracting meaningful information from the large amount of data that can be produced by a successful sensor network deployment can be a challenging task, but a consistent framework of multi-dimensional data analysis helps to make the problem tractable.

8. ACKNOWLEDGEMENTS

This work was supported by the Intel Research Laboratory at Berkeley, DARPA grant F33615-01-C1895 (Network Embedded Systems Technology), the National Science Foundation Center for Information Technology Research in the Interest of Society (NSF CITRIS) grant EIA-0122599, and the California MICRO fellowship and project grant 04-021.

9. REFERENCES

- [1] M. A. Batalin, M. Rahimi, Y. Yu, D. Liu, A. Kansal, G. S. Sukhatme, W. J. Kaiser, M. Hansen, G. J. Pottie, M. Srivastava, and D. Estrin. Call and Response: Experiments in Sampling the Environment. In *Proceedings of the Second ACM Conference on Embedded Networked Sensor Systems*, 2004.
- [2] P. Buonadonna, D. Gay, J. M. Hellerstein, W. Hong, and S. Madden. TASK: Sensor Network in a Box. In *Proceedings of the Second IEEE European Workshop on Wireless Sensor Networks and Applications*, 2005.
- [3] R. Cardell-Oliver, K. Smettem, M. Kranz, and K. Mayer. Field Testing a Wireless Sensor Network for Reactive Environmental Monitoring. In *Proceedings of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 2004.
- [4] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao. Habitat Monitoring: Application Driver for Wireless Communications Technology. In *Proceedings of the First ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean*, 2001.
- [5] J. de Rosnay. *The Macroscope: a New World Scientific System*. Harper & Row, 1979.
- [6] D. Ganesan, B. Greenstein, D. Perelyubskiy, D. Estrin, and J. Heidemann. An Evaluation of Multi-Resolution Storage for Sensor Networks. In *Proceedings of the First ACM Conference on Embedded Networked Sensor Systems*, 2003.
- [7] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, and K. Pister. System Architecture Directions for Networked Sensors. In *Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-IX)*, pages 93–104, Cambridge, MA, USA, Nov. 2000. ACM Press.
- [8] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. The Design of an Acquisitional Query Processor for Sensor Networks. In *SIGMOD*, 2003.
- [9] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson. Wireless Sensor Networks for Habitat Monitoring. In *Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications*, 2002.
- [10] M. J. Mariscal, S. N. Martens, S. L. Ustin, J. Chen, S. B. Weiss, and D. A. Roberts. Light-transmission Profiles in an Old-growth Forest Canopy: Simulations of Photosynthetically-Active Radiation by Using Spatially Explicitly Radiative Transfer Models. *Ecosystems*, 7:454–467, 2004.
- [11] K. Martinez, J. K. Hart, and R. Ong. Environmental Sensor Networks. *IEEE Computer*, 38(8):50–56, August 2004.
- [12] P. C. Miller. Bioclimate, Leaf Temperature, and Primary Production in Red Mangrove Canopies in South Florida. *Ecology*, 53(1):22–45, Jan 1972.
- [13] U. Niinemets, O. Kull, and J. Tenhunen. Within-canopy variation in the rate of development of photosynthetic capacity is proportional to integrated quantum flux density in temperate deciduous trees. *Plant, Cell and Environment*, 27:293–313, 2004.
- [14] E. Osterweil and D. Estrin. Tiny Diffusion in the Extensible Sensing System at the James Reserve. <http://www.cens.ucla.edu/eoster/tinydiff/>, May 2003.
- [15] E. Osterweil and D. Estrin. ESS: The Extensible Sensing System. <http://www.cens.ucla.edu/eoster/ess/>, Apr. 2004.
- [16] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler. An Analysis of a Large Scale Habitat Monitoring Application. In *Proceedings of the Second ACM Conference on Embedded Networked Sensor Systems*, 2004.
- [17] R. Szewczyk, J. Polastre, A. M. Mainwaring, and D. E. Culler. Lessons from a Sensor Network Expedition. In *Proceedings of the First IEEE European Workshop on Wireless Sensor Networks and Applications*, 2004.
- [18] A. Woo, T. Tong, and D. Culler. Taming the Underlying Challenges of Reliable Multihop Routing in Sensor Networks. In *Proceedings of the First ACM Conference on Embedded Networked Sensor Systems*, pages 14–27, Los Angeles, CA, USA, Nov. 2003. ACM Press.