A wireless soil moisture smart sensor web using physics-based optimal control: Concept and initial demonstrations

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

As Published
http://dx.doi.org/10.1109/JSTARS.2010.2052918

Publisher
Institute of Electrical and Electronics Engineers

Version
Final published version

Citable link
http://hdl.handle.net/1721.1/62148

Terms of Use
Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.
A Wireless Soil Moisture Smart Sensor Web Using Physics-Based Optimal Control: Concept and Initial Demonstrations

Mahta Moghaddam, Fellow, IEEE, Dara Entekhabi, Senior Member, IEEE, Yuriy Goykhman, Member, IEEE, Ke Li, Mingyan Liu, Senior Member, IEEE, Aditya Mahajan, Ashutosh Nayyar, David Shuman, Member, IEEE, and Demosthenis Teneketzis, Fellow, IEEE

Abstract—This paper introduces a new concept for a smart wireless sensor web technology for optimal measurements of surface-to-depth profiles of soil moisture using in-situ sensors. The objective of the technology, supported by the NASA Earth Science Technology Office Advanced Information Systems Technology program, is to enable a guided and adaptive sampling strategy for the in-situ sensor network to meet the measurement validation objectives of spaceborne soil moisture sensors. A potential application for this technology is the validation of products from the Soil Moisture Active/Passive (SMAP) mission. Spatially, the total variability in soil-moisture fields comes from variability in processes on various scales. Temporally, variability is caused by external forcings, landscape heterogeneity, and antecedent conditions. Installing a dense in-situ network to sample the field continuously in time for all ranges of variability is impractical. However, a sparser but smarter network with an optimized measurement schedule can provide the validation estimates by operating in a guided fashion with guidance from its own sparse measurements. The feedback and control take place in the context of a dynamic physics-based hydrologic and sensor modeling system. The overall design of the smart sensor web— including the control architecture, physics-based hydrologic and sensor models, and actuation and communication hardware—is presented in this paper. We also present results illustrating sensor scheduling and estimation strategies as well as initial numerical and field demonstrations of the sensor web concept. It is shown that the coordinated operation of sensors through the control policy results in substantial savings in resource usage.

Index Terms—Control systems, in-situ validation, radar, radiometer, remote sensing, sensor webs, soil moisture, wireless networks.

I. INTRODUCTION

The long-term vision of Earth Science measurements involves sensor webs that can provide information at conforming spatial and temporal sampling scales, and at selectable times and locations, depending on the phenomena under observation. Each of the six strategic focus areas of NASA Earth Science (climate, carbon, surface, atmosphere, weather, and water) has a number of measurement needs, many of which will ultimately need to be measured via such a sensor web architecture. Here, we develop technologies that enable key components of a sensor web for an example measurement need, namely, soil moisture. Soil moisture is a measurement need in four out of the six NASA strategic focus area roadmaps (climate, carbon, weather, and water roadmaps) [1]. It is used in all land surface models, all water and energy balance models, general circulation models, weather prediction models, and ecosystem process simulation models. Depending on the particular application area, this quantity may need to be measured with a number of different sampling characteristics. It is therefore necessary to develop sensor web capabilities to enable flexible and guided sampling scenarios, as well as calibration and validation strategies to support them.

In-situ networks are used in the calibration and validation of remotely sensed variables [2]–[4]. Sparsity of network nodes, i.e., instruments, within the satellite footprint leads to differences between the satellite measurement and the in situ network estimates for the geophysical variable areal mean. This is particularly a problem for highly heterogeneous fields such as soil moisture. Soil moisture varies in space due to intermittency in precipitation, heterogeneity in soil type and vegetation cover, and in response to topographic redistribution. The NASA Soil Moisture Active and Passive (SMAP) mission [5], in particular, faces this problem in calibrating and validating its estimates of soil moisture. SMAP uses low-frequency microwave radar and radiometer to sense surface moisture conditions over global land surfaces.

The ground footprints of remote sensors such as SMAP are often coarser than the scale of variations of the variables they seek to measure. As a result, the remote sensing estimate is only a coarse-resolution representation of a field mean [6], [7]. A key challenge is how to calibrate and validate the satellite footprint estimate, for example from SMAP, which is an average of the field that may be tens or hundreds of km² for the radar and radiometer, respectively. This broad spectrum of variability and

Digital Object Identifier 10.1109/JSTARS.2010.2052918
multiple causes is not unique to soil moisture, but is a characteristic of many Earth system variables.

In-situ sensors often sample a point location in the heterogeneous field. The true mean of soil moisture fields is a function of time and of the state of the soil surface on a wide spectrum of scales ranging from meters (e.g., topography) to several kilometers (e.g., precipitation) [8]. Its determination requires a very fine sampling of the area within the satellite footprint, both spatially and temporally. This, however, is cost prohibitive; manually installing these sensors is expensive, and their battery power does not allow continuous sampling, as we need them to last a reasonably long period of time (months or even years). These considerations pose severe limitations on how many sensors can be installed, and how frequently they can be used/activated. The overall objective is thus to place and schedule the sensors so as to minimize a total expected cost consisting of the accuracy of the estimates of the surface-to-depth profiles of soil moisture and the energy consumed in taking the measurements.

There are two elements to the above problem; one is the determination of the best set of locations within the sensing field to place a limited number of sensors (sensor related cost constraint), and the other is the optimal dynamic operation of these sensors (when and which to activate) once they are placed, based on energy and accuracy considerations. These two elements are coupled. For instance, if energy of operation is a more dominant concern than placement costs, then one can choose to place more sensors to compensate for a desired, reduced sampling rate. The reverse may hold as well. In addition, activation and sampling decisions can influence where sensors should be placed and vice versa. But jointly considering and optimizing both elements leads to a problem whose complexity is prohibitive both analytically and computationally. We therefore decompose these problems and solve them sequentially. For the purposes of this paper, we assume pre-determined placements and focus on the control policy driven by physical models for sensors and time evolution of soil moisture fields. We still exploit the spatial correlations of soil moisture, in order to make the best estimates of previous measurements, as well as the uncertainties that are inherent in the soil moisture evolution and the sensor measurements. Recent studies such as [9] have considered ad-hoc approaches to sparse sampling of soil moisture, but the treatment presented in our work is based on rigorous optimization and control system theories.

The system architecture is as follows. The sensors are placed at multiple lateral locations. At each lateral location, multiple sensors at different depths are wired to a local actuator. The actuator is capable of wirelessly communicating with a central coordinator, and actuating the in-situ sensors. The central coordinator dynamically schedules the measurements at each location, transmits the scheduling commands to each local actuator, and subsequently receives the sensor measurement readings back from the actuators. The coordinator then forms an estimate of the soil moisture at all locations and depths, and schedules future measurements. A diagram of this control architecture is shown in Fig. 1. The physical implementation of this wireless communication and actuation system is discussed further in Section V.

The coordinator’s task is to leverage the spatial and temporal correlations of soil moisture, in order to make the best estimates of its evolution with as few measurements as possible. In order to do so, it must take into account the following:

1) the physics-based models of soil moisture evolution, which are described in more detail in Section III;
2) the soil moisture sensor models, which are described in more detail in Section IV;
3) all measurements to date;
4) ancillary data (e.g. rainfall, soil properties).

Fundamental issues in selecting a sensor configuration are the following:

- the energy consumption cost of different sensor configurations;
the expected effect of measurements taken with different configurations on the quality of the current state estimate;
• the expected effect of measurements taken with different configurations on future decisions for sensor configurations and their effect on quality of future state estimates.

Specifically, there is a fundamental tradeoff between the cost of taking a measurement and the expected gain from the information yielded by a measurement. Even if a measurement may improve current and future soil moisture estimates, the optimal decision may be to not take the measurement if it is too costly in terms of power usage.

We formulate the problem of choosing a sensor configuration and estimation strategy as a partially observed Markov decision problem (POMDP). We use the physics-based models of Section III to derive a statistical description of the soil moisture evolution. Namely, we model the soil moisture evolution as a first order Markov process, with transition statistics appropriately inferred from the physics-based models. The uncertainty included in the sensor models and the fact that we may not always take measurements implies that the underlying Markov process is only partially observed.

In addition to the physical and sensor models, the third key component of the POMDP is the performance criterion. The performance criterion consists of energy costs associated with each sensor configuration, and a distortion metric that measures the expected quality of the soil moisture estimates at each time.

Standard numerical methods exist for solving POMDPs [10]–[13]. These methods work well for small instances of our problem. For larger-scale instances, we have developed problem-specific techniques and approximations, which are discussed in detail in [14]. Section VI includes some numerical examples showing how generating scheduling and estimation strategies in this fashion allows the central coordinator to form good estimates of surface-to-depth profiles of soil moisture, while also conserving energy by taking sparser measurements.

III. SOIL MOISTURE EVOLUTION MODEL

As mentioned in the previous section, we model the time-evolution of soil moisture as a Markov process. In this section, we discuss two physical models of the evolution of soil moisture fields that serve as the basis for the development of the Markovian transition statistics. In one model the focus is on the temporal behavior and variations in soil depth. Because the vertical dynamics of soil moisture are governed by advection and diffusion with source/sink at the surface, the variance in soil moisture decreases with increasing depth. In the second model, we capture the heterogeneity in soil type and vegetation as well as redistribution over sloped landscapes, which cause significant spatial variations in the evolution of soil moisture fields. Descriptions of the two models and the context in which they are used are given below.

A. Temporal Evolution Model

The time and space evolution of soil moisture fields can be expressed via a pair of coupled partial differential equations (PDE) in space and time. This model has a number of parameters associated with soil characteristics and meteorological conditions. The solution to the coupled differential equation is an estimate of future states of soil moisture fields with the knowledge of the current state and the model parameters.

The Soil-Water-Atmosphere-Plant (SWAP) model [15] is a community standard solver for such a model, developed in the Netherlands. SWAP incorporates surface energy balance by including micrometeorological data such as precipitation, winds, air temperature, and humidity. It also incorporates soil physics properties such as amplitude and phase characteristics of flow dynamics. It then solves the coupled differential equations numerically.

The SWAP model is an open-source package and is publicly available. However, the inputs to the model are not provided, and therefore users must build their own interface to the core of the PDE solver. Once the user interface is built for defining the input parameters, the program can be run in an ensemble mode via another user-defined interface so that the statistical variations of the output can be investigated as a result of the statistical variations of the input parameters and variables.

We have built the user interface and simulated the SWAP model for some hypothetical scenarios. We performed comprehensive simulations of SWAP over long (up to 20 year) time periods for realistic environmental conditions. The example shown in Fig. 2 is taken from the results of SWAP simulated for a nominal location near Tampa, Florida, using actual rainfall and micrometeorological measurements. The soil surface is assumed bare. The results are shown for a representative 6-month interval (out of the 20 years simulated) at three different soil depths.

We observe that soil moisture is a strong function of rainfall, especially immediately after the rain events. This dependence is strongest at the surface, and diminishes for deeper locations. There is also a clear delay associated with moisture change at depth. For small amounts of rainfall, even the surface soil moisture is not significantly changed. Therefore, rainfall presents a trigger to discernible soil moisture change only for rainfall amounts exceeding certain values.

The variations of soil moisture also follow different patterns at different depths: the dissipation is much more rapid for shallower regions, and more damped at larger depths. Diffusion details depend on specifics of the area under study, such as soil texture, topography (assumed flat here), and vegetation cover. While it is now relatively straightforward for us to perform SWAP simulations with varying sets of rainfall and meteorological conditions (as were done in the example mentioned above according to actual observed data), we assume that rainfall is the only environmental variable while developing the sensor web control strategy. Note also that the variations of soil moisture are bound between roughly 5% and the saturation level of approximately 42% (volumetric). This information guides the selection of quantization levels of soil moisture during the development of the control algorithm. Extending the formulation of the problem to include other environmental variables such as temperature and solar radiation is out of the scope of this paper, but one that will be implemented in the future.

Once the soil moisture quantization levels are fixed, we use the SWAP simulations to determine the frequency of state transitions between quantization levels. These frequencies are used to generate a matrix of transition probabilities that describes the
Fig. 1. Control architecture. Each sensor measures variables over a finite period of time. Variables are correlated with the soil moisture field. Data are compressed at each sensor node and transmitted to the coordinating center, which derives an optimal control instruction set for the sensors, as well as soil moisture estimates.

conditional probability distribution of the soil moisture quantile at the next time, given the current soil moisture quantile. This matrix provides the statistical information of the physics-based soil moisture evolution model that is used by the controller.

B. Temporal and Spatial Evolution Model

In addition to the temporal dynamics of soil moisture along the depth of the soil column, the second model of soil moisture evolution takes into account the spatial redistribution of soil moisture along the lateral plane of a watershed. These spatial variations have different statistical behaviors depending on factors such as topography, soil texture, vegetation cover, amount of rainfall, and time after rainfall. Measurements such as those by Western and Grayson [16] confirm these dependencies.

We have started transitioning to a new soil moisture evolution model that is capable of simulating soil moisture dynamics across a three-dimensional field (as opposed to the depth-only SWAP model). The TIN-based Real-time Integrate Basin Simulator (tRIBS), under development at MIT since 2001, is such a model [17], [18]. The model includes the dissipative infiltration (like SWAP) of water through soil, but also incorporates gravity dominated lateral redistribution and overland and channel routing. The latest version of the model, tRIBS-VEGGIE, is capable of dynamic inclusion of vegetation canopy radiation interception and storage, and will be used in the future for inclusion of vegetation effects into the sensor web control strategy.

Fig. 3(a) shows a nominal 2 km × 2 km basin with arbitrary topography and drainage channels, used for tRIBS simulations. Fig. 3(b) and (c) shows sample time evolutions of soil moisture at different depths [Fig. 3(b)] and at different basin locations at a fixed depth [Fig. 3(c)]. The simulations have been performed for a nominal location whose climatology is consistent with Oklahoma.

The information from these and similar simulations are used to generate the spatially correlated Markovian statistics in the same way as described for the SWAP model. Note, however, that the resulting statistics represent both temporal and spatial correlations.

IV. SOIL MOISTURE SENSOR MODELS

Validation sensors make observations that are translated into estimates of unknown soil moisture. For a given observation time and for a given sensor, the sensor measurement is related to the value of the variable soil moisture via a physical model that includes sensor parameters. These parameters could be
frequency, polarization, power level, etc. Measurement noise is added to the true signal. Sensor models do not include any time evolution or dynamic nature. They can, however, include the probabilistic nature of the unknowns at any given time. The models and unknowns could be scalar (one dimensional) or vector (multidimensional), depending on how many variables are being measured and the number of sensors. Different sensors allow estimates of the unknowns such as soil moisture and surface roughness at different spatial scales. Sensors could be in-situ (moisture probes) or remote (tower-based, airborne, or spaceborne radars and radiometers).

In general, the estimation of unknowns is a complex task, depending on the degree of model nonlinearity, measurement noise, and sensor calibration. It is assumed that each sensor is calibrated independently of the rest of the sensors in the web.

Deriving physics-based remote sensor models to relate their measurements to estimates of soil moisture is generally rather complicated. The in-situ sensors, on the other hand, offer an opportunity for accurate measurements that are related to soil moisture values via simple empirical models. As an example, we have chosen to work with an in-situ soil moisture probe making highly localized measurements, namely, capacitance probes from Decagon, model ECH2O EC-5 (www.decagon.com).

These sensors are commonly used in the field, and have been used in previous wireless soil moisture sensing applications [19], [20]. The manufacturer provides standard calibration curves for these sensors with nominal accuracy of 1–2%, but due to variability of dielectric properties among different soil types, we decided to produce our own calibration curves. We developed the calibration curves through a standard procedure: we started with various soil samples that were fully dried, and added water in known proportions. With each addition of water (which results in a known value of moisture content), voltage measurements were taken. This produced a graph of probe voltage vs. water content, which was subsequently used to fit a polynomial. These polynomials, shown in Fig. 4 along with the probe and experimental data points, were used as the initial sensor model inputs to the control system. The models generated with the empirical data represent a calibration accuracy of better than 1%. We note, however, that depending on the type of soil, different calibration curves are obtained. Therefore, it is important to ensure proper sensor calibration in the field. If the wrong calibration curve is used, the sensor model (retrieved soil moisture value) could be in error by as much as 4% (Fig. 4), effectively amounting to measurement noise.

For remote sensors, which could be tower-mounted, airborne, or spaceborne, the physics-based retrieval models of soil moisture involve solutions to nonlinear optimization problems. Considering a tower-mounted radar as an example [21], its measured backscattering coefficients could be related to the profiles of soil moisture via models derived from Maxwell’s equations. A number of models that relate radar backscattering coefficients to soil moisture have been recently developed (the “forward” problem), including analytical [22] and hybrid...
analytical–numerical [23] models. An example of model simulations is shown in Fig. 5 [23].

The sensor retrieval (“inverse”) problem has also been solved for this application using various techniques. The basic strategy is to first simplify the sensor forward model to make it suitable for retrieval (inversion). In the case of analytic models, this task is already inherent in the solution. For the solutions that involve numerical techniques, the approach we have adopted is to derive multi-dimensional polynomial expressions that are derived from the more complicated numerical solutions. The closed-form nature of the resulting model allows us to apply a number of optimization techniques, both local and global. The statistical properties of the unknowns are systematically included in development of the optimization algorithm. Both of these classes of techniques are reported elsewhere [24], [25].

Extensive noise sensitivity analyses have been performed, an example of which is shown in Fig. 6 for a global optimizer to retrieve soil moisture [25] for two subsurface layers. The global optimizer used in this example is simulated annealing. In the figure, the expected model retrieval error (“noise” to the control system of the sensor web) is studied with the Cramer–Rao lower bound (CRLB). The CRLB is an indication of how well a variable (e.g., soil moisture) can be estimated from a sensor forward model in the presence of noise. As shown in Fig. 6 using 50 numerical experiments per point, the retrieval has low sensitivity to measurement noise. The retrieval errors are less than 4% for most cases studied, even in the presence of substantial noise [25]. We have assumed that the measurement channels (for example at different frequencies and polarizations) have the same noise statistics but that the noise is multiplicative and depends on the value of the sample measured. Details of the retrieval technique and the CRLB analysis can be found in [25].

V. WIRELESS COMMUNICATION AND ACTUATION SYSTEM

A. Design Requirements and Constraints

To achieve the objective of collecting surface-to-depth soil profiles at distributed locations, we need a network consisting of soil-moisture probes and ground wireless transceiver modules (referred to below as nodes or sensor nodes). The sensor nodes actuate and control the sensor probes and send collected data back to a base station. These devices are deployed in the field and are expected to operate for long periods of time (on the order of at least months) without direct human intervention. In this section we present Ripple-1, the ground wireless sensor node we designed for this project, as well as a ZigBee based wireless communication network we designed using Ripple-1.

Our system shares some of the requirements common to many other systems. These include long lifetime, high reliability, ease in deployment and maintenance, ability to support multi-hop communication, scalability, and relatively long-range
wireless communication. Long-range for our application means distances on the order of hundreds of meters to a mile, as we need to cover a sufficiently large area to be able to observe spatial variability in soil moisture.

In addition, our system has the following distinguishing features. First, in terms of data flow, it operates in a “data pull” mode rather than a “data push” mode, since the measurement decision is made at the base station using antecedent data and \textit{a priori} statistical information. This makes many data push (or clock-driven or event-driven data collection) paradigms [26] unsuitable. Our sensor nodes need to be highly responsive to base station commands. At the same time, our system potentially has a very wide range of sampling and data rates, sampling from once per minute to once per hour or tens of hours depending on exogenous weather conditions and antecedent moisture values. Both of these features make duty cycling mechanisms very challenging to design.

Finally, we want to have a low-cost design and a relatively easy-to-maintain system. Some of the more specific requirements include: large network size (more than 30 nodes) and extensibility; low cost (< $100 per node) and small form; up to eight sensor channels on each node to enable measurement of soil moisture at multiple depths, as well as temperature, precipitation, or other environmental variables; and the ability to work in extreme temperature environments.

The requirements listed in the preceding three paragraphs rule out most (if not all) of existing sensor platforms available on the market. These include MICA2 [27], TelosB [28], BNode [29], and Fleck 3 (CSIRO ICT Centre), to name a few, which, in particular, do not meet the requirement for long-range operation. We used the Narada [30] sensing and actuation boards to collect some initial results in the early phase of this project. However, it was too energy-consuming and had insufficient communication range for our scenario.

B. Ripple-1 System Overview

Fig. 7 shows the architecture of a Ripple-1 system. At the network level, the system consists of a number of sensor nodes deployed over a target field, a base station that performs data collection and sensing control, also deployed in the field, and an off-field database used to store data that also allows remote data access, e.g., from office/home or on the move. At each sensing site (where a sensor node is placed), a number (3–5) of moisture probes are also deployed vertically underground with wire connection to the sensor node on the ground. This forms the configuration of a single sensor location.

A web site (hosted on a server on the U. Michigan campus) has been developed to provide an interface for users to access and visualize collected data, and to override scheduling algorithms run on the base station. The connection between the base station, the database, and the web server is through a 3G Internet card installed on the base station. Thus any device with Internet access, including PCs and smart phones can browse the web server and access data and control.

C. Sensor Network Operation

In searching for a low-power, low-cost, reliable, and multi-hop solution, we converged on the ZigBee technology [31]. Currently, ZigBee is the only standards-based technology on the market that targets low-cost and low-power networking applications (e.g., home networks). It is built on the IEEE 802.15.4 standard that specifies the physical (PHY) and media access control (MAC) layers. Specifically, ZigBee specifies the network, security, and application layers, and defines three types of logic devices:

- Coordinator: this is the most capable device that establishes the network and assists in routing data. A single network only has one coordinator.
- Router: it supports data routing and can talk to the coordinator, end devices, and other routers.
- End device: it has just enough functionality to talk to its parent node (either the coordinator or a router).

The topology of a typical ZigBee network can be a star, mesh or cluster tree (also called star-mesh hybrid). Our field-deployed network is shown in Fig. 8; it consists of a single coordinator/base station, 2 router nodes, and 11 end devices.

D. Node Design

Having identified ZigBee as the network solution, we surveyed currently available chips for building our sensor node. Among these, we decided that the XBee PRO ZB module by Digi International [32] (this design is based on EM250 system-on-chip from Ember) is a good candidate that has relatively long battery life, is reliable, low-cost, and industry-standard. The key characteristics of this module are shown in Table I.

To provide superior communication range (up to 1 mile), the XBee PRO ZB module is equipped with a built-in low noise amplifier and a power amplifier. An XBee PRO ZB module with different firmware versions can act as one of the three logic device types in a ZigBee network.

- **Coordinator**: Of the three types of logic devices, the coordinator is the most straightforward to set up. The coordinator in our ZigBee network is essentially wire-connected...
TABLE I
KEY CHARACTERISTICS OF XBEE PRO ZB MODULE [32]

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range (outdoor line-of-sight)</td>
<td>up to 1600m</td>
</tr>
<tr>
<td>Supply Voltage</td>
<td>3.0 - 3.4V</td>
</tr>
<tr>
<td>Tx Peak Current</td>
<td>295mA (at 3.3V)</td>
</tr>
<tr>
<td>Rx Current</td>
<td>45mA (at 3.3V)</td>
</tr>
<tr>
<td>Idle Current</td>
<td>15mA</td>
</tr>
<tr>
<td>Sleep Current</td>
<td>&lt; 10μA</td>
</tr>
</tbody>
</table>

Fig. 9. End node design block diagram.

(through a USB port) to the base station computer. It builds the ZigBee network and ensures information flows from base station to the network and back.

- **End Device:** The end-device is the most challenging part of the design. The XBee PRO module can be battery-powered and can have a typical lifetime of several months to a couple of years (depending on the system). In our case, to support the long radio range (up to a mile) and heavy load (up to 8 channels that can support soil moisture probes, precipitation and temperature sensors, etc.) it is necessary to attach renewable energy sources (solar) and/or rechargeable batteries to each node.

- **Router:** In our network the router has basically the same features as the end device, but with a larger solar panel and larger rechargeable batteries. In the next version of the Ripple system (currently under development), more sophisticated sleep scheduling mechanisms will significantly reduce a router node’s power consumption, whereby allowing it to operate using exactly the same battery and energy solutions as an end device.

During regular operation, an end device alternates between high-powered active periods and low-powered sleep periods, the latter of which account for about 99% of the time. As the on-board radio and sensors consume most of the power, our electrical design principle is to set into the low power mode or power off all components during sleep periods.

The hardware block diagram for a Ripple-1 node is shown in Fig. 9. During sleep periods, voltage regulator 1 and Xbee PRO module are set into low power mode; voltage regulator 2, analog switch, and sensors are powered off.

We measured the power consumption of Ripple-1 node without sensors, which was around 95 mW in active mode (including Tx, Rx, and idling modes) and 0.18 mW in sleep mode despite the addition of amplifiers in the XBee PRO ZB module. With sampling rate at two samples (1.5 seconds in active mode) every ten minutes (an obvious over-estimate for our application), the daily energy requirement of the node is about 10 mWh. Equipped with sensors, the total consumption of a node will be slightly higher than this number. Energy storage elements that can provide energy for more than 30 days of operation without recharging are practical candidates for Ripple-1 nodes.

In addition to energy capacity, we also considered other characteristics including lifetime, charging method, safety, size, environmental aspects, and cost. We then narrowed down choices to the following types: nickel metal hydride (NiMH), lithium ion (Li-ion) and super capacitor. Compared to rechargeable batteries, the energy density of super capacitors is very low, making them insufficient for our nodes. Li-ion batteries were also ruled out because of complicated charging method.

On the other hand, common NiMH batteries present two major drawbacks as well. The first is a high self-discharge rate of 30% per month, but this problem has been solved by Sanyo in their NiMH battery design that has a less than 10% per month self-discharge rate [33]. The second is the low charging efficiency of roughly 66%. This means that NiMH batteries store only 2 out of 3 units of input energy. Our solution is to use a relatively high power solar cell. Our final battery and solar cell selection are shown in Table II.

Two fully charged AAA NiMH batteries with 800 mAh capacity in series provide 1920 mWh energy. In other words, even having self-discharge rate of 10% per month, the two batteries can provide energy for around 4 months for a node without sensors and without charging. Other options such as supercapacitors are not appropriate for our application due to their low energy density and high self-discharge rate.

A picture of the completed module, along with the battery pack and solar cell, is given in Fig. 10.

We next consider the lifetime issue and explain the energy management mechanism in Ripple-1. Rechargeable batteries have a finite lifetime. The batteries’ lifetimes directly determine the lifetime of the system, and it is therefore important to maximize the batteries’ lifetimes. The natural aging process of rechargeable batteries is gradual and results in a gradual reduction in capacity over time. Battery manufacturers often provide cycle life as the aging parameter of a battery product; this is defined as the number of complete charge-discharge cycles a battery can perform before its nominal capacity falls below 80% of its initial rated capacity [34]. However, such
cycle life estimates are obtained from standard aging tests that do not capture many factors that in practice influence the life of a battery. The most important factors are extreme temperatures, overcharging/over-discharging, rate of charge or discharge, and the depth of discharge (DOD) of battery cycles [35]–[37]. We discuss some of the more relevant ones below, and in doing so explain the reasons behind the energy management mechanism in Ripple-1.

E. Overcharging/Over-Discharging

In our scenario, both charging current and discharging current are relatively small. A small amount of overcharging or over-discharging will not cause premature failure of the batteries but can significantly shorten their lives [38]. For example, tests show that continuously over-discharging NiMH cells by 0.2 V can result in a 40 percent loss of cycle life [34].

As mentioned above, the two batteries can provide a node with energy for around 4 months, without charging. Therefore, as long as we charge the battery to a relatively high level each time we charge, over-discharging is unlikely. To avoid overcharging, we charge the batteries to about 90% of the state of charge (SOC), the capacity ratio remaining in a battery [39].

F. Depth of Discharge

Depth of discharge (DOD) is the ratio of the quantity of electricity (usually in ampere-hours) removed from a battery to its rated capacity [39]. The DOD is the inverse of SOC: as one increases, the other decreases. For example, the DOD is 0% for a fully charged battery, and 100% for an empty battery. Tests show that the number of cycles yielded by a battery is exponentially decreasing in the DOD; this can be seen in Fig. 11 [34].

In this example, the battery can be used for 15,000 cycles if it is discharged by 5% in each cycle, and 7000 cycles if the DOD is 10%, but only 500 cycles if the DOD is 100%.

We would like to maximize the energy throughput of a battery during its lifetime, which means we need to maximize the total amount of energy taken out of a battery over all the cycles in its lifetime. Total energy throughput can be calculated by the product of DOD in each cycle and total cycles of the battery. In this example, suppose the capacity of the battery is C, then the energy throughput of 5% DOD in each cycle is 750 C, 10% DOD is 700 C and 100% DOD is only 500 C. For this reason, we decided to restrict the possible DOD in each cycle, so as to improve the total energy throughput of the batteries.

Combining the above, our overall battery management strategy is shown in Fig. 12. One practical challenge is that the SOC is both difficult and costly to measure exactly. We therefore use the discharge curve (voltage vs. SOC) of the batteries to measure the SOC by reading the output voltage of the batteries. This is because it is relatively simple and reliable to design a voltage-controlled charging circuit.

The measured discharge curve of two Sanyo NiMH AAA batteries in series with 50 mA discharging current at 21°C; the results are shown in Fig. 13. The output voltage of the batteries is between 2.7 V to 2.8 V when their SOC is 90%. We therefore designed our circuit to charge the batteries up to that value.

Fig. 14 shows the rechargeable battery voltage level of a node placed outside over a representative period of three days, with a sampling interval of 5 minutes. The final version of the Ripple-1 node, including a weatherproof enclosure, is shown in Fig. 15. This module is field-deployable and has been tested and used in initial demonstrations of our sensor web technology.
VI. SIMULATIONS AND FIELD EXPERIMENTS

A. Numerical Simulations

We consider an arrangement of in-situ sensors at two different locations, and three depths (25, 67, and 123 mm) at each location. We obtain soil moisture evolution statistics for these locations from the tRIBS model described in Section III. We assume that when a measurement is scheduled at a given location, the sensors at all three depths are used. We also assume that the time between measurements at each location cannot exceed 30 time steps. The objectives are to conserve energy and estimate the soil moisture at both locations and all three depths. We assume a nine-level quantization of soil moisture at each location/depth pair, and penalize estimation errors by the absolute difference between the quantile index of the true moisture and the estimated quantile index. Relative to one unit of estimation error, the energy cost of taking measurements at all depths at a given location is 1.5. We assume these measurements are noiseless. We use a discount factor of 0.95, and a time horizon of 200 steps.

We consider three different scheduling and estimation strategies. The first is to take measurements at both locations at every time step. The second is to optimally schedule the sensors and estimate the soil moistures at each location independently of the other location. The third is to optimally schedule the sensor measurements at each location independently of the other location, but to have the coordinator jointly estimate all soil moisture quantiles using the measurements from both locations. The resulting expected costs are shown in Table III.

Note that the second and third strategies result in an over 80% reduction in the number of measurements, as compared to a continuous sampling strategy. With the relative weight of the estimation and measurement costs used, this reduction results in a significant improvement in the total expected cost. The above example also demonstrates that the coordinator can reduce the expected estimation cost by leveraging not only the correlations of soil moisture at different depths at the same location, but also the correlations of soil moisture across different locations. This explains the reduction in expected estimation cost between the second and third strategies, despite the fact they both call for taking the same number of measurements.

B. Field Experiments

To test and validate all aspects of the new sensor web technology, we deployed a network of in-situ ECH2O EC-5 soil moisture sensors at the University of Michigan Matthaei Botanical Gardens in Ann Arbor, Michigan. The sensors were arranged at seven locations (nodes) throughout the field, covering a range of up to 250 m. At each location, soil moisture was sampled at up to three distinct depths (25 mm, 67 mm, and 123 mm).
Fig. 16. Aerial view of field validation site at the University of Michigan (UM) Matthaei Botanical Gardens. The seven nodes are shown by yellow asterisks.

Fig. 17. Sample soil moisture data measured at UM Matthaei botanical gardens using ECH2O EC-5 probes, showing the time and depth variations of soil moisture after three rain events.

Fig. 18. Performance of the closed loop sensor web, as measured by the accuracy of soil moisture estimates. Figure shows the comparison between the true values of soil moisture from continuous time samples (red) and values estimated by the sparse samples of the sensor web (blue).

| TABLE IV |
| COMPARISON OF DIFFERENT CONTROL AND ESTIMATION STRATEGIES USING DATA. SENSORS ARE AT 2 DIFFERENT LOCATIONS AND 2 DEPTHS AT EACH LOCATION. THE PARAMETERS ARE THE SAME AS DESCRIBED IN THE NUMERICAL EXPERIMENTS SECTION |

<table>
<thead>
<tr>
<th>Control and Estimation Strategies</th>
<th>Total expected cost</th>
<th>Expected measurement cost</th>
<th>Expected estimation cost</th>
<th>% of time measurement taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take measurement at every sensor at every time</td>
<td>57.0</td>
<td>57.0</td>
<td>0.0</td>
<td>100%</td>
</tr>
<tr>
<td>Independent measurement scheduling and estimation</td>
<td>33.4</td>
<td>7.5</td>
<td>25.9</td>
<td>13%</td>
</tr>
</tbody>
</table>

Fig. 16 shows an aerial view of the field site, with sensor node locations identified.

Initially, eight sensors at three lateral locations were used to collect a near-continuous record of soil moisture variations. The reason for collecting these data was to calculate the moisture evolution statistics for this particular location, instead of relying on simulations as was done initially in the development of the sensor control policy. Fig. 17 shows an example of time variations of soil moisture at three depths at one of the locations.

Using the field soil moisture data, we derived the true transition probabilities matrix for this site, and subsequently used it to derive the sensor control policy. Fig. 18 shows the performance of the closed loop sensor web system, as measured by the accuracy of the values of soil moisture estimates using the sparse measurements, compared with the true measured values. For brief periods following rainfall, soil moisture changes rapidly and the sparse measurements produce inaccurate estimates. But after a short time, the estimates recover and become quite accurate. This problem will be mitigated in future versions of the control algorithm by an automatic dense sampling policy triggered by a rainfall sensor, collecting dense field samples to collect more comprehensive statistics and therefore a more robust control policy, or implementing higher-order Markovian models.

We used the first half of the field soil moisture data to derive the transition probabilities matrix, and the second half to test the control policy. Table IV shows the performance of the closed loop sensor web system. As with the numerical simulations, the performance criteria consist of energy costs associated with measurements and the distortion costs reflecting the accuracy of the soil moisture estimates.

VII. CONCLUSIONS: VIEW TO THE FUTURE

The technology introduced here for integrating a physics-based modeling framework into a sensor web control system to achieve a dynamic and sparse sampling strategy is fundamentally new. The sensor web considered here aims to use in-situ sensors to sample three-dimensional soil moisture fields as a function of time, as part of a validation system for future large-footprint satellite observations of soil moisture. The NASA SMAP mission is the primary target application for this technology, where it is envisioned that sparse sampling of
soil moisture by the in-situ sensors will provide the required validation data at a minimum cost.

We have shown that it is not necessary for the sensors to collect data continuously (or with dense time sampling), but rather they can take sparse measurements. The measurement schedule is based on prior statistics of soil moisture evolution, rainfall, and the antecedent data from the sensors. The accuracy of the antecedent data has a direct impact on the effectiveness of the control policy. We have shown that the in-situ sensors are highly accurate and can be considered noise-free if well calibrated. The measurement schedule (“policy”) is derived through rigorous concepts of optimal control and delivered to the in-situ sensors via a wireless communication and sensor actuation system.

We have developed the wireless communication and actuation system using COTS components but through a novel system design that has optimized power handling, sleep cycling, cost, robustness, and communications range. The wireless system, named here as Ripple-1, has been fabricated and field-tested.

We have tested the closed-loop operation of the entire system at the UM Matthaei Botanical Gardens, and verified the utility of (1) the control system in generating sensor scheduling policies, (2) the Ripple-1 system in delivering the control policy to the sensors and actuating them, and (3) the on-demand sensor control and data transmission via the same wireless link.

This closed-loop sensor web is under continued development, and has a number of improvements planned before it is made fully operational.

• We continue to enhance the computational capabilities of the control system to enable the control of larger and larger number of sensors.

• We are in the process of developing an optimal placement policy for the sensors within the landscape, instead of assuming an arbitrary placement.

• We are enhancing the multiphot features of the Ripple-1 node so that the router has more efficient power handling capability.

• We continue to improve our sensor retrieval models so that the sensor data fed to the control system have as little noise as possible.

We note that the methodology for data collection and data processing described here is also applicable to several other technological areas including transportation systems, wireless sensor networks, and Mobile and Ad hoc Networks.

ACKNOWLEDGMENT

The authors would like to thank A. Flores of Boise State University for providing tRIBS simulations.

REFERENCES


Yuriy Goykhman (S’06–M’10) received the B.S. degree in electrical and computer engineering in 2005 from Carnegie Mellon University. He received the M.S. degree in 2007 from the University of Michigan, where he is currently working on his Ph.D. degree.

His research interests include forward and inverse scattering, radar systems, radar data processing, and sensor models.

Ke Li received the B.S. degree in mechanical engineering from Beijing University of Aeronautics and Astronautics, Beijing, China. He is currently pursuing the Ph.D. degree in the Department of Precision Instruments and Mechatronics at Tsinghua University, Beijing, China.

He is also currently a Visiting Researcher in the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. His research interests include architectures, protocols, and performance analysis of wireless networks.

Mingyan Liu (M’00) received the B.Sc. degree in electrical engineering in 1995 from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, and the M.Sc. degree in systems engineering and Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 1997 and 2000, respectively.

She joined the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, in September 2000, where she is currently an Associate Professor. Her research interests are in optimal resource allocation, performance modeling and analysis, and energy efficient design of wireless, mobile ad hoc, and sensor networks.

Dr. Liu is the recipient of the 2002 NSF CAREER Award, and the University of Michigan Elizabeth C. Crosby Research Award in 2003. She serves on the editorial board of IEEE/ACM TRANSACTIONS ON NETWORKING and IEEE TRANSACTIONS ON MOBILE COMPUTING. She is a member of the ACM.

Aditya Mahajan (S’06–M’09) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, India, in 2003, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of Michigan, Ann Arbor, in 2006 and 2008, respectively.

He is currently an Assistant Professor of electrical and computer engineering, McGill University, QC, Canada. From 2008 to 2010, he was a Postdoctoral Researcher in the Department of Electrical Engineering at Yale University, New Haven, CT.

His research interests include decentralized stochastic control, team theory, real-time communication, information theory, and discrete event systems.

Ashutosh Nayar received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 2006, and the M.S. degree in electrical engineering and computer science in 2008 from the University of Michigan, Ann Arbor, where he is currently working towards the Ph.D. degree in electrical engineering and computer science.

His research interests include decentralized stochastic control, stochastic scheduling and resource allocation, game theory, and mechanism design.
David Shuman (M’10) received the B.A. degree in economics and the M.S. degree in engineering-economic systems and operations research from Stanford University, Stanford, CA, in 2001. He also received the M.S. degree in electrical engineering: systems, the M.S. degree in applied mathematics, and the Ph.D. degree in electrical engineering: systems from the University of Michigan, Ann Arbor, in 2006, 2009, and 2010, respectively.

He is currently a Postdoctoral Researcher at the Institute of Electrical Engineering at Ecole Polytechnique Fédérale de Lausanne in Lausanne, Switzerland. His research interests include stochastic control, stochastic scheduling and resource allocation problems, energy-efficient design of wireless communication networks, and inventory theory.

Demosthenis Teneketzis (M’87–SM’97–F’00) received the diploma in Electrical Engineering from the University of Patras, Patras, Greece, in 1974, and the M.S., E.E., and Ph.D. degrees, all in electrical engineering, from the Massachusetts Institute of Technology, Cambridge, in 1976, 1977, and 1979, respectively.

He is currently a Professor of electrical engineering and computer science at the University of Michigan, Ann Arbor. In winter and spring 1992, he was a Visiting Professor at the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. Prior to joining the University of Michigan, he worked for Systems Control, Inc., Palo Alto, CA, and Alphatech, Inc., Burlington, MA. His research interests are in stochastic control, decentralized systems, queueing and communication networks, stochastic scheduling and resource allocation problems, mathematical economics, and discrete-event systems.