Title

Rapid Deployment with Confidence:Calibration and Fault Detection in Environmental Sensor Networks

Authors
N. Ramanathan
L. Balzano
M. Burt
et al.

Publication Date
2006

Rapid Deployment with Confidence: Calibration and Fault Detection in Environmental Sensor Networks

Nithya Ramanathan[†⊎], Laura Balzano[†⊎], Marci Burt[†], Deborah Estrin[†], Tom Harmon[†], Charlie Harvey[‡], Jenny Jay[†], Eddie Kohler[†], Sarah Rothenberg[†], and Mani Srivastava[†]

[⊎]Primary Authors

[†]Center for Embedded Networked Sensing, UCLA

[‡]Department of Civil and Environmental Engineering, MIT

{nithya, kohler, destrin}@cs.ucla.edu, {sunbeam,mbs}@ee.ucla.edu
{marciburt,srothenb}@ucla.edu, tharmon@ucmerced.edu, jjay@seas.ucla.edu
charvey@mit.edu

July 4, 2006

ABSTRACT

Rapidly deployable sensor networks are portable, reusable, and can take advantage of a human user in the field attending to the deployment. Unfortunately, even small disruptions or problems in collected data must be addressed quickly, as the overall quantity of data gathered is small relative to long-term deployments.

In this paper we describe a procedure for calibration and a system for online fault remediation. Care in the calibration process for ion selective electrodes used for water quality assists interpretation of the data. Scientists will have more confidence in the data obtained from a rapid deployment if in-field users can detect and compensate for problems as they occur. We have designed and implemented a tool for use in the field to detect potential faults and provide actions to remedy or validate the faulty data. In January of 2006 we deployed 48 sensors over a period of 12 days in Bangladesh in order to aid in validating a hypothesis on the mass presence of arsenic in the groundwater. Our system is based on the the approximately 25,000 measurements we collected.

1 INTRODUCTION

The presence of arsenic in groundwater has led to the largest environmental poisoning in history; tens of millions of people in the Ganges Delta continue to drink groundwater that is dangerously contaminated with arsenic. In Bangladesh alone, if consumption of contaminated water continues, the prevalence of arsenicosis and skin cancer will be approximately 2,000,000 and 100,000 cases per year, respectively, and the incidence of death from cancer induced by arsenic will be approximately 3,000 cases per year [23].

A current working hypothesis is that the influx of dissolved arsenic into the ground water is greatly enhanced where irrigation for rice cultivation provides the primary source of aquifer recharge [15]. To aid in validating this hypothesis, we accompanied a group of scientists from MIT, Stanford, and the Bangladesh University of Engineering and Technology, to undertake a rapid deployment of a wireless sensor network in a rice paddy in Bangladesh in January of 2006. We deployed 48 sensors over a period of 12 days, collecting approximately 25,000 measurements. The deployment setup is illustrated in Figure 1.

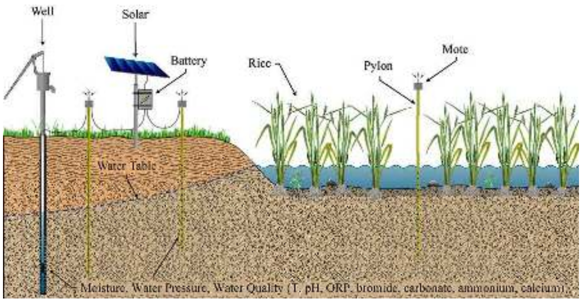

Figure 1: Depiction of deployment (drawing by XXXXXX).

Rapid deployment We discuss this Bangladesh experiment as a case study in the *rapid deployment* of a wireless sensor network (WSN). This model, which holds great promise for environmental monitoring, has emerged as one alternative to the traditional long-term, autonomous, and static WSN deployment model. Rapidly deployable networks are designed to be quick and simple to deploy; also they may only be left in place for a relatively short period of time [2]. Water quality sensing can benefit greatly from rapidly deployed sensor networks. Although good water quality is critical for public health, "analysis is still primarily conducted in a laborious manner by physical collection of a sample that is analyzed back in a laboratory." [22] This kind of data collection and analysis is time consuming, mostly undirected, and, in many instances, misses the contaminant events of interest. While a long-term deployment could simplify collection, it would not be able to respond quickly to contamination events, and would be expensive and difficult to maintain.

Rapidly deployable sensor networks have several advantages. First, because the deployment duration is relatively short, on the order of days or weeks instead of months, a user may continually be present during the course of the deployment. This human in the loop enables enhanced flexibility and network functionality challenging or impossible to achieve within the confines of complete network autonomy; for example, a user can manually replace a broken node or sensor. Second, sensors that require frequent calibration and maintenance can be infeasibly expensive to deploy for extended periods of time. Minimizing the deployment time can facilitate the maintenance and replacement of such sensors.

However, there are many issues to address in order to achieve a robust rapid deployment.

Detecting and remediating faults Due to the short duration of rapid deployments, in many instances the focus is on collecting a sufficient amount of useful data as quickly as possible. Thus, all faults affecting the *quantity* and *quality* of data should be immediately addressed and fixed to minimize their impact. Reliable data transmission [3] increases the quantity of recovered data, but ensuring that scientists have confidence in the quality of the data collected by the network is more challenging.

A large variety of faults can impact data quality, including sensors affected by aging, biofouling, or leaking of internal solution, or simply sensors with a bad or no connection to the board. Moreover, during a deployment, often there is too much data to manage; manually detecting faults online can be difficult. Our experience in Bangladesh exemplifies these issues. Many of our sensors reported data outside of their operating range. In addition, data from many sensors were punctuated with anomalous patterns, some that are physically impossible. Despite careful pre-deployment preparation, we were not able to detect all of these issues in the field. A systematic *in-field* approach is required, one that helps a user find and fix faults as quickly as possible in the limited time available.

In this paper, we describe two ways to increase confidence in data from a rapid deployment for environmental monitoring. Careful pre-deployment calibration procedures are crucial for correctly interpreting data from ion selective electrode sensors. Mistakes at this stage can result in data interpretation errors. More importantly, online fault detection can allow the user to alleviate system faults in the field. Our system identifies unreliable data along with an associated remediating action that is likely to fix the fault's cause. The goal of the system is *not* to identify all unreliable data in the field; instead it only identifies data that has an associated validating or remediating action. However, as we discuss further in Section 6, our system associates an action with most data identified as unreliable by an independent system.

Our contributions are three-fold. First, we describe procedures we followed before leaving for Bangladesh for calibrating and testing the entire system extensively (Section 4). Our second contribution is the fault-detection and remediation system we developed upon returning from Bangladesh (Section 6). The system applies a simple set of rules, identifying anomalous data patterns that indicate potential faults in the system. Each rule has an associated action that can help the user remedy the fault or validate the data. Our third and final contribution is that of our experience with a rapidly deployed sensor network to gather water quality data in Bangladesh. We describe the application and sensor network in Section 3, and the experience in Bangladesh itself in Section 5.

Validation Before beginning, it is necessary to spend a moment on the issue of data validity. Our system identifies faults associated with almost 40% of the data points we collected, but there is no guarantee that those faulty points are actually unreliable. We argue that this issue—a lack of ground truth—is common to sensor deployments, and that a system such as ours will help develop consensus on expected behavior over time.

We carefully calibrated and tested our sensors before going to the field. This step was critical for data interpretation, but in the field the sensors frequently exhibited quite different behavior than in our pre-deployment testing. What can we do with the resulting data? How do we know when the network is reporting garbage, and when it is reporting unusual phenomena worthy of study?

Sensor networks are deployed precisely to produce data that was previously unobservable; since this data, such as our soil chemistry readings, was difficult to obtain before with equal spatial and temporal density, consensus on ground truth—what the network *should* report—remains unachieved. Additionally, as WSN technology is in its infancy, best deployment practices for obtaining reliable data have not yet been developed. Even as WSNs become more common, new sensor types and deployment scenarios will continue to erode consensus on ground truth and best practices. We based our system on the one area of consensus we did find: domain experts were able to tell us which data is most likely faulty. Our system builds on these classifications, our deployment experience, and physical mote characteristics to decide when a fault has likely occurred.

Lack of ground truth means we cannot, for example, measure false positives and false negatives in our evaluation. However, human-attended deployments are well-suited to refining fault judgements and achieving ground truth consensus through in-field fault remediation. For instance, our system may indicate that a measurement is faulty and that replacing the sensor board would likely help. If replacing the board does bring the measurement back into realistic bounds, then we gain confidence in our system's judgments. Likewise, consider an out-of-range datum that our system *does not* identify as faulty. A human standing by to sample the corresponding area with more precise, expensive methods, again acting to confirm or deny our hypothesis that our assignment of actions to data points covers most truly unreliable points in the data set. This sort of active, directed investigation requires a system such as ours.

2 Related Work

2.1 Fault detection techniques

Fault detection is a critical step in data analysis, especially when the data are being used for scientific purposes. In [7], Bertrand-Krajewski et al. develop a set of seven criteria derived from physical processes underlying the data and measurement system to determine validity of the data. These criteria are: status of sensor, physical range, locally realistic range, duration since last maintenance, signal gradient, material redundancy, and analytical redundancy. If a data point fails a single criterion, the point is considered unreliable. The authors advocate a significant initial effort to gather data and select parameters for the criteria.

This set of criteria is very useful for eliminating unreliable data offline, but in a rapid deployment a user is available to fix problems and salvage data while the system is online. Independently of the work in [7] we have developed a set of rules that relate directly to potential causes, such as a poor connection or a broken sensor. These rules and their implementation are tailored for online use during a rapid deployment. The data identified by our rules are similar to that identified by the offline reliability criteria, which is an affirmation of our online system.

A significant difference between our system and these criteria in [7] is the assignment of actionable causes to each rule. Our system is designed to run in the field and attempts to direct the user to remediate and validate anomalous data. Developing rules specifically for cause identification and remediation or validation is an important contribution of our work.

There exists a large body of work on expert systems for fault diagnosis. The importance of interaction between human operators and expert systems is motivated in [18]. This research is based on the idea that fault diagnosis and repair are knowledge-intensive and experiential tasks [20]. In our work we apply some of these techniques to sensor networks. We explore rules that are specific to data we collected in Bangladesh.

2.2 Deployment papers

The authors of [19] discussed their experiences instrumenting a redwood tree with a wireless sensor network to collect data about the climate directly around the tree. They gathered data for 44 days from 33 sensor nodes with temperature, humidity, and incident and reflected photo-synthetic radiation sensors. They defined data yield as the percentage of expected points which they actually received and reported a yield of 49%. This is a very good yield for a long-term deployment; in this paper we address rapid deployments which must have higher yields. They found that the accuracy of their sensors' factory calibration was not much improved by a thorough calibration before deployment. They did not, however, verify that the calibration accuracy endured throughout the deployment. Our work extends to chemical sensors, where calibration and maintenance issues are more significant than with physical sensors. Additionally, our contribution builds on this work by discussing ways to increase quantity and quality of data online during the deployment.

A rapid deployment to record acoustic activity at a volcano in central Ecuador is described in [21]. Three nodes with infrasonic microphones were deployed, collecting over 54 hours of signals. Their sensors sampled at 102Hz per node, thus their main challenges were that of time synchronization of the data and transmission bandwidth management. Our challenge is instead handling data from many individual sensors and ensuring the quality of that data with calibration and fault detection.

The Networked Infomechanical Systems (NIMS) [13] project is a rapidly deployable system built of infrastructure-based robotics. The NIMS node provides extremely high resolution spatiotemporal mapping of stream contaminants and properties. However, a long-term continuous presence is both unnecessary and impractical due to the cost of the node and its vulnerability to theft. More importantly, even brief installations are sufficient to elucidate water quality trends at an unprecedented level of detail. One rapid deployment of a mobile NIMS node, investigating urban drainage in Medea Creek [11], has been deployed in an urban stream monthly over the past seven months [9]. The system requires only about 2 hours to install and collects samples for 24 hours. In this deployment, a bucket of a known solution was placed on the bank and the NIMS node periodically moves to this bucket for in-situ calibration [9]. In our deployment, the sensors are not only immobile but buried underground, so we must take a different approach to in-situ calibration. Our experience with a non-mobile rapidly deployable system is a unique look at rapidly deployable sensor networks.

In [1], the authors describe a WSN to monitor soil ecology (including soil temperature and moisture). The moisture sensors were calibrated individually and they had set up a database to apply soil temperature and individual moisture sensor calibration equations to provide calibrated moisture readings.

3 Studying Arsenic in Bangladesh

The factors controlling arsenic mobilization to ground water in Bangladesh are not fully understood. A group of scientists has been studying the Munshiganj district for the past 5 years. Their current working hypothesis is that the influx of dissolved arsenic into the ground water is greatly enhanced where irrigation for rice cultivation provides the primary source of aquifer recharge[1]. Aquifers can be recharged by various sources such as rain-water or water used for irrigation. The hypothesis (simplified for our purposes)

[1] An aquifer is a body of geologic material that can supply useful quantities of ground water to natural springs and water wells. Aquifer recharge is the process by which water seeps down through the soil into an underlying aquifer (http://www.nj.gov/dep/njgs/enviroed/aqfrchrg.htm)

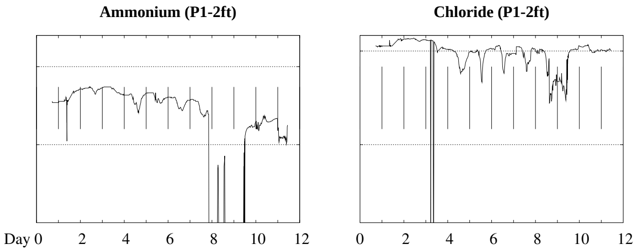

Figure 2: Ammonium and chloride recorded on Pylon 1 with sensors buried 2 feet deep (P1-2ft) over the course of 12 days showing diurnal and diel cycles. Horizontal lines show the upper and lower limits of the sensor's operating range; the extents of the vertical lines show the more limited *linear detect range* (Section 4). There is one vertical line per day.

states that infiltration of irrigation water rich in organic matter can change the oxidation-reduction potential just below the surface of the field. Microbial respiration of the organic matter in the irrigation return flow causes reduction (gain of electrons) and solubilization of arsenic-containing minerals [12, 15, 17]. Our sensor network deployment aimed to collect data that might validate or invalidate this hypothesis.

The scientists wanted to use the dense temporal and spatial sensing provided by a sensor network to better characterize the subsurface environment. Soil is spatially heterogeneous, requiring modeling [14] or dense sampling. Traditional subsurface water sampling entails drilling holes and placing slotted tubing through which to withdraw samples. Samples are brought to the surface using pumps and transported to laboratories or analyzed using a field kit. Thus, there is a substantial effort in taking dense samples and a time lag between sampling and observing the environmental phenomena. Our sensor network deployment would potentially avoid both these problems.

One goal of our deployment was to test for diurnal changes in the groundwater chemistry. Indeed, the temporally dense measurements revealed changes (shown in Figure 2) that were previously unobserved and may be due to photosynthetic activity of either algae or the rice plants themselves.

Several chemicals are integral to the mobilization of arsenic in the soil. Since off-the-shelf arsenic sensors are not available, the scientists hope to further validate work showing that proxy geochemical measurements such as ammonium and calcium can indicate elevated arsenic concentrations. These chemicals are directly measured more easily than arsenic. In addition, they plan to use the data collected from the network to develop a reactive transport model for arsenic mobilization. This model will inform future well placement decisions and deep well construction.

Ion-Selective Electrodes We used 3 types of sensors in our network: Decagon[2] moisture sensors which measure the dielectric constant of the soil, Digi-key[3] thermistors, and Sentek[4] ion-selective electrodes (ISEs). The ISEs and moisture sensors were used to characterize the groundwater chemistry. The output from ISEs is temperature dependent but can be adjusted for temperature using the Nernst equation[5]. In order to use this equation, temperature sensors must be deployed with the ISEs in order to record temperature at each point in time. We also deployed 2 pressure transducers with local logging capabilities to monitor water depth.

ISEs are designed to measure the concentration of a specific ion in a solution. They have an ion-selective membrane treated to allow through the ion of interest. Combination ISEs report the electric potential between ions that pass through this membrane and an internal reference voltage (provided by a gel probe).

ISEs are the most suitable sensor available for in-situ water chemistry measurement with a sensor network. The sensor we chose from Sentek is the cheapest acceptable version of this technology. ISEs are designed for measuring water chemistry and not necessarily for soil environments, but we showed in the laboratory that they are effective for some levels of partially saturated soils and for fully saturated soils. ISEs require frequent maintenance and calibration, and are thus poorly matched with the demands of a longer-term deployment; they are at least theoretically suitable for rapid deployments, however.

Our deployment used 7 types of ISEs:

- **ORP** measures changes in oxidation-reduction potential of the soil as the organic-rich irrigation water flows through;
- **ammonium and calcium**, which have been shown to correlate with arsenic in this region in work done by the group in past years [12];
- **chloride, pH, nitrate, and carbonate** help better characterize the geochemical parameters.

In order to get a stable reading from an ISE in the lab, we had to average over multiple values to minimize the noise in the signal. We define a **noise window** as the number of samples required for a reliable average. We set the noise window to 15 for our deployment, and samples are taken once per second in this noise window. In addition, often the first several measurements in a noise window are unrepresentative, so we throw away the first 5 measurements at the server when calculating an average. All of these parameters were experimentally derived and chosen conservatively for our deployment.

Of all of the sensors we used in our network, the ISEs required the most attention in calibration and fault detection, and also provided the most interesting data. Thus, the focus of this paper is on the data produced by the ISEs.

[2]http://www.decagon.com
[3]http://www.digikey.com
[4]http://www.sentek.co.uk/direct.htm
[5]The Nernst equation is described at http://www.nico2000.net/Book/Guide4.html

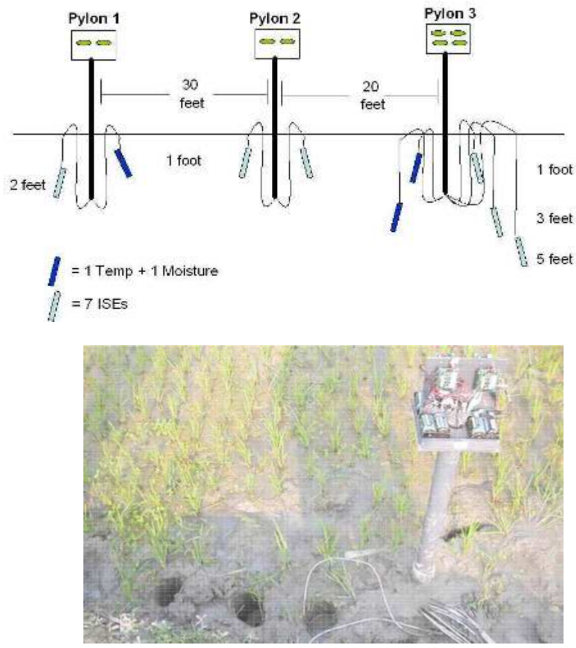

Figure 3: Top panel is the layout of our deployment in the field. Depths below ground are indicated on the diagram. Light rectangle corresponds to a full suite of 7 ISEs, and dark rectangle corresponds to 1 temperature and 1 moisture sensor. Bottom panel is an image of a pylon without a lid deployed in the rice paddy in Bangladesh.

Pylon Design In addition to choosing sensors for our deployment, we had to design an enclosure for the mote system, implement the software, and test the hardware and software that we would deploy. The first challenge was to design an enclosure that would protect the motes from the environment, be easy to deploy, and minimize disturbance of the soil during the deployment process. At each location, we wanted to deploy a full suite of sensors at 3 different depths, in order to characterize the chemistry above, in middle, and below an iron band that the scientists suspected was located at an approximate depth of 3 feet. We designed and deployed the PVC enclosure which houses all the networking hardware needed for three depths and sits on top of a column (Figure 3). One suite of sensors included the 7 ISEs listed above and temperature and moisture sensors. The layout of the sensors and pylons is in Figure 3.

Initially our plan was to deploy all sensors in a single hole beneath the pylon column. However, placing sensors at multiple depths disturbed the soil too much, making it hard to pack down. Thus, we settled on deploying a single depth of sensors in a hole, and placing the holes as close together as possible. We could not fit more than 4 ISEs in one hole, and the moisture sensors were isolated so that their electromagnetic radiation would not interfere with the electric potential measured by the ISEs. Thus, we dug three holes per depth to accommodate a full suite of sensors. When the pylon is deployed, the sensor cables come out from the bottom of the pylon and extend to the satellite holes.

To aid in ease of deployment, we are developing javelin pylons [4] to replace the pylons we used in Bangladesh. These pylons are even easier to deploy as the pylon column itself contains the sensors. The javelin narrows at the bottom so that it can be driven into the ground, minimizing the impact on the soil and avoiding the need to dig holes for the sensors or for the pylon structure itself.

Networking The enclosure of the pylon housed the networking and sensor-related hardware. We used Mica2 motes connected to a MDA300 sensor-board to collect data from the sensors in the pylon. The base-station, a Stargate[6] powered by a car battery, collected data from the network. We used the Extensible Sensing System [5] for our network stack; this included multihop data collection at a centralized sink, time synchronization, a network debugging tool [6], and a disruption tolerant networking layer [3] based on delay tolerant networking [8].

Since improving the quantity of data is especially important for rapid deployments, the disruption tolerant networking layer was critical for our success. While this layer does not provide end-to-end reliability, it can handle longer-term route disruptions that MAC-layer retransmissions cannot. If a valid route to the base station is not present or the MAC layer fails to successfully transmit a packet to its next hop, the disruption tolerant networking layer saves the packet to local storage [3]. Writing data to local flash consumes power, but the additional reliability justified the tradeoff in practice. For example, many nights we were not able to deploy our base station due to security issues (even the car battery was vulnerable to theft) and various software problems. However, we lost minimal data as a result of these issues or any other base station outages, eventually receiving 76% of the expected packets—a relatively high yield in the spectrum of sensor network deployments [19].

4 Calibration and Testing

Before deploying our sensor network in Bangladesh we spent 2 months in the lab calibrating and testing our system. Calibration is the process of mapping a sensor's measured output to an estimate of the property being sensed. The calibration process for the ISEs is the most involved of all the sensors we used in Bangladesh, so we focus our discussion on them.

Mistakes in the process of calibrating the sensor can result in large margins of error when translating sensor readings. Thus, proper pre-deployment calibration is a critical step in enhancing a user's confidence in the subsequent collected data.

The accuracy requirements of the application must be considered during this step. As described in Section 3, the purpose of our deployment was to collect data to learn more about the groundwater chemistry in the shallow soil of the rice paddies. We were interested in diurnal behavior of the ionic content. Thus, we needed a good characterization of the sensor's response to changing ionic concentrations.

[6] All of our networking hardware is manufactured by Crossbow, Inc.

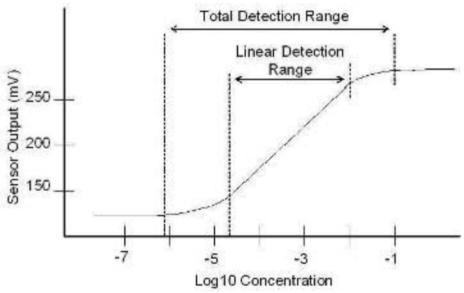

Figure 4: Idealized calibration curve. Calibration parameters are represented by the slope and offset of the equation for the linear portion of the line, and are referenced as such in the paper. Linear and total detection ranges for the sensor - concepts used throughout this paper - are labelled in the diagram.

In this section we describe first the process of calibrating the ISE sensors and the relevance to our deployment and post-deployment work. Then we describe the differences between traditional ISE calibration and calibrating the sensors with the sensor network.

4.1 Ion Selective Electrode Calibration

ISEs are calibrated by exposing the sensor to a range of standard concentrations and developing a function relating millivolt (mV) readings to concentration. Traditionally, the sensor is hooked up to a data acquisition device, such as a pH meter, which displays the millivolt reading from the sensor and a signal as to when that reading stabilizes.

For ISEs, the electric potential relates linearly with the logarithm of the concentration. The R^2 value of the resulting linear regression gives a measure of the confidence to be placed in concentration values translated from voltage readings. The slope and offset of the linear regression line can also be used to identify problems with the sensor. Each type of ion has an expected slope[7]. If the actual slope is lower than expected, the sensor has a sluggish response and translated values will not be as accurate. Additionally, large changes in the slope and offset over the course of the deployment can be indicative of a problem with the sensor. We discuss ways to address this in Section 7.

The calibration curve for a sensor is usually created over the expected operating range of the sensor, or its **total detection range** (TDR). An idealized curve is shown in Figure 4. The TDR is further divided into the **linear detection range** (LDR), and **non-linear detection range** (NLDR) as shown in Figure 4.

Linear detection range The LDR is the linear portion of the calibration curve. The cutoff for this range depends on the accuracy desired in the application. Prior to our deployment, we performed a 12-point calibration curve for each sensor. Data points near the non-linear detection range were sequentially omitted from the graph until the R^2 value for the regression line in the LDR was greater than 0.97. We used a model for the groundwater in Bangladesh based on previous work done at our site [12]. In this way, we were able to identify both the approximate TDR and the sensor's response to expected concentrations in Bangladesh.

Non-linear detection range The NLDR is the portion of the calibration curve where the sensor has a non-linear response to the measured phenomenon. Data that falls in this range may still be useful, but has a lower associated confidence because it indicates a region of lower sensitivity for the sensor. Piecewise linear calibrations can extend the sensor LDR, but both the regressions and the knots between the regressions must be carefully developed for each sensor individually. Below the LDR, the calibration curve decreases in slope where the ISE response cannot be distinguished from water devoid of the targeted ion. An NLDR may also be present at elevated concentrations, where corrective calculations using ion activity coefficients may be necessary.

Based on the calibration results, we characterize each sensor with respect to its TDR, LDR, NLDR, slope, and offset. Correctly calculating these parameters is extremely important, not only to accurately translate millivolt readings into concentrations, but also to identify faulty sensors. Using the TDR as the expected operating range of the sensor, readings from a sensor that occur outside of its TDR can be discarded as faulty. Incorrectly defining the TDR or the slope and offset of the calibration equation could easily result in classifying measurements as being outside of the detection range of the sensor when they are not.

4.2 Calibrating the mote system

Calibrating an ISE sensor for a sensor network involves more than just replacing the pH meter with a mote and repeating the calibration process. The interaction between the hardware of the sensor-network and the sensors necessitate changes and additions to the traditional calibration process.

The goal of this process is to characterize the entire system, not just the sensor, because different hardware can result in different offsets. Even with high quality instruments, issues such as different acquisition hardware and wire-lengths can change the parameters of the calibration curve. In our experiments we encountered a difference of approximately 100 mV, which corresponds to about 2 orders of magnitude in concentration for most of our sensors, when taking readings from a sensor using a pH meter instead of the mote. We also found variations in offset across different sensor boards, though they were much smaller. We attempted to minimize the impact of such issues by calibrating sensors with the exact hardware and wiring they would be deployed with.

Calibrating one sensor with the mote and the software interface we used consumes about 25 minutes. This includes the time to send a query each time the sensor is placed in a standard concentration, retransmit lost queries, sample for a noise window, move sensors between concentrations, and

[7] Monovalent, or single charge ion sensors, are expected to have a slope of approximately 59 mV/decade of concentration, and divalent, or double charge ion sensors, are expected to have a slope of approximately 29 mV/decade of concentration.

wait 60 seconds to account for hysterisis of the sensor in changing concentrations.

While low quality hardware of sensor networks means problems with channel isolation and impedance mismatch can cause additional problems, we found that the best approach is to use the same hardware configuration during calibration as during the deployment. While similar to the previous idea of calibrating with the exact hardware that will be deployed, this idea is different. More specifically, when calibrating a sensor, all other sensors that will be connected to that board during deployment should be connected. For example, in lab we calibrated a chloride sensor first with only the sensor being calibrated connected, and then after connecting the other 6 sensors that the sensor will be deployed with. While the R^2 for both curves is .99, the slope increases by 20% and the intercept increases by 4%. Not all sensors have the same response.

Finally, it is important to test the system exactly as it will be deployed. After we returned from Bangladesh, the sensor manufacturer suggested that some of our faulty data could have resulted from defects in the manufacturing process of the sensors that only showed up when the sensors were buried in saturated soil. To confirm this, we fully submerged some of the sensors in lab and found that there were leaks in some of the connection joints, visible due to air bubbles that slowly leaked out. These leaks could result in changes in the reference voltage or interference in the sensor readings as other ions seep through the cracks.

5 Deployment in Bangladesh

Our field site in Bangladesh consists of a series of rice paddies separated by irrigation troughs built of mud and clay from the field. A motor-driven pump ran for several hours each day to irrigate the fields. Farmers who wanted their fields irrigated break the clay wall separating their field from the bordering irrigation trough, allowing their fields to be flooded.

Next to the irrigation well was a small field where women worked and where our tent was situated. The village was uphill from this field, as were most villages, to minimize damage caused by flooding from the monsoons.

As we worked, people from the village were always present, asking questions and offering all forms of help. They were extremely concerned about the arsenic in the water and curious about the sensor network. Skin lesions on some of the children were a grim reminder that we were working in this region specifically because the arsenic content in the groundwater was extremely high.

Data collected The data collected are shown in Figure 5. We deployed 3 pylons containing 48 sensors in a rice field in Bangladesh over a period of 12 days, collecting approximately 25,000 measurements of data, about 76% of the expected data. We deployed one fully-equipped pylon consisting of 3 suites of sensors at 1, 3, and 5 feet, and two partially equipped pylons with one and two suites of sensors (as shown in Figure 3). The main challenges consisted of performing operations such as testing, calibrating and identifying broken hardware real-time while working in a muddy field environment.

Continuous testing Deploying a pylon involves multiple steps. In order to identify and fix failures at each step we continuously tested the system end-to-end – not just the sensors, and not just the networking hardware.

One pylon includes 27 sensors, and ensuring that all sensors are connected to their correct ports (e.g. if the moisture sensor does not receive an excitation voltage, it will not provide readings) and operational is essential.

We began by calibrating the sensors in a mobile chemistry lab we set up in a tent to identify faulty sensors. We then constructed the pylon and did a 1-point calibration of each sensor in the pylon to identify issues occurring during the construction process.

We needed a way to test the sensors once they had been buried in the ground to ensure that a membrane had not been displaced or a connection had not gone bad during the burying of the sensors in the mud. So, we dipped the sensors in the surface water of the field, took a reading, and compared this to the reading we took immediately after burying the sensor in the ground. Because surface water gathers in the hole immediately before burying the sensor, we assume the reading should be similar.

6 Fault Detection System

Upon returning from Bangladesh, we found that a significant amount of the data collected by the network was uninterpretable by the scientists due to the prevalence of anomalous patterns in the data. A goal of sensor networks is to collect sufficient data to provide a detailed picture of a phenomena. But as a result, there is often too much data for a user to monitor manually in the field. While a tool that can detect and notify a user of faulty data can be helpful, we designed a tool that can also suggest actions a user can take in order to remediate or validate data that appears faulty.

This section describes the rule-based system we designed based on our experience and data from Bangladesh. We developed the system for online use during a rapid deployment. Thus the design relies on the assumption that users will be in or close to the field for short deployments and will be willing to tradeoff their labor and time in order to increase confidence in the data quality.

The system design is based on the end-to-end principle [10, 16]. More specifically, there are many things that can go wrong in the data path, but instead of trying to identify all of these things the system focuses on identifying bad outcomes, represented by anomalous patterns in the data. The system first uses a small set of rules to identify these data patterns. Then, using an associated set of root causes, it suggests actions a user can take to remedy or validate the data.

We separate this process of identifying anomalous data and associating actions into two steps because the first step

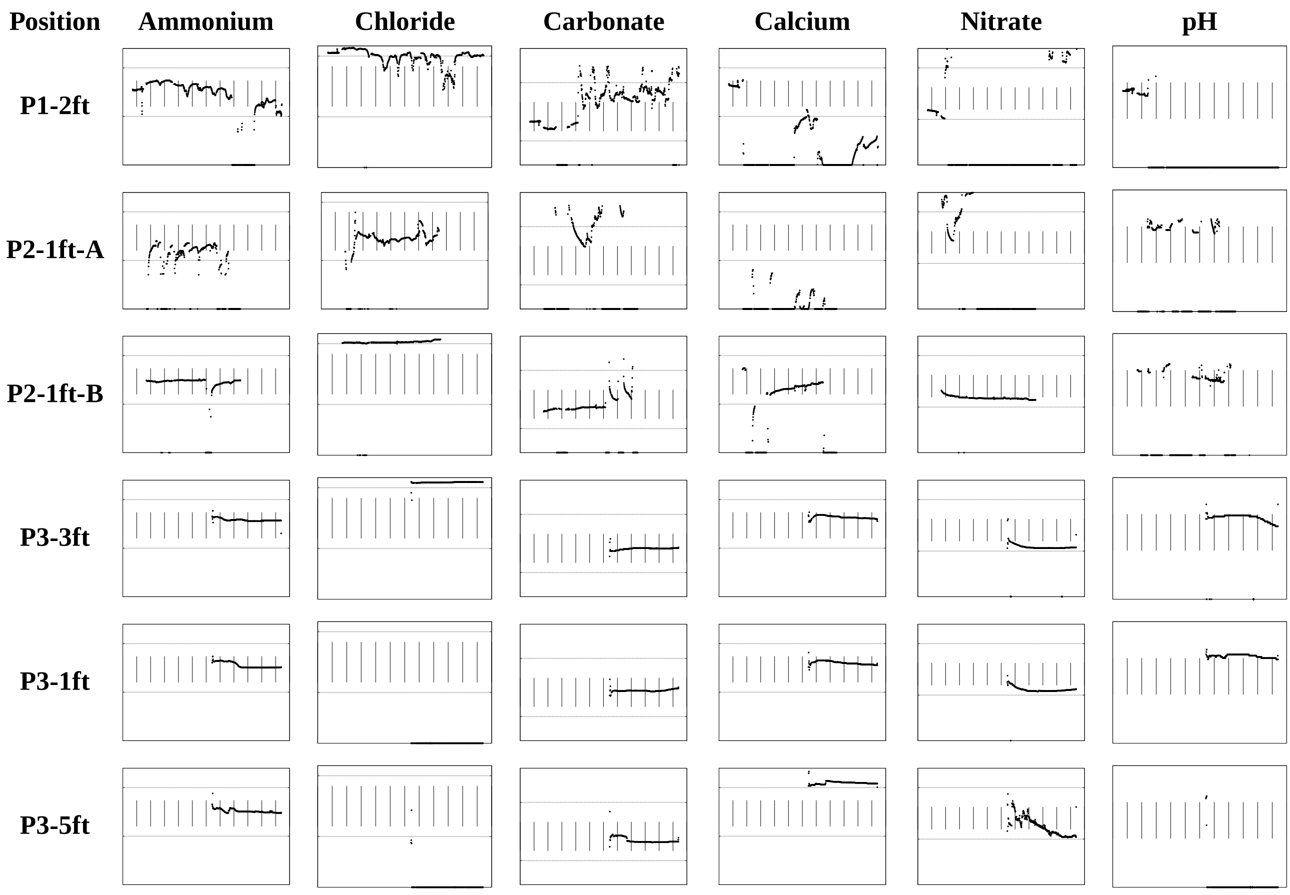

Figure 5: Time-series graphs of concentrations reported by 36 of the 42 ion-selective electrodes we deployed. Each row corresponds to one mote deployed at the specified pylon and depth, each column corresponds to a sensor type. Due to space constraints we did not plot data from any of the redox ion-selective electrodes, or from the temperature, moisture, or pressure sensors. Horizontal lines show the limits of the total detection range; the extent of the vertical lines delimit the linear detection range. There is one vertical line per day. The pH graphs have no horizontal lines as every possible sample point is within their TDR.

can be generalized to other systems with site- and system-specific adjustment to the parameters. The final step of associating actions with these rules, however, is likely to be much more hardware or application specific.

While there are many approaches for detecting faults, we chose this rule-based system as an expedient first step. This initial design framework allows us to do more than just outlier detection. Within the rules, we can aggregate data in particular ways that clearly reveal faults. Additionally, this design framework facilitates later incorporating other specific techniques, such as filtering or machine learning algorithms, as additional rules to the system.

6.1 Rules

Upon receiving a data point, the system applies the set of rules developed to recognize anomalous data. We wanted to avoid a fault-tree approach due to the dependence on ordering of rules often resulting in an artificially imposed hierarchy. Thus all rules are applied to all data points. Each data point receives a vector of bits corresponding to the rules; if a data point meets the criteria specified by a rule, then it receives a 1 in the corresponding bit location of the vector.

There are 5 rules. Several of the rules require setting a parameter - which we discuss below. Some rules are run

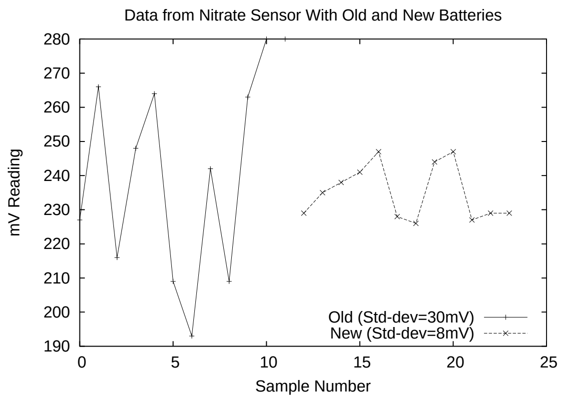

Figure 6: Samples taken from a nitrate sensor with a bad battery and a new battery. An old battery can result in significantly noisier data, demonstrating one instance of a NOISE fault; standard deviation of data within a noise window decreases more than 3X when an old battery is replaced.

directly on the raw millivolt data collected from the sensors, and others are run on the concentration converted from the average of the millivolts in a noise window.

NOISE *This rule applied to 1% of all points.* Because there is always some amount of noise in the measurements, we take the average of readings in the noise window. If

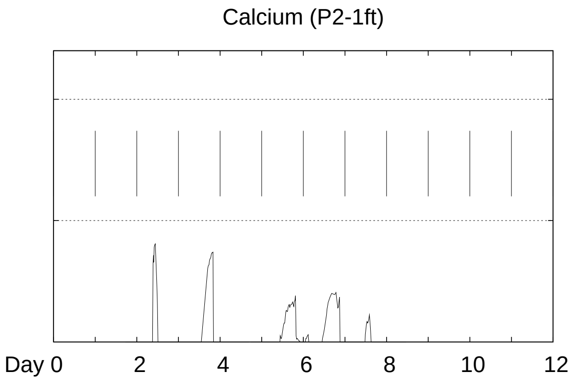

Figure 7: Calcium data from pylon 2 (1 ft)(A). All data are below the linear detection range, indicated by the horizontal line. Graph demonstrating a sensor that falls under the BROKEN sensor rule.

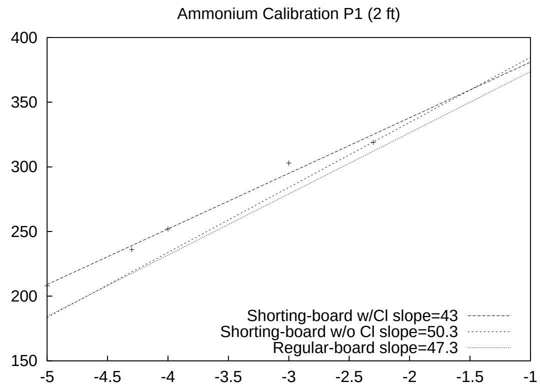

Figure 8: Regression lines for calibration curve taken when ammonium calibrated with all sensors connected, calibrated when the chloride port was reading a short, and calibrated when the chloride port was reading a short and the chloride was connected (for which the points are shown). R^2 for all calibration curves is $>= .97$.

the standard deviation of points within the noise window is greater than some threshold specified by the *noise* parameter, then the points fail this rule.

In lab we found that the standard-deviation of samples within a noise window increases more than 3x when used with a lower voltage battery. Figure 6 is a graph of the samples with the old and new battery. While identifying a limit for an acceptable battery voltage is also another option, it would be inconsistent with our design philosophy of identifying the smaller set of bad outcomes instead of the larger set of bad scenarios that may or may not lead to a negative result.

An instance of a NOISE rule violation in the deployment occurred on the middle of the 5th day of the deployment. The cable for the redox sensor became disconnected from the connector resulting in readings with standard-deviations as high as 60 times the standard-deviation for most readings.

This rule is designed to identify large standard deviations over a small time scale, but very small (or 0) standard deviations over a larger time scale can also indicate fault. A low threshold on the data over long periods of time could also be implemented. However, more careful parameter choice is necessary for differentiating between these stuck-at faults and a truly steady phenomena, and we leave this case for future work.

If measurements from a sensor are identified as noisy, either check the battery or the connectors on the sensor and to the sensor-board.

INVALID NLDR *This rule applied to 11% of all points.* According to our rules, we trust data that occurs in the linear detection range of the calibration curve. The non-linear detection range (NLDR) represents the boundary region of the detection range, thus readings in this range are subject to further examination. Because points in the NLDR are close to the boundary, they could either be potentially faulty, or simply representative of a very low or very high concentration. We hypothesize that if multiple sensors report readings in the same NLDR, then it is more likely a result of a reading in the boundary of the calibration curve and not indicative of a potential fault in the sensor. We define the number of sensors that must validate a reading in the NLDR as the *nldr_redundancy* parameter.

If a sensor reports a reading in the NLDR, and this reading is not validated by *nldr_redundancy* sensors, then this point fails the rule. In order to determine if two sensors agree, the base-station must have received the points from the two sensors within 25 minutes of each other, chosen because the sampling period is 20 minutes, with an additional buffer.

If a point fails this rule, the associated action is to collect manual samples to validate the measured concentration with a more accurate instrument in lab.

BROKEN *This rule applied to 12% of all points.* We calibrated many of the sensors in the field immediately after retrieving them from the ground. When calibrating the calcium sensor from the second pylon, the mote reported 0 millivolts for all concentrations. Checking with the pH meter showed the sensor reading -221 millivolts. While not 0, this reading is still an unexpected reading for a sensor whose normal operating range is approximately between 200-400 millivolts.

Upon examining the data from this calcium sensor (shown in Figure 7), we saw that every one of the 471 points collected from this sensor were below its total detection range. Ideally this sensor would have been identified and changed earlier so that we could have gathered some good data during the 6 days that it was deployed.

A sensor is considered BROKEN if a set of contiguous points are below the total detection range of the sensor. The number of contiguous points required to fail this rule is specified by the *broken* parameter defined below.

If identified as BROKEN, replace the sensor.

SHORT *This rule applied to 26% of all points.* The data from many of the sensors exhibited a common pattern of a sharp change between two successive data points (as specified by the *rate_of_change* parameter) followed or preceded by a string of 0 millivolt readings. Examples of SHORT rule violations can be seen repeatedly in the data from an ammonium sensor in the second pylon in Figure 16.

There could be many causes for this data pattern. We propose two potential causes. First, as seen in Figure 14, a majority of the SHORT violations occur on a subset of the motes. The sensors connected to these motes were manufactured with shorter cables (4 feet long) and were extended with extra wire soldered to a BNC connector. Due to the cable length, the BNC connectors attached to these sensors were exposed to the mud because they were not long enough to reach the housing of the enclosure. This is unlike our second set of sensors which were manufactured with 30 feet of cable. Because the internal reference voltage of the ISE is electrically connected to the BNC connector, we hypothesize that a short circuit was created between the probe of the ISE and its reference voltage.

Second, we experienced a violation of the SHORT rule in lab. A mote started reporting 0 millivolts from a chloride sensor. Independent testing of the sensor with a pH meter showed that the sensor itself was not reporting 0 millivolts. Even after disconnecting the sensor, continued sampling from the port returned 0 millivolts. Under normal circumstances a port without a sensor connected reports noise.

This sensor-board behavior occurred just after we had calibrated the ammonium sensor with the chloride sensor connected. In order to determine if other ports were impacted by the shorting port, we then re-calibrated the ammonium sensor twice more, once with the chloride sensor connected to the shorting port and once without any sensor connected. The three calibration curves are seen in Figure 8. When the ammonium sensor is calibrated with the shorting board, the curve with no chloride connected is closer to the calibration curve on a normal board with chloride connected. This example is important because it demonstrates that in some instances, when multiple sensors are connected to one board, a fault on one channel can impact the readings from other sensors. In cases such as these, remediating faults on one sensor is not only necessary to address issues on that sensor, but also to limit the impact of that fault on other potentially good sensor readings.

Rules like these illustrate the flexibility of the end-to-end approach. In early deployments we want to treat the system mostly as a black-box. In future deployments, the pattern in the data represented by the SHORT rule can be refined to create several rules to match similar patterns caused by different electrical components.

If a short persists, disconnect and check the sensor. Ensure the connector is not in contact with the phenomena being measured. Or, if after disconnecting the sensor the mote continues to report 0 millivolts, change the sensor board.

BAD BATTERY SENSOR Extremely low battery voltage readings can be indicative of a low battery, but they can also indicate a broken battery sensor on the mote. In our deployment we experienced the latter and were forced to replace the mote to resolve the problem. A working battery sensor is critical because battery voltage can be a quick and easy way to diagnose problems in a sensor network [19].

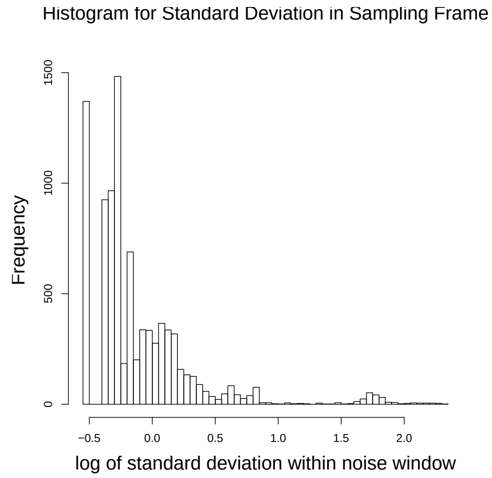

Figure 9: Histogram of the log of the standard deviation in the noise of the measurements. The bimodal distribution allowed for a simple choice of the NOISE parameter.

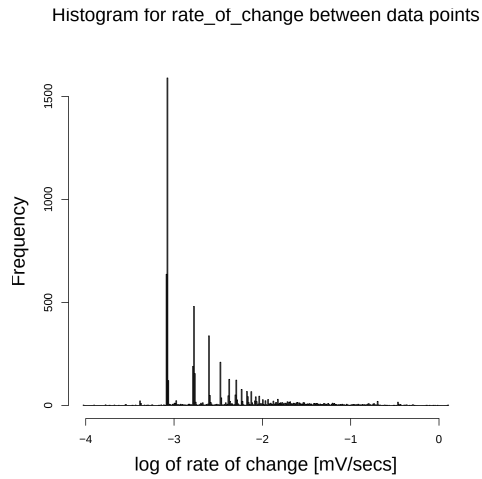

Figure 10: Histogram of the log of the gradient of neighboring data points. The discretization occurs is a result of the resolution of the ADC. Peaks occur at a change of 1, 2, 3, and 4 mV over a period of approximately 20 minutes. Intermediate points occur because samples do not come strictly every 20 minutes due to mote clock imprecision.

If a mote reports a very low battery voltage check or replace the battery or change the mote.

6.2 Assigning an ordered list of actions

Once the rules have been applied to each data-point, the user is notified of the actions associated with each rule that applies to that data point. In the simple case when a data point only fails one rule, the action associated with that rule is reported for that data point.

If a data point fails multiple rules, then a decision must be made of which action to apply. A running tally is kept of the number of contiguous points from an individual sensor that fails a rule. The system then orders the actions based on this count. A rule that has applied to the most number of past, contiguous points will be reported first.

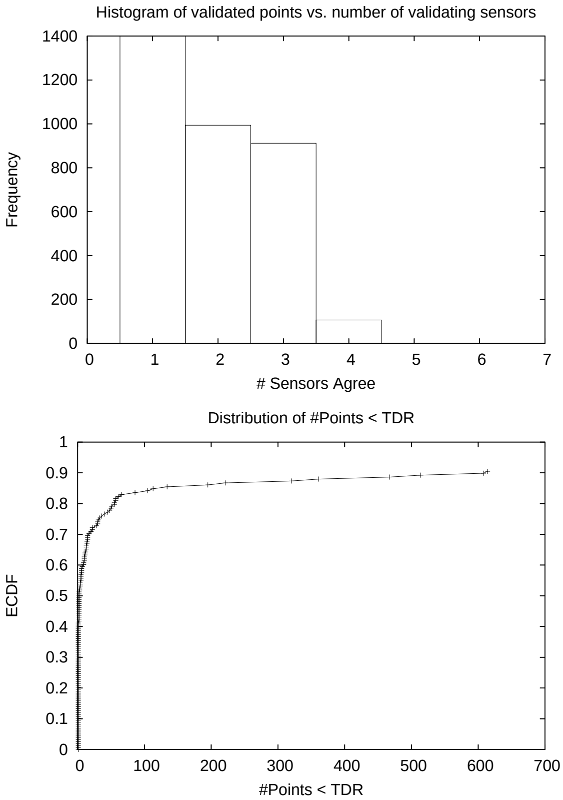

Figure 11: The top panel shows the number of sensors that agree at any one time on reading in the non-linear detection range. The bottom panel shows the empirical cumulative distribution function of the number of contiguous points that any sensor reports outside the total detection range.

6.3 Setting parameters

The parameters required for the fault rules are: *nldr_redundancy*, *rate_of_change*, *broken*, and *noise*. To set these parameters we got an approximate value from the scientists we work with based on the calibration, the dataset, and experiments we have done with the sensors. We examined a variety of approaches before settling on a set of several techniques to set a threshold value. The system also uses the *detection limits* for the sensor as defined during the calibration step.

noise This value, the acceptable standard deviation in millivolt readings within a noise window, is set to 10mV. We chose this value based on the distribution of noise in our dataset. Unlike the other parameters, we know fundamentally that, over short time scales, small noise variance is good and large noise variance is bad. When we plotted a histogram of the variance (Figure 9), there were two clear modes, one clustered less than 10 mV and one clustered greater than 30 mV.

nldr_redundancy The number of sensors that must agree to validate a reading in the non-linear detection range of a sensor. This value is set to 3 sensors (Figure 11).

broken The *broken* parameter represents the threshold for the number of contiguous points a sensor reports outside of the TDR before it is deemed BROKEN. Once a sensor is classified as BROKEN, we do not want it to remain in the system, as we are doing a rapid deployment and want to get useful data from all sensors. However, we do not want to prematurely replace a sensor because replacing a sensor is disruptive and a significant amount of work. Some functional sensors may transiently have readings outside their detection range.

For assigning *broken* we chose a value of 150 contiguous points, or approximately 2 days of sampling, as being the division between most of the points and the smaller set of outliers. Using the ECDF in Figure 11 we can also see that these values identify less than 10% of the points on the graph as being broken.

rate_of_change The threshold for an acceptable rate of change between neighboring points. This value is set to .01 mV/second. As opposed to the histogram for standard deviation, there is no clear place to draw a threshold line, as seen in Figure 10. However, in choosing a value for the *rate_of_change* parameter, the scientists we work with specified that in our field concentrations in soil should not change more than one order of magnitude over the course of an hour without some outside perturbation. Using the histogram, we found the point where the values flatten out near this specified value.

detection limits While not strictly variable parameters, the limits of the linear and total detection ranges for the sensors were also used in the rules. These were set using our calibration curves. We did not stress the upper limits of the detection ranges in the calibration curves, so we used the reported values in the sensor manual for the upper limit of the linear and total detection range.

6.4 Evaluation of Fault Detection

The goal of our system is to associate actions with faulty data; it is not, however, to identify all unreliable data. More specifically, data that does not have an associated action is not necessarily reliable. A complete process may involve running a system like ours in the field to identify and remediate faults in the system, and then run a system like that defined by Bertrand-Krajewski et al. [7] after the deployment to remove unreliable points from the final data set that could not be remediated in the field. However, our system's utility requires that most unreliable data have an associated action. We thus validate our system by testing the hypothesis that it can associate an action with most of the data identified as unreliable by the BK system [7]. Unfortunately, we are not able to analyze the data for false positives and false negatives: deployment conditions made it impractical to confirm our pylons' measurements independently.

We ran both systems on the data set from Bangladesh. Our implementation of the BK system included 3 of its 7 rules, namely physical range, analytical redundancy, and signal gradient. The BK system designers noted that not all rules would apply to every deployment [7]; the missing 4 rules did not apply to us—for example, our sensors did not report status, preventing the use of rule status of sensor. They

	Reliable [7]	Unreliable [7]
Not Faulty	12,138	**581**
Faulty	**82**	8123

Figure 12: Output comparison between our fault detection system (Faulty/Not Faulty) and Bertrand-Krajewski's reliability analysis (Reliable/Unreliable) [7]. Both systems were ran on 20,924 points. Disagreements between the tests are in boldface.

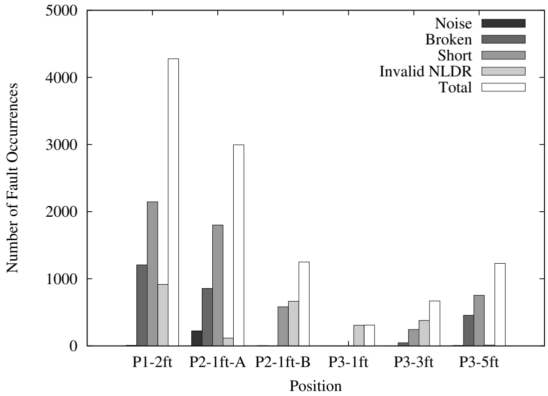

Figure 14: Number of occurrences of faults separated by rule-type and mote.

also state that parameters must be set by gathering data on the system. We set the parameter for their signal gradient test to our *rate_of_change* parameter and the limits of the physical range rule to our total detection range; we define their analytical redundancy rule as our INVALID NLDR rule.

The results of the comparison are summarized in Table 12. We find that our system identifies actionable causes for 93% of the data labeled unreliable by the BK system. About half of the remaining unreliable points fail the BK system's physical range rule; our system cannot explain them because the number of contiguous points is less than *broken*, the parameter for our BROKEN rule. The other half fail their signal gradient rule but are not identified by our SHORT rule because the gradient is not preceded or followed by a series of 0 millivolt readings.

Our system identified 82 points as having an actionable cause despite these points being labeled as reliable by the BK system. This set is also the complete set of points that only fail our NOISE, and 84% of them were caused by a redox sensor that became disconnected from a mote; the BK system was not designed to catch such faults.

6.5 Fault Analysis

We used our system to better understand the faults that occurred in our Bangladesh experiment in order to improve the design for our next deployment. While simply identifying faulty data is helpful, we use the actions suggested by the system and correlations in the faults in order to analyze the deployment.

The rules applied to 39% of the data from the ISEs in our Bangladesh deployment.

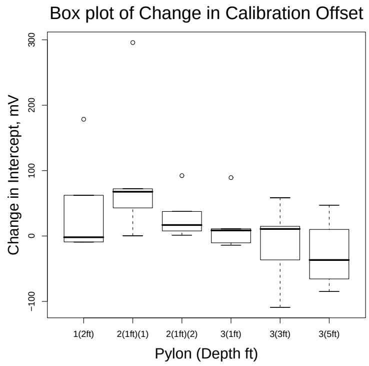

Figure 15: Box plots quantifying the change of slope of calibration curves for each sensor on each mote before and after deployment. The box represents the middle 50% of the data, with the median at the line. The range of this box is known as the inter- quartile range. The whiskers extend to the max and min of the data unless there are outliers, defined as outside 1.5 times the inter-quartile range. Those points are then represented with dots.

A histogram of each rule violation, divided by mote, is shown in Figure 14. We show a 2x6 matrix of *fault graphs*, or graphs of faults identified by the system over time for individual sensors, in Figure 13. The top row of Figure 13 is representative of the ISEs connected to 3 motes that experienced the most faults in the deployment, and the bottom row is representative of the ISEs connected to the remaining 3 motes that experienced the fewest faults.

Using this layout we are able to identify correlations. We describe three of the most useful correlations here.

First, the columns for chloride and nitrate in Figure 13 are a representative cross section of graphs showing the predominance of the INVALID NLDR rule violation for those sensors. By looking at the graphs of the data in Figure 5 we can see that nitrate and chloride data from most sensors tends to be in the lower and upper NLDR, respectively.

The second correlation we found is that the short rule is isolated predominantly on two motes. As mentioned before, these sensors had shorter cables, thus many of the sensor cables connected to these motes were in the mud for the duration of the deployment– a potential explanation for violation of the SHORT.

Third, most often a BROKEN violation is preceded by a SHORT violation. A short that extends for *broken* contiguous points will look like a broken sensor, as in the three top right fault graphs in the matrix. Our system does not distinguish between the two at this point.

7 Discussion and Future Work

In this section we discuss three ideas relating to data integrity in sensor network deployments based on our experience in Bangladesh: calibration that changes with time and exposure to the environment, in-situ calibration and validation techniques, and meta-data storage issues.

Calibration Scheduling Sensors such as ISEs require considerable maintenance and calibration. The calibration

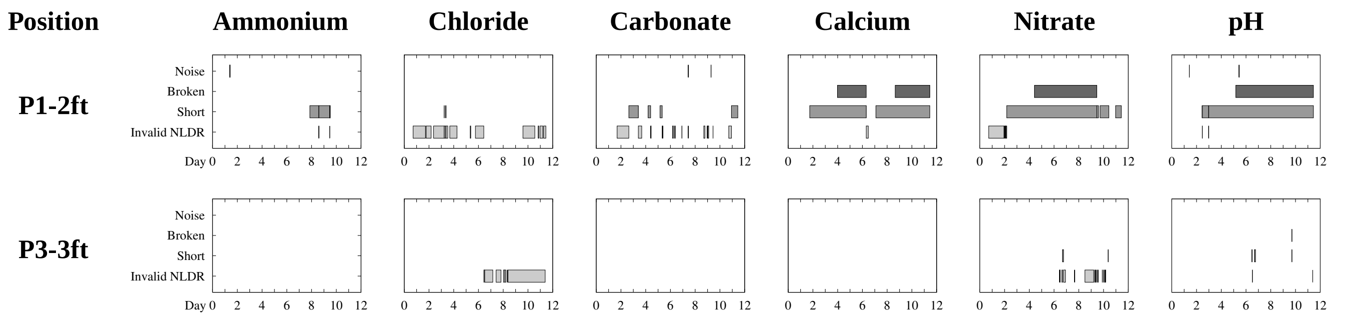

Figure 13: Matrix of fault graphs for individual sensors. Rows in the matrix correspond to motes, columns correspond to sensors. Within each graph, the tics on the x-axis correspond to days in the deployment, and tics on the y-axis correspond to one of the 4 fault rules that applied to the ISE data from Bangladesh. From the top the order is: NOISE, BROKEN, SHORT, INVALID NLDR. Due to space limitations we could not include the graphs for the remaining 30 ISEs.

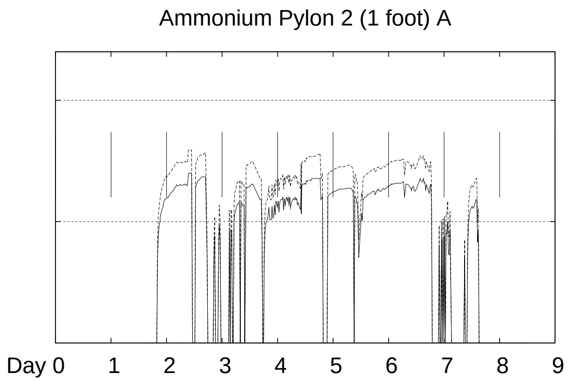

Figure 16: Data from the ammonium sensor in pylon 2 at a depth of 1 foot. Top curve on the graph corresponds to data translated using the pre-calibration equation, and bottom curve corresponds to data translated using the post-calibration equation. Data that may be in-range using one equation, may not be in-range using another equation. Further motivation for calibration scheduling.

equation for these sensors changes over time and with exposure to harsh environment. Figure 15 is a box plot quantifying the changes in our calibration equation parameters between pre- and post-deployment calibrations. Many of the parameters changed significantly over the course of the deployment.

As a result, data that may be in the LDR of a sensor using calibration equations calculated at pre-deployment may no longer be in the LDR of the sensor using the calibration equation from post-deployment and vice-versa. Figure 16 is a graph of ammonium data from Pylon 2; the two lines correspond to data translated using the pre-deployment (top line) and post-deployment (bottom line) calibration equations. Data using the post-deployment equation is more likely to be outside of the LDR than data using the pre-deployment equation. Thus, as the calibration equation parameters change over time, the number of points that apply to a rule may vary, in turn impacting the actions a user should take. For example, using the pre-deployment equation to translate data, a sensor may seem fine because all data is within the TDR, but may appear broken using the post-deployment equation.

We are interested to find if the change in calibration parameters can be predicted even with loose accuracy using time, moisture and temperature of the soil, and the chemical measurements themselves as predictors. If so, we can use this information to predict confidence intervals for the sensor's measurements. The fault detection system can take this into consideration when choosing the appropriate action to remedy fault. A user should remove and calibrate sensors that may have had a large change in calibration parameters.

In-Situ Validation In order to use sensors whose calibration changes cannot be easily predicted, in-situ calibration and validation techniques must be developed. Manual samples can always support environmental measurements on soil and water. Additionally, we have briefly experimented with deploying our sensors underground with a small plastic tube tied near the tip of the sensor. A known concentration can then be injected from the surface down to the sensor to validate that the sensor is responding appropriately.

Meta-Data Storage In order to understand the temporal dynamics of the system and the data, annotation of the data is important. Many important meta-data need to be associated with the data, such as what calibration equation should be used in between calibrations, what faults have been seen, and what actions have been taken. For example, if a user changes some hardware or calibrates a sensor, the data should be annotated with this information. More importantly, when translating the data from millivolts to concentration, given multiple calibration events, the correct calibration equation varies over time. Events in the field should also be associated with the data. For example in our deployment it would be important to know times and durations of rainfall or irrigation events.

8 Conclusion

Rapid deployments are a useful model for environmental monitoring. However, in order to ensure confidence in the relatively small quantity of data, fault detection and remediation is necessary.

Our sensor network was rapidly deployed, and so there was little time to deal with those faults during the deployment. The in-field tools we subsequently designed will allow us to interact with future deployments and greatly improve our confidence in the quality of data collected.

ACKNOWLEDGMENTS

This material is based in part upon work supported by the National Science Foundation under Grant No. BES-0605515.

REFERENCES

[1]

[2] Anonymized.

[3] Anonymized. Master's thesis.

[4] Anonymized.

[5] Anonymized. Technical report.

[6] Anonymized.

[7] J. Bertrand-Krajewski, J. Bardin, M. Mourad, and Y. Beranger. Accounting for sensor calibration, data validation, measurement and sampling uncertainties in monitoring urban drainage systems. *Water Science and Technology*, pp. 95–102, 2003.

[8] K. Fall. A delay-tolerant network architecture for challenged internets. In *Procs. of ACM SIGCOMM*, August 2003.

[9] R. Gilbert. Personal communications. Email messages to authors, 2005-2006.

[10] T. Grandin and C. Johnson. *Animals in Translation*. Scribner, 2005.

[11] T. C. Harmon, R. F. Ambrose, R. M. Gilbert, J. C. Fisher, M. Stealey, and W. J. Kaiser. High resolution river hydraulic and water quality characterization using rapidly deployable networked infomechanical systems (nims rd). Technical Report 60, CENS, February 2006.

[12] C.F. Harvey, C. Swartz, A.B.M. Badruzzman, N. Keon-Blute, W. Yu, M.A. Ali, J. Jay, R. Beckie, V. Niedan, D. Brabander, P. Oates, K. Ashfaque, S. Islam, H. Hemond, and M.F. Ahmed. Arsenic mobility and groundwater extraction in bangladesh. *Science*, (298):1602–1606, 2002.

[13] W. J. Kaiser, G. J. Pottie, M. Srivastava, G. S. Sukhatme, J. Villasenor, and D. Estrin. Networked infomechanical systems (nims) for ambient intelligence. Technical Report 31, CENS, December 5 2003.

[14] A.B. McBratney and M.J. Pringle. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture*, 1:125–152, 1999.

[15] M. Polizzotto, C.F. Harvey, S.R. Sutton, and S. Fendorf. Processes conducive to the release and transport of arsenic into aquifers of bangladesh. *Proceedings of the National Academy of Sciences*, December 2005.

[16] J.H. Saltzer, D.P. Reed, and D.D. Clark. End-to-end arguments in system design. *ACM Transactions in Computer Science*, 2:277–288, November 1984.

[17] C.H. Swartz, N.E. Keon, B. Badruzzman, A. Ali, D. Brabander, J. Jay, S. Islam, H.F. Hemond, and C.F. Harvey. Subsurface geochemistry and arsenic mobility in bangladesh. *Geochemica Acta*, 2004.

[18] M. Thorn. The role of knowledge based systems in fault diagnosis inapplications using human operators. *IEE Colloquium on Fault Diagnosis in Process Systems*, 1997/174:7/1–7/3, 1997.

[19] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, and W. Hong. A macroscope in the redwoods. In *Proc. SenSys*, November 2005.

[20] S.G. Tzafestas. *Fault Diagnosis in Dynamic Systems*, chapter 15, pp. 509–572. Prentice Hall International, 1989.

[21] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh. Monitoring volcanic eruptions with a wireless sensor network. In *Procs. of EWSN*, 2005.

[22] X. Yang, K.G. Ong, W.R. Dreschel, K. Zeng, C.S. Mungle, and C.A. Grimes. Design of a wireless sensor network for long-term, in-situ monitoring of an aqueous environment. *Sensors*, 2:455–472, 2002.

[23] W. Yu. Socio-hydrologic approaches for managing groundwater contamination problems: strategies for the arsenic problem in bangladesh. doctoral thesis, division of engineering and applied sciences. 2003.