Watching the Grid: Utility-Independent Measurements of Electricity Reliability in Accra, Ghana

Noah Klugman[†§]
University of California, Berkeley
nklugman@berkeley.edu

Joshua Adkins[†§]
University of California, Berkeley
adkins@berkeley.edu

Emily Paszkiewicz[§]
University of California, Berkeley
emilypasz@berkeley.edu

Molly G. Hickman[§]
Virginia Tech
mollygh@vt.edu

Matthew Podolsky
University of California, Berkeley
podolsky@berkeley.edu

Jay Taneja
University of Massachusetts, Amherst
jtaneja@umass.edu

Prabal Dutta[§]
University of California, Berkeley
prabal@berkeley.edu

ABSTRACT

In much of the world, electricity grids are not instrumented at the customer level, limiting insights into the power quality experienced by utility customers. Moreover, to understand grid performance, regulators and investors must depend on utilities to self-report reliability data. To address these challenges, we introduce PowerWatch, an agile methodology to directly measure customer experience and aggregated grid performance without relying on the utility for deployment or management. PowerWatch employs a system of distributed sensors coupled with cloud-based analytics. We evaluate the PowerWatch methodology by deploying 462 sensors in homes and businesses in Accra, Ghana for over a year, yielding the largest open-source data set on electricity reliability at the customer-level in the region. We describe the architecture, design, and performance of PowerWatch, as well as the data that are collected, explaining how we determine the accuracy and coverage of our methodology without ground truth. Finally, we report on grid performance issues, finding nearly twice as many outages as the utility observed, suggesting a need for better grid performance monitoring.

CCS CONCEPTS

• **Hardware** → **Sensor applications and deployments**; **Energy metering**.

IPSN '21, May 18–21, 2021, Nashville, TN, USA

ACM ISBN 978-1-4503-8098-0/21/05.
https://doi.org/10.1145/3412382.3458276

ACM Reference Format:
Noah Klugman, Joshua Adkins, Emily Paszkiewicz, Molly G. Hickman, Matthew Podolsky, Jay Taneja, and Prabal Dutta. 2021. Watching the Grid: Utility-Independent Measurements of Electricity Reliability in Accra, Ghana. In *The 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021) (IPSN '21), May 18–21, 2021, Nashville, TN, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3412382.3458276

1 INTRODUCTION

Electric grids power economic activity throughout the world. However, frequent power outages and voltage fluctuations leave many consumers and businesses with only a fraction of the benefits promised by electrification [7, 9, 30, 46]. To remedy this, governments and global development organizations are prioritizing investments to improve electricity reliability in low- and middle-income countries (LMICs) [49, 50].

While utilities in high-income countries have augmented grids with increasingly advanced sensors, utilities in LMICs have limited instrumentation due to budget constraints [77]. Even when LMICs have grid-monitoring equipment, it typically resides only at the transmission tier of the grid, providing limited insight into performance issues like outages and voltage sags at the distribution tier [30]. Without measurement at the distribution tier, utilities struggle to improve reliability, and economies, institutions, and livelihoods suffer.

The value of high-resolution reliability data in LMIC settings is to date unquantified. Without complete instrumentation, it is difficult to know the frequency and extent of outages that occur throughout the grid. The right observations could enable utilities to enhance day-to-day operations (e.g., where to dispatch repair trucks) as well as long-term

† Co-primary authors § Also affiliated with nLine, Inc.
University of California Berkeley IRB protocol CPHS 2017-12-10599

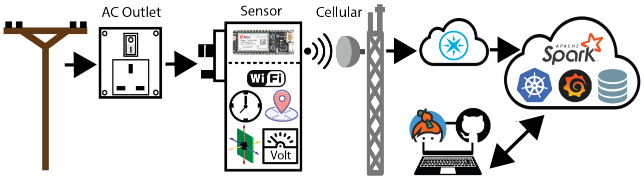

Figure 1: PowerWatch System Architecture. PowerWatch measures the grid by plugging in at outlets in homes or businesses, transmitting data about power quality over the cellular network, and clustering the data based on temporal and spatial characteristics of power outages.

planning (e.g., where to add transformers) [66, 74]. Regulators, who play an important role in enforcing national reliability standards, can use these measurements to hold utilities accountable for system performance [30, 65]. While investors seek these measurements as a key input in deciding to enter a market [2, 20, 50], the best publicly-available measurements are typically only at the country scale [35, 78].

To begin to quantify this observation gap, we present PowerWatch, a system that combines plug-in sensors and an outage-detection algorithm to provide high-resolution, utility-independent measurement of distribution grids. To evaluate our system's performance, we deployed 462 PowerWatch sensors in Accra, Ghana, and collected grid performance data for a period of 14 months. These measurements are supporting the monitoring and evaluation of a $498 million USD investment by the Millennium Challenge Corporation (MCC), a U.S. Government organization chartered to invest in infrastructure to reduce poverty and encourage economic growth. We believe this work represents the largest high-resolution examination of the lived electricity environment for consumers in Sub-Saharan Africa.

Figure 1 shows the PowerWatch system architecture. A PowerWatch sensor plugs into an electrical outlet in a home or business, and periodically reports whether it has power, the local voltage and frequency, the current time and reported location, and other meta-data. We place multiple PowerWatch sensors in a deployment "site" (typically the homes under common grid infrastructure, e.g. a transformer) and detect space-time clusters of power events in the data stream. Such clusters, when present, suggest an outage or restoration in the site covered by the sensors. Similarly, when outages and restorations across sites form a cluster, this suggests more widespread problems at a higher level of grid infrastructure. Our design allows us to deploy PowerWatch without utility involvement or approval–keeping the sampling locations and resulting data streams independent from utility influence–which is critical for PowerWatch's role in monitoring and evaluation functions [51].

Our central claim is that inexpensive plug-in sensors installed in end-users' outlets can detect power outages and restorations at each level of the distribution grid with better granularity and fidelity than existing methods used in the region. We demonstrate that endpoint sensing—in households and businesses—can capture significant grid events that are currently unobserved, assessing the scale and scope of grid performance issues while maintaining utility-independence. Further, by taking measurements more frequently than the typical 15-minute resolution of smart meters, our system can capture important pre-outage and post-restoration grid conditions that may inform additional outage measurements.

With no existing high-resolution electricity reliability data for consumer end-points in Ghana, evaluating PowerWatch directly against ground truth is unfeasible. We evaluate PowerWatch in the absence of ground truth by: (1) examining whether the sensor itself performs well enough in the field to produce a meaningful data stream; (2) examining the deployment's ability to sense outages by finding spatial and temporal patterns across multiple sensors that would be unlikely to occur for reasons other than power outages or restorations; and then (3) examining our coverage model by demonstrating that the information contributed by each site is often redundant, which suggests our deployment is sufficiently dense to detect outages and estimate grid reliability.

In this work, we: (1) introduce a new sensing methodology and data-collection system for taking utility-independent reliability measurements of all levels of an electricity grid (Section 4.4); (2) deploy this system at scale in a developing, urban environment and evaluate our ability to extract true outages from the resulting data streams (Section 6.1); (3) show that a sparse deployment of end-point sensors can detect medium- and high-voltage outages as accurately as the utility and can measure performance at all levels of the grid, providing evidence that the utility may be under-sensing power problems within the low-voltage system (Section 6.2); and (4) present the largest existing high-resolution, open-source dataset on the experience of electricity consumers in a low- or middle-income country, enabling regulators, researchers, and ratepayers to take data-driven steps toward improving reliability (github.com/lab11/powerwatch-ipsn2021).

2 RELATED WORK

Related work encompasses research related both to monitoring electric grid reliability and to deploying sensor systems at scale in uncontrolled environments.

2.1 Monitoring Grid Reliability

Smart meters have existed commercially for decades, but global adoption has been slowing. Utilities in low- and middle-income countries (LMICs) have been particularly slow adopters (estimated at 5% in Africa and the Middle East by 2020 [60]), citing high costs, procurement delays, and difficult integration with existing systems [36]. Instead, many utilities have deployed less-expensive pre-paid meters, which activate only when a customer purchases credit (reducing the need for meter readers). While pre-paid meters do not communicate with the utility [16], large recent pre-paid meter purchases in LMICs still decrease the likelihood of rapid smart-meter adoption in the near future [57].

2.1.1 Sensing the Grid. The distribution grid has many measurement points beyond the high-voltage transmission level, creating challenges for metering at scale. Some prior works address this problem with innovations around new sensor front-ends better able to scale, including sparsely deployed micro-synchrophasors [63, 73], circuit and load-level meters [11, 15, 22], and mobile-phone-based side-channels [45, 62]. Other techniques leverage large, lower-resolution data sets, including satellite nightlight imagery and internet outages, to expose basic measurements for grid stability [32, 54, 58, 64]. While our work does innovate at the sensor front-end–creating a new but relatively simple sensor–we primarily leverage our deployment methodology and back-end analysis for measurement power. We have previously published on our deployment methodology, lessons learned from deploying sensors at scale, and the meta-systems necessary to handle large deployments [44]. In this work, we evaluate the ability of our deployment methodology and analysis techniques to observe the grid.

2.1.2 Resolution of Metering. Micro-synchrophasors sample voltage magnitude and phase angle for each channel and transmit this information at 120 Hz. They require a GPS fix to achieve sufficient time resolution to perform analysis across a deployment [73]. The resulting large data stream, which is sent to the cloud for analysis without sub-sampling, has required innovations in time series databases to ingest and process [8]. Smart meters average and transmit consumption, voltage, and frequency data in relatively infrequent (e.g., 15 minute) reports [76]. PowerWatch performs periodic (2 minute) high-frequency sampling to calculate grid voltage and frequency, and transmits these measurements, along with other data including the number of active WiFi networks and accelerometer readings. Subsequent versions of the PowerWatch sensor more precisely record the time of power state changes using interrupts. By reducing transmission rates compared to micro-synchrophasors, we lower cost and complexity of measurements while retaining the ability to detect power outages and report on voltage quality.

2.1.3 Analytics on Energy Meter Data Streams. A variety of applications require extracting deeper insights from energy meters. Application classes include modeling consumer demand [37], load disaggregation [75], and state estimation [55, 63]. Meta-papers on systems tuned to process the larger data streams from smart meters take different approaches, introducing meter-specific cloud architectures [47], clustering techniques [29], and databases [8]. Our deployment benefits from lower data rates than those requiring specialized time series methods [8] and from the maturity of cloud-based tools like Apache Spark [80] that provide support for our data processing methods.

2.2 Large Sensor Deployments

Evaluations of large-scale sensor networks deployed in the wild discuss challenges related to reliability, networking, node placement, security, and filtering noise introduced by leaving the lab [14, 17, 21, 23, 40, 48, 72]. We evaluate our system against similar considerations, particularly reliability, node placement, and filtering noise from our data stream.

PowerWatch is not the first deployment that lacks ground truth. Techniques to overcome this deficiency include methods that give confidence that a sample is representative [43] and unsupervised learning techniques that extract patterns. Both methods are common in Earth Science, where, like the grid, large-scale phenomena such as forest growth [18] or ocean eddy tracking [28] cannot be directly verified.

Dawson-Haggerty et al. explore the reliability of a long-running deployment of 455 plug-load meters similar in design to PowerWatch [21]. However, their focus is on observing appliances and benchmarking the performance of a low-power wireless mesh network, while ours is in detecting power problems across nodes connected to shared electrical infrastructure and reporting through cellular networks.

Buevich et al. developed and deployed 52 meters on a microgrid in rural Haiti [14]. Their system was installed at the service connection, requiring tight coordination with the electric utility. Much of their work describes the networking difficulties faced in rural environments where a network backhaul is not guaranteed. While PowerWatch explores a utility-independent grid monitoring solution, Beuvich's discussion of a rural deployment methodology would inform future, more rural PowerWatch deployments.

3 ELECTRICITY CONTEXT IN GHANA

Ghana is a West African country with a population of 30.8 million and a per-capita GDP of US$2,266 in 2020 [31]. The grid has roughly 4,740,000 connections and experiences a peak load of 2,881 MW, a supply capacity of 4,695 MW, and an estimated 24.7% distribution loss rate as of 2019 [26]. The Electricity Company of Ghana (ECG) is the distribution utility in Ghana's capital city of Accra [3]. It is important to consider both the electrical and social constraints in Accra to contextualize the PowerWatch system design.

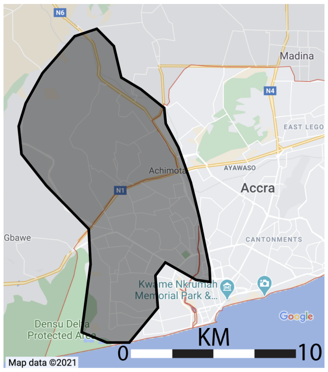

Figure 2: PowerWatch Deployment Area. Sensors are deployed in three of 26 districts in Accra. The deployment covers an area of approximately 130 square kilometers.

3.1 History of Poor Electricity Reliability

Electricity has the potential to provide substantial social and economic benefits [13, 39, 49]. In Ghana, however, the grid at times falls short of providing these promised benefits, resulting in customer frustrations that have culminated in civil unrest [6, 7]. From 2013 until 2015, the country experienced drastic electricity shortages, resulting in outages of six to 24 hours during 159 days of 2015. This period is known as "Dumsor," a Twi word meaning "off-on." While Dumsor has been largely mitigated with the introduction of new generation capacity [19], Ghana still reports longer and more frequent outages than countries with similar GDPs [53].

Partially in response to the Dumsor crisis, the country recently embarked on significant reforms to the entire electric grid, including adding new generation capacity, expanding the transmission network, and re-configuring the distribution network. These efforts have multiple goals, including cutting operational costs, reducing transmission and distribution losses, increasing affordable access to grid connections, and improving reliability. The country's current work to improve grid reliability motivated our selection of Ghana as the deployment venue for PowerWatch [50].

3.2 Ground Truth is Not Available

To improve reliability, it is important to measure it [1, 2]. To understand how well infrastructure investments improve reliability, it is important to have baseline measurements [50].

In Accra, however, high-resolution measurements are limited. The highest spatial- and temporal-resolution measurements come from the ECG SCADA system. This system covers only high-voltage transmission lines and some portion of the medium-voltage distribution network [56].

Measurements of low-voltage outages come primarily from customer calls. Analysis of data collected from the national call center suggests this data stream is sparse and noisy: dips in reporting occur during the day when people are at work and few reports occur during the night when people are asleep. While some of these patterns may reflect an underlying reality (a grid may fail more often when it is operating at capacity, which is more likely in the middle of the day [41]), customer-call data is still likely under-sampling outages, particularly in the context where frequent outages and slow repair response times may reduce a willingness to report.

To improve monitoring, ECG has recently started to deploy smart meters, but economic and social challenges create barriers to achieving broad smart-meter coverage in the short term [5, 10, 52]. ECG recently completed a much larger effort to install pre-paid meters, but these meters do not collect or communicate power quality measurements [25, 61].

3.3 Value of Independent Measurements

An independent audit of reliability is often desirable to regulators and investors. Utilities often have incentives to report strong reliability metrics. Therefore, even if ECG were to deploy a wide-scale roll-out of advanced metering infrastructure, PowerWatch's ability to provide independent data provides value. This independence, along with PowerWatch's low-voltage monitoring capacity, contributed to MCC and their local implementing partner, the Millennium Development Agency (MiDA), choosing to use PowerWatch as a primary source of data for their monitoring and evaluation efforts [51].

4 POWERWATCH SYSTEM

PowerWatch consists of plug-in sensors that take a set of relatively simple measurements—primarily measuring whether the sensor is powered—and an outage-detection algorithm that combines and analyzes the sensors' measurements to form meaningful insights about the performance of the electric grid. In this section, we discuss the performance metrics used to evaluate the grid, the system requirements to collect those metrics, and the system that meets those requirements.

4.1 Key Estimate Informed by PowerWatch

Electric utilities commonly quantify grid reliability by calculating the system average interruption duration index (SAIDI) and the system average interruption frequency index (SAIFI), defined in Equation (1) and Equation (2) [1]. These metrics are calculated for a unit of time (often monthly) with the number of electricity meters served in an area often used for the number of consumers [9].

$$SAIDI = \frac{\text{Total duration of sustained interruptions}}{\text{Total number of consumers impacted}} \tag{1}$$

$$SAIFI = \frac{\text{Total number of sustained interruptions}}{\text{Total number of consumers impacted}} \tag{2}$$

(a) PowerWatch PCB. **(b) PowerWatch sensors.** **(c) Installation.**

Figure 3: PowerWatch as deployed. (a) PowerWatch PCB with cellular radio, SD card, and sensing circuits. (b) Assembled PowerWatch sensors with QR code scanned at installation to associate the sensor with a participant. (c) A field officer installs a PowerWatch sensor at a household outlet.

Accurately calculating SAIDI and SAIFI requires information about the grid's performance (the numerator) and underlying electrical configuration and customer make-up (the denominator). PowerWatch improves the estimate of the numerator; improving the accuracy of the denominator remains future work. In Ghana, the denominator cannot be easily determined due to a lack of accurate grid-infrastructure and customer maps. These information gaps are not uncommon, and many projects are ongoing around the world to map infrastructure and customers to improve the operation of utilities and the reporting of reliability metrics [27, 32, 69].

To quantify grid performance without estimating the number of customers impacted by a sensed outage, we define two new indices: Subsampled SAIDI (S-SAIDI) and Subsampled SAIFI (S-SAIFI) as Equation (3) and Equation (4).

$$\text{S-SAIDI} = \frac{\text{Total duration of sustained interruptions in subsample}}{\text{Total size of subsample}} \tag{3}$$

$$\text{S-SAIFI} = \frac{\text{Total number of sustained interruptions in subsample}}{\text{Total size of subsample}} \tag{4}$$

We note that as the size of the subsample increases and becomes more proportionate to population density, S-SAIDI and S-SAIFI approach SAIDI and SAIFI.

4.2 System Requirements

The high-level goals of PowerWatch are to measure S-SAIDI, S-SAIFI, and other grid-health indicators, while maintaining independence from the utility. These high-level requirements inform the following design decisions.

4.2.1 Deployment. To maintain independence from the utility (see Section 3.3), we cannot rely on the utility to attach sensors to their infrastructure; doing so could introduce sampling bias if the utility makes only some infrastructure available [51]. Instead, we design our sensors to be deployed and debugged by non-experts, allowing us to install sensors at consumer locations we select.

4.2.2 Sensing. PowerWatch sensors must detect the loss of power to calculate S-SAIFI and must additionally detect the restoration of power to calculate S-SAIDI. To accurately capture power restorations, the sensor should be able to keep time while not receiving power from the grid. For all timestamps, the sensor should maintain temporal resolution in seconds. We assume that this is sufficiently fast to observe grid behavior (outages impact the grid on the order of minutes [79]). Sensors should also report their location within tens of meters to allow PowerWatch to estimate the extent of an outage without relying on maps of underlying grid infrastructure. Finally, PowerWatch sensors must detect grid voltage and frequency, features requested by stakeholders.

4.2.3 Communication. Because PowerWatch sensors will not necessarily be collected at the end of the deployment, and the data they collect may be used in real-time in the future, the sensor should have a reliable wide area network connection, with capacity measured in the low megabytes per month. This connection is used to collect data, to track system health parameters, and to perform over-the-air firmware updates. Short network outages are tolerable because data can be stored locally and sent when the network returns.

4.3 Sensing Methodology

To maintain independence from the utility (Section 3.3), we deploy PowerWatch sensors in households and businesses. We deploy sensors with the goal that all sensors in a site are served by the same transformer. This is difficult to ensure due to the high density of the grid in Accra.

While the deployment in customer homes and businesses introduces significant noise—participants unplug sensors, individual prepaid meters run out of credit, and generators artificially restore power—we hypothesize that, with careful filtering, we can extract patterns from our data that give us confidence that a sensor is part of a true outage and is roughly measuring both the spatial extent of the outage and the grid voltage level at which the outage occurred.

Specifically, we look for spatially- and temporally-related changes in power state across two or more sensors to classify an event as an outage, and we deploy three sensors at each site so outages can still be detected if a single sensor fails. By requiring a space-time cluster before classifying an outage, we filter out noise created by placing sensors with end-users. However, by not considering single-sensor reports as true outages, we reject small outages that only affect one sensor.

To determine where in the grid hierarchy an outage occurred, we measure the number of sensors that observe an outage. Given a cluster of only two or three sensors, the point of failure is most likely on low-voltage infrastructure. Given an outage spanning multiple deployment sites, the point of failure is likely in a higher tier of the grid.

4.4 Architecture

The PowerWatch architecture, shown in Figure 1, consists of: (1) an outage-detection sensor that is deployed in utility customer homes and businesses, (2) a cellular network link to the cloud, and (3) cloud-based data analytics that cluster reports from multiple sensors into outages. Field staff help with deployment and management.

4.4.1 Sensor. The PowerWatch sensor, shown in Figure 3, plugs into an outlet and reports the state of the grid over a cellular backhaul through a Particle Electron modem [59]. We selected a cellular backhaul before entering the field given the relatively high percentage of mobile phone users in Accra. Residents have an average of 1.37 mobile subscriptions and 90.0% own a mobile phone [33, 71].

Every two minutes, the sensor takes a high-frequency sample of the voltage waveform at the outlet to calculate grid RMS voltage and frequency. It also records other signals including GPS location and time, cellular quality, and number of nearby WiFi signals (for evaluation as a potential side-channel). Newer sensor versions interrupt on power-state change and record the timestamp (from an RTC) and acceleration (to help filter out user-unplug events). These measurements are stored locally on an SD card and transmitted to the cloud when a cellular connection is available. The sensor contains a 2000 mAh battery, which can run the sensor for several days, longer than most outages in Accra.

The two-minute sampling interval was chosen as a trade-off between data resolution and communication costs. We note that it is higher than the 15-minute sampling rate used by most smart meters, which are considered state-of-the-art in calculating SAIDI and SAIFI. Additionally, for the sensors containing outage timestamping functionality, the sensor can report outages with second-level precision and of less than one minute duration, sufficient for detecting the IEEE's definition of a sustained interruption [1].

The PowerWatch sensor was originally optimized for reliability and ease of manufacturing at small quantities, rather than cost, and was made available to funding agencies for US$187 per unit. The largest contributors to this cost were the populated PCB with power supply, GPS, and sensing circuitry ($87), the Particle Electron ($38), the enclosure ($20), and assembly ($10). Unoptimized communication and cloud infrastructure currently costs $8 per sensor per month. Newer sensor designs, which maintain the same or greater functionality, are projected to cost $30-$40 and optimized communication and cloud costs should be less than $1 per sensor per month.

4.4.2 Deployment. Our deployment methodology is described elsewhere in greater detail [44] and is presented here only briefly to inform our discussion of system performance.

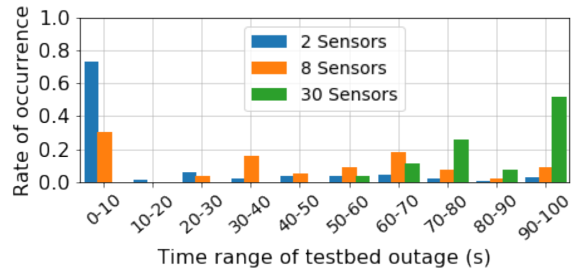

Figure 4: Time range of testbed outages. A testbed of sensors and programmable outlets generated two hundred outages of various sizes in a controlled setting. We observed the precision of outage timestamping, noting that for any given outage sensors may report that the same outage occurred up to 100 s apart. This allows us to parameterize clustering algorithms used to detect outages in the field. Newer firmware reduces temporal variance to less than 10 s.

Accra is segmented into 26 districts; we deployed in three, shown in Figure 2. We chose our deployment sites based on the anticipated locations of new transformers, provided by the utility [52], to allow our data to be used in a formal impact evaluation being done by our collaborators.

PowerWatch sensors are deployed by a staff of 15 temporary local employees (called field officers) and maintained by a staff of four full-time local employees. The field officers screen the occupants of potential sensor locations (called participants) to ensure their home or business is connected to and using the grid. After being informed about the collection, storage, and use of data, participants provide consent to our collection of their personal data and data from PowerWatch. All participant interactions are approved by our IRB protocol. After a participant consents, the field officer uses SurveyCTO [4] to collect information about the participant, the deployment location, and the sensor being deployed.

Field officers plug in the sensor at the participant's home or business or, more recently due to COVID-19, instruct participants to plug the sensor into an available outlet. Participants are provided a power strip to ensure they do not need to sacrifice an outlet, and they are automatically transferred 5 GHC (around US$1) of airtime monthly as an incentive to keep the sensor installed and to offset any electricity costs incurred by participating.

Deployment costs were $100,000 for just over one year of operation, including fieldwork for deployment and maintenance, a full-time project manager, and participant incentives. To reduce participant-incentive and maintenance costs, we plan to pilot an app that shows participants reliability information (e.g., alert them when there is an outage at their home) as an incentive to (continue to) participate and keep their sensors plugged into the grid.

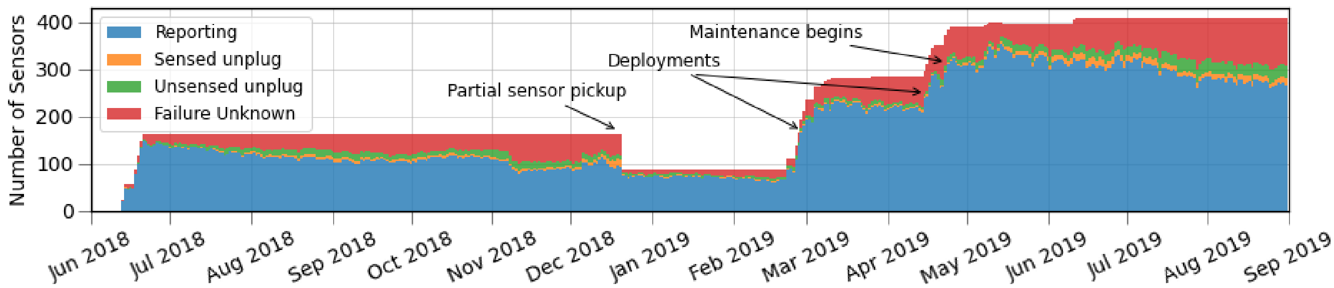

Figure 5: Number of sensors reporting throughout the deployment. Failures are either user unplugs (sensed by the accelerometer), sensors dying due to unsensed unplugs (such as those that occur when the wall switch is flipped), or unknown failures (which are likely also due to participants unplugging or turning off the sensors, as we observe no hardware or long-term software failure in collected sensors). Initial deployments occurred in June 2018, with some sensors retrieved in December 2018, and additional sensors were deployed in February and April 2019. Field staff actively attempted to maintain reliability from April to June 2019, greatly reducing the rate of sensor failure. Even without field staff support, the rate of failure lessens over time, demonstrating that our deployment methodology is sustainable if properly over-provisioned.

4.4.3 Cloud. The core of the PowerWatch cloud receives data from PowerWatch sensors and stores that data in a PostgreSQL/TimescaleDB database [70]. Data is then joined with deployment metadata for further analysis. Data is not deleted from the sensor until the sensor receives confirmation the data was stored in the database.

Additional cloud services supporting PowerWatch include dashboards to monitor the deployment and inform field officers of non-functioning sensors, systems to transfer incentives to participants, and visualizations of outage data.

4.4.4 Outage Clustering. As discussed in Section 4.3, outage reports from multiple sensors are combined to ensure the validity of an outage. We consider two co-reporting sensors sufficient to indicate an outage. To perform this clustering we use STDBSCAN [12], which clusters outage reports into density-based clusters in both time and space.

STDBSCAN requires parameters to specify the minimum number of points within a density-based cluster and the maximum distance between points in both time and space.

To select the time parameter for STDBSCAN, we created a testbed to generate artificial outages of various sizes and observe the time distribution of outage reports in this controlled setting. The testbed consists of three programmable outlets, with 2, 8, and 30 sensors connected to each outlet respectively. Because sensors are connected to the same outlet, we can ensure they experience an artificial outage at the same time. Testbed sensors are programmed with the firmware version that contains the least precise outage timestamping.

The resulting times from this experiment are shown in Figure 4. In 200 artificial outages, all sensors set up to experience an outage successfully reported that an outage occurred. For a given outage, the time range of the outage reports varied up to 100 s (that is, a sensor reported that the outage occurred not more than 100 s after it actually occurred). We therefore conservatively use 100 s as the time parameter for STDBSCAN, allowing the algorithm to cluster two sensors with a reporting-time discrepancy of up to 100 s. We find that data from more recent versions of the sensor report all outages within 4 s of one another, which will allow us to further reduce the clustering time parameter.

To derive the spatial parameter for STDBSCAN, we explored the spatial distribution of sensors within our deployment. We observed that adjacent sites would likely, but not necessarily, experience an outage at the same time [7, 79]. Therefore, for all sites we calculate the maximum distance between any site and its second nearest site. Doing this—and excluding outliers whose second nearest site is beyond 3x the inter-quartile range of the distribution—yields a spatial clustering parameter of 2.4 km.

5 DEPLOYMENT PERFORMANCE

Our evaluation of PowerWatch spans two dimensions: the performance of the system; and the ability of our sensors, deployment methodology, and outage-detection algorithms to measure S-SAIFI and S-SAIDI. In this section, we evaluate the performance of the sensors by examining the sensor uptime, packet reception rate, and spatial and temporal accuracy of the sensors and the deployment. The sensor instrument performed as designed, staying alive and precise over long periods of time in a challenging environment.

Uptime. Uptime across the PowerWatch deployment is shown in Figure 5. We measure average uptime across the deployment to be 73.6 % with suspected unplug failures occurring 2.3 % of the time, suspected sensor switch-offs occurring 5.2 % of the time, and unknown failures occurring the remaining 18.9 % of the time. Further, we find that at least two sensors

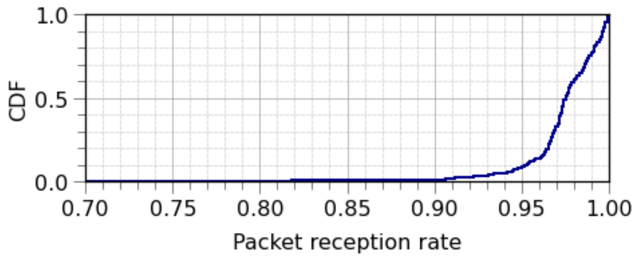

Figure 6: Packet Reception Rate (PRR). PRR is calculated by comparing each sensor's expected reporting interval and sequence numbers with data received. Jumps in sequence number, or periods sensors did not report when expected, indicate a transmission failure due to lack of cellular connection or bugs in the firmware. Sensors are not included after permanent failure, and PRR is increased by local queuing.

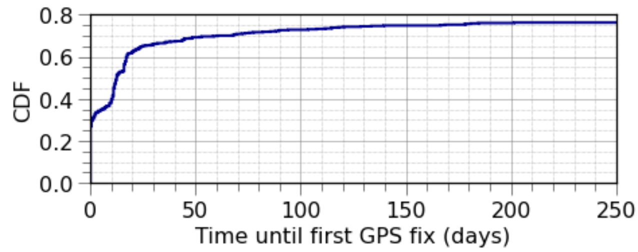

Figure 7: Time to acquire first successful GPS fix. Note CDF axis stops at 0.8. From 462 sensor deployments, over 17% achieve a fix within the first hour after their deployment began, and over 29% within the first day. Over 65% achieve a fix within 30 days. The remaining 11% that achieve a fix are spread over 300 additional days. In 23.2% of the deployments the sensors never achieved a GPS fix.

(the minimum number for our outage detection algorithm to detect an outage localized at that site) are reporting per-site on 85.3 % of site-days.

While we would like to collect more information about the causes of unknown failures, we note that when our field team calls participants and asks them to re-plug-in their sensors, such as in May 2019, the sensor reporting rate increases significantly. This, along with the fact that we find all re-collected sensors to be functional when operated in a controlled setting, leads us to believe that most unknown failures are due to participant behavior.

Packet Reception Rate. To measure the quality of the cellular backhaul we calculate the per-sensor packet reception rate (PRR) based on packet sequence numbers and their expected reporting interval, excluding sensors if they permanently fail. We see a mean PRR of 97.4 %, and that 95 % of sensors have a PRR over 95 %.

GPS Performance Indoors. Because our sensors are deployed indoors, reception is a concern. The rate of GPS fix is low, with 44 % of reports containing a GPS fix sufficient to get GPS time and 42.9 % of reports containing a GPS fix sufficient for localization. Because the sensors are stationary, infrequent fixes are acceptable, especially when paired with a GPS point taken with a tablet at the time of deployment. 78.0 % of sensors get a valid GPS fix at some point during their deployment. The wide variance in the time for each sensor to acquire its first GPS fix is shown in Figure 7. We conclude that while GPS is moderately successful indoors, it should not be depended on to be quick or universally present.

Spatial Resolution. The location of a sensor is collected both at deployment by the field team and during the deployment by the sensor's on-board GPS. As discussed above, 78.0 % of sensors get a valid GPS fix at some point during their deployment, which is used to verify and correct locations collected by the field team. All locations are collected to 10 m accuracy.

Temporal Resolution. In the experimental setup for Figure 4, we show that we detect 100 % of our more than 100 simulated outages of various sizes while using the first generation of our sensor, which had the least precise outage timestamping ability of our sensors (no RTC). To measure the accuracy of our timestamping in the wild, we compare the reported timestamp to the GPS timestamp reported when a GPS fix is acquired, true for 44.0 % of sensor reports. We find that over 99 % of timestamps are within 10 s of GPS time and over 99.9 % are within 60 s of GPS time.

6 APPLICATION PERFORMANCE

Between June 2018 and September 2019, PowerWatch detected 3,123 outages ranging from large outages that stem from high- and medium-voltage faults upstream of the consumer, to small outages stemming from failures in the low-voltage network near the consumer. The full set of outages are shown in Figure 8.

We further examine the performance of PowerWatch by considering both the methods used for extracting outages from a noisy datastream of outage reports, and the suitability of our sample for estimating S-SAIFI and S-SAIDI.

6.1 Extracting Outages

PowerWatch extracts outages from aggregated single-sensor reports by finding space-time clusters. Evaluating this technique would be a comparatively simple task with the presence of ground truth measurements; however, only limited ground truth exists from the utility at the high- and medium-voltage levels, and nearly no usable ground truth exists at the low-voltage levels (see Section 3).

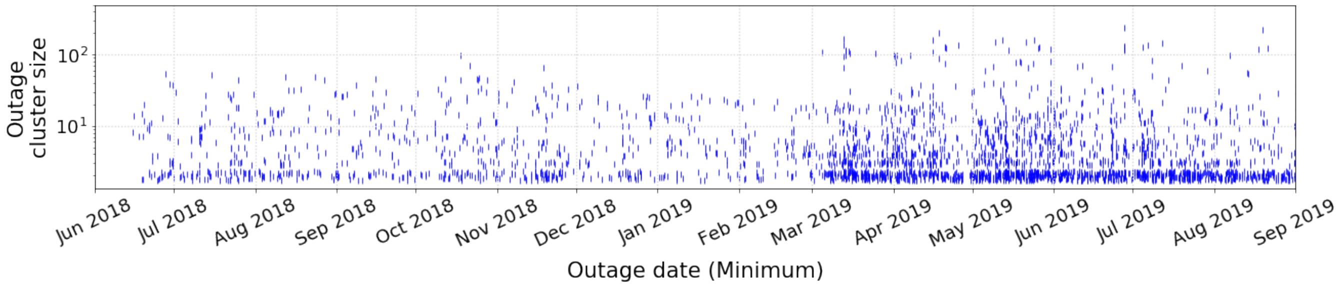

Figure 8: All outages PowerWatch detected from June 2018 to September 2019. The outages are visualized on a timeline where the y axis shows the size of the outage (as the number of sensors impacted) on a log scale. Small perturbations are added to the location of the lines to make it easier to distinguish outages of the same size. PowerWatch detected 3,123 outages with an average duration of 1.7 hours. The longest outage lasted over 48 hours. The largest outage impacted nearly an 80 km^2 area, representing two-thirds of our deployed sensors.

The limited data that do exist—SCADA data for high- and medium-voltage outages, and customer calls and truck logs for medium- and low-voltage outages—are either not made available to our team or do not contain precise-enough space and time information to validate individual outage events.

Therefore, our evaluation augments the limited ground-truth data that we do have with evaluations of the spatial-temporal relationships between sensor reports that we would only expect to see in true outages.

6.1.1 Initial Checks. We first search for anecdotal confirmation that outages similar to those sensed are occurring in Accra. Large outages are sometimes reported in the news and can be detected by PowerWatch (shown in Figure 9)[34].

Additionally, while the utility-reported repair logs are not precise enough to validate individual outages, we can compare their relative number to the outages sensed by PowerWatch. The repair logs we obtained indicate 1,449 repairs in just over five months in one district, while in that same time and district PowerWatch detected 575 outages. While PowerWatch detected fewer outages, this is as expected because PowerWatch is only covering part of the grid. When linearly extrapolating our coverage to the entire district we predict PowerWatch would have detected 1,801 outages in that period, similar to the number recorded in the logs.

6.1.2 Temporal Patterns within Sensor Reports. When there is a power failure, the entire downstream network rapidly de-energizes and should trigger individual PowerWatch sensor outage reports that are very close to one another in time. Conversely, we expect false positive outage reports that are caused by participants unplugging a sensor or by prepaid meters expiring to be relatively randomly distributed in time. The transition between these two temporal groupings should occur around the maximum time cluster of sensors reporting a true outage, about 100 s, as explored in Figure 4.

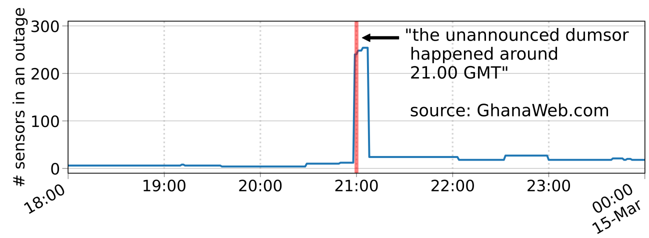

Figure 9: A power outage event ("dumsor") reported by GhanaWeb, a popular news source, to have occurred "around 21:00" on March 14 is perfectly captured by PowerWatch sensors and clustering algorithms [34].

To test this hypothesis we consider the CDF of the time between an outage report and its next closest outage report, shown in Figure 10. In line with expectations, we see a bi-modal distribution of inter-sensor outage reporting times, with a transition between these two modes occuring around 100 s. This supports our hypothesis that true outages are distinguishable in the time domain.

6.1.3 Spatial Patterns within Sensor Reports. We would expect most true outages to be spatially dense, as the spatial distribution of outages (especially small outages) is contiguous. Further, we would not expect to see many powered sensors within the extent of a detected outage, although some are possible due to generators and concave grid structures.

To test whether these properties are true in the PowerWatch dataset, we evaluate the number of powered sensors within the convex hull of the detected outages in Table 1. We find that for all sizes of outages the number of powered sensors within the convex hull is low, with not more than two powered sensors within the convex hull of any outage. The lack of powered sensors within the convex hull indicates that PowerWatch is sensing true outages.

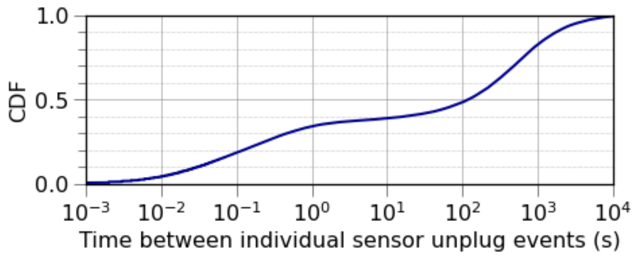

Figure 10: Distribution of times between individual sensor unplug reports. Over 40 % of sensor unplugs occur within 100 seconds (10^2) of another unplug report. Additionally, the flat section in the middle of the graph indicates that sensor unplug reports occur largely in two modes: those highly correlated in time with other unplug events, and those occurring much more randomly in time. We believe the temporal correlation is due to outages, and that the presence of this correlation can be used to separate true unplug events from those not caused by grid failure.

6.1.4 Other Corroborating Signals. To further confirm that the outages extracted by our clustering algorithm are true outages, we examine other signals collected by PowerWatch for signs that an outage occurred. In Figure 11 we analyze the voltage, frequency, and number of WiFi networks detected by sensors near an outage and by sensors not near, and thus not impacted by, the outage.

In sensors near small outage events we see a distinct rise in voltage after an outage and a distinct drop in voltage at the time of restoration. In larger outages we see similar effects impacting the entire network of sensors, and also an increase in frequency throughout the entire network right after an outage occurs. These shifts in voltage and frequency are in line with expectations when a sudden change in electric load, such as an outage, occurs. For both large and small outages we see a drop in the number of WiFi networks at the time of an outage and an increase in the number of WiFi networks at the time of a restoration, consistent with the loss of power to nearby WiFi access points.

6.1.5 Outage Extraction Summary. While we do not have ground truth to tell us the exact accuracy of our outage detection algorithms, the presence of both spatial and temporal relationships between reports collected by individual sensors in the field and the presence of expected voltage and frequency changes near an outage are most easily explained by a failure of the grid. Using the presence of these relationships to bolster the assumption that outage events detected by PowerWatch are true outages, we move forward to evaluate the ability of PowerWatch to sample the grid sufficiently to estimate S-SAIFI and S-SAIDI.

Outage Size	Number of powered sensors within convex hull of an outage			
	Mean	Mean %	Max	Max %
3-10 Sensor Outages	0.03	0.33 %	2	20 %
10-30 Sensor Outages	0.09	0.51 %	2	11.76 %
30+ Sensor Outages	0.31	0.60 %	2	4.65 %

Table 1: Number of powered sensors within the convex hull of an outage. We see that across all sizes of outages very few powered sensors–at most 2–fall within the convex hull of a detected outage. This gives confidence that outages detected by PowerWatch are true outages as we would not expect sensors within an outage area to be powered beyond anomalies such as the presence of a generator or concave grid shapes where separately-powered infrastructure encroach into the convex hull of an outage.

6.2 Sampling Evaluation

An optimal sampling strategy would place sensors such that they capture a representative view of the grid. Unfortunately, this cannot be achieved easily in Accra, as there are few available high-resolution observations of the grid's performance and the available infrastructure maps are incomplete. Therefore, we evaluate our deployment methodology post-hoc, attempting to answer the following two questions: (1) whether we have deployed enough sensors to correctly detect and capture the extent of most high- and medium-voltage outages, and (2) whether we have deployed a sufficient subsample to trust our S-SAIDI and S-SAIFI calculations.

To answer these questions, we first compare our S-SAIFI measurements against the best measurement of SAIFI, gathered by the ECG SCADA system, and then use both numerical and statistical methods to evaluate the predictive power of our subsample for estimating S-SAIDI.

6.2.1 Comparing Against Ground Truth. We compare S-SAIFI against the SAIFI reported by the Electric Company of Ghana (ECG) in Q3 2018, the only SAIFI information we were able to collect at the time of submission. The ECG report includes few low-voltage outages because there is no low-voltage automated monitoring. Some medium-voltage feeders are also not monitored by ECG's SCADA system. ECG reports are aggregated by district, allowing us to directly compare with the one district we had instrumented at the time of this analysis. Finally, ECG's calculation of SAIFI depends on their knowledge of customer service connections in each district, but this data is not available to us.

To compare against the ECG Q3 report, in Figure 12 we compare the district-wide ECG-measured SAIFI to S-SAIFI measured by PowerWatch. The measurement is split into contributions from small clusters of fewer than ten sensors

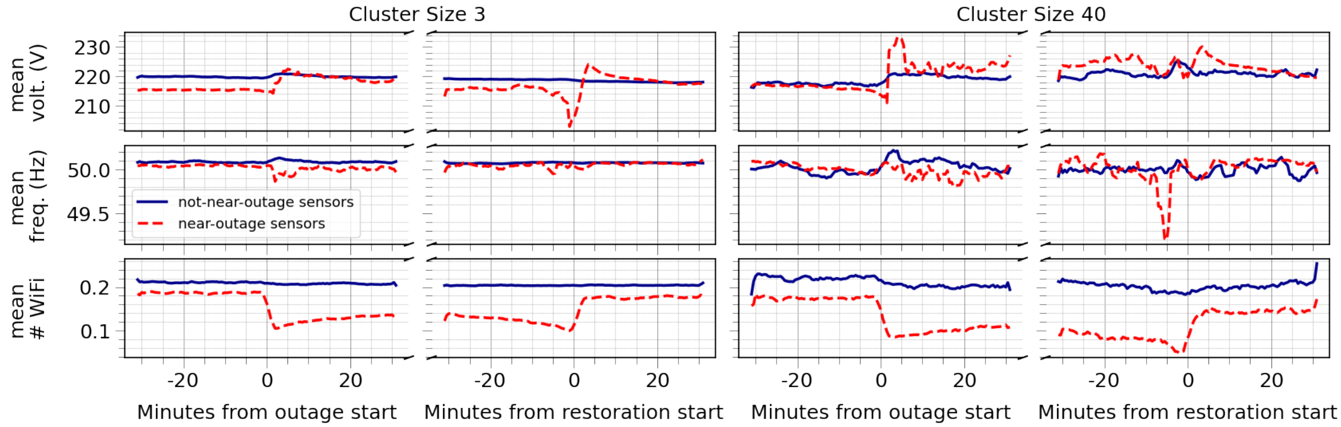

Figure 11: Voltage, frequency, and number of WiFi networks before and after an outage. We time-align and average the voltage, frequency, and number of WiFi networks observed by PowerWatch sensors during small (clusters of 3 sensors) and large (clusters of 40 sensors) power outages and restorations. Sensors are "near" an outage if they are in the same site as a sensor in the outage. Voltage and frequency are not measured for sensors experiencing an outage. As cluster size increases, we observe that sensors not near an outage detect changes in frequency and voltage in response to the change in demand associated with an outage or restoration event. The change in number of nearby WiFi signals is similar–decreasing on outage and increasing on restoration. Together these signals corroborate that outages detected by PowerWatch are true outages.

and clusters of more than ten sensors. We expect the large cluster sizes to correlate with high- and medium-voltage outages included in the ECG report, and the small cluster sizes to correlate with low-voltage outages.

When PowerWatch's S-SAIFI is calculated for larger outages, we see that it closely matches the SAIFI reported by ECG. We also observe that PowerWatch detects a substantial number of smaller outages that are not detected by ECG. This data suggests ECG is under-sampling the grid and under-reporting smaller outages that affect customers.

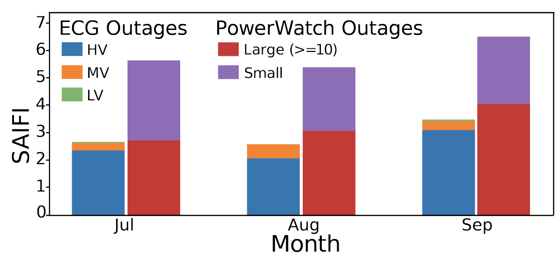

Figure 12: Comparison of PowerWatch S-SAIFI to the utility (ECG) reported SAIFI in quarter 3 of 2018. We see that our large outage clusters closely compare to the combined medium- and high-voltage SAIFI reported by ECG, while low-voltage outages (small outage clusters) sensed by PowerWatch greatly exceed low-voltage SAIFI reported by ECG. This shows evidence of the extent of under-sampling by the utility at the low-voltage level of the grid.

6.2.2 Combinatoric Method for Evaluating Coverage. While agreement with ECG's SAIFI figures increase confidence in our sampling methodology, we further explore the power of our sample with statistical and numerical methods. We begin by verifying that our deployment can cover a significant portion of the high- and medium-voltage grid.

We do not have accurate maps of the infrastructure in our deployment areas, but we do have the relative counts of infrastructure elements at each level of the grid. This provides sufficient information to construct a simple model that estimates the likelihood that our deployment would observe any given high- or medium-voltage failure.

To create this model, we assume: (1) within each district, transformers are evenly distributed between substations, and (2) for every site, each transformer not yet instrumented has an equal chance of being chosen. We then frame the coverage question as an urn problem that yields the likelihood we have chosen at least one site from each substation (high voltage) and/or feeder (medium voltage) after selecting x sites, without replacement, in a given district.

The probability that our site excludes one or more substations is $1 - P(\overline{S_1} \cup \overline{S_2} \cup \cdots \cup \overline{S_n})$, where S_x represents the proposition that the sample includes the x^{th} site and $\overline{S_x}$, that the sample excludes that site. Since the union of n propositions can be expressed as a sum of their intersections [38], we can write:

$$P(\bigcup_{i=1}^{n} A_i) = X_1 - X_2 + X_3 - \ldots + (-1)^{n+1} X_n,$$

where X_k, in our case, is the sum, taken over all combinations of precisely k substations, of the probabilities that all those k

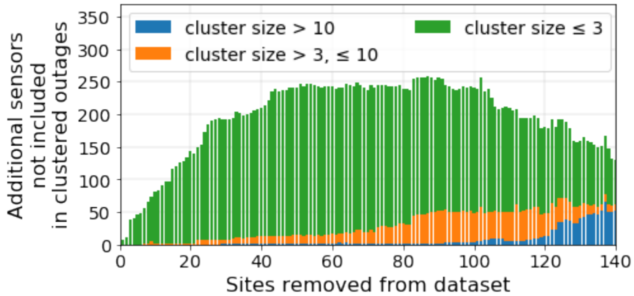

Figure 13: Coverage dropout study from June to August 2019. To evaluate the outage detection coverage of PowerWatch, we perform a dropout study, removing sites from our dataset and observing the impact of those removals. Specifically, we look at the number of "additional sensors" that had been part of an outage cluster prior to the dropout, but which were no longer after a site was dropped. Intuitively, if removing a site causes many outages to either not be formed or shrink significantly in size, that indicates that the site is necessary to detect the correct extent of an outage and that we may be undersampling. During this time period, with no sites removed, there are 1,383 reports from sensors involved in outages of size ≤ 3; 1,030 reports from sensors involved in outages of size > 3 and ≤ 10; and 3,969 reports from sensors involved in outages of size > 10. We observe that for outages containing more than three sensors, nearly 20 sites can be removed from our dataset before we start missing reports from additional sensors. This indicates we have deployed sufficient sensors to detect medium- and high-voltage outages, but, as expected, we do not have a high degree of coverage on the low-voltage network and must rely on sampling to estimate its reliability.

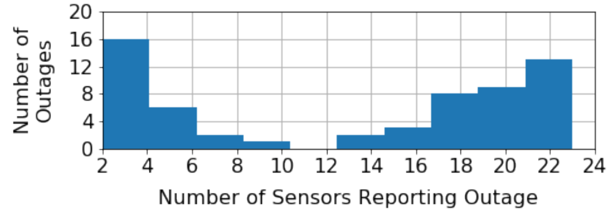

Figure 14: Number of sensors reporting outages in a densely instrumented site. To better understand the limits of our low-voltage sampling we deploy 25 sensors in a single site (under a single transformer) for two months and observe the results. We see two groups of outages - larger outages, which impact all or a significant portion of the site, and smaller outages, which may be a single phase or smaller. The larger outages comprise about 60% of the outages at this site, while smaller outages make up about 40%. This suggests that our primary deployment strategy of three sensors per site detects many, but not all, low-voltage outages.

substations were excluded from the sample. Since we assume that each substation has the same number of sites beneath it and that sites are IID, this collapses to:
$\binom{n}{1}P(\overline{S_1}) - \binom{n}{2}P(\overline{S_1} \cap \overline{S_2}) + \cdots + (-1)^{n+1}\binom{n}{n}P(\overline{S_1} \cap \cdots \cap \overline{S_n})$,
where $P(\overline{S_n}) = \binom{\text{transformers}}{x}^{-1}\binom{\frac{\text{transformers}}{\text{substations}} \times (\text{substations}-n)}{x}$.

We find that of the many ways to choose 85 sites in Achimota ($\binom{461}{85} \simeq 2.4 \times 10^{94}$), Dansoman ($\binom{157}{38} \simeq 4.0 \times 10^{36}$), and Kaneshie ($\binom{343}{28} \simeq 1.0 \times 10^{41}$), in all three districts, fewer than .01% excluded any substations, giving confidence that we should observe all high-voltage outages. Substituting feeders for substations in the equation above, we find that a random draw of 85 sites in Achimota will include all medium-voltage feeders with 44% probability.

6.2.3 Dropout Study for Evaluating Coverage. We verify our coverage model from our data by observing that if there is sufficient sensor coverage of a certain level of the grid, removing a small number of sensors from our dataset should not significantly impact either the number or size of outages detected at that level. We test this hypothesis by performing a dropout study: removing sites from our dataset and observing the impact of the removals on the number and extent of outages detected by PowerWatch.

When removing a site in this study, we expect one of three outcomes: an outage may no longer be detected, the cluster size may become smaller, or a larger outage may become partitioned into two or more smaller ones. In each case, without sufficient coverage, we would expect that removing a single site would cause an outage to shrink by more than just the site dropped for the study, indicating that the removed site played a non-redundant role in our coverage. We increasingly remove sites, count the additionally-impacted sensors, and show results in Figure 13.

We see that for all but the smallest outages, we must remove more than 20 sites before the removed sites are no longer redundant, suggesting that we have sufficient coverage to cluster high- and medium-voltage outages in our deployment areas.

For small outages with three or fewer sensors, we see signs of insufficient coverage immediately—as soon as a single site is removed from our dataset, small outages that were occurring outside of that site are no longer detected. This is unsurprising since, as shown in Figure 14, outages commonly occur that only impact parts of a site. We know our deployment is not dense enough to detect all low-voltage outages; however, we can show it is sufficient for estimating S-SAIDI and S-SAIFI.

6.2.4 Dropout Study for Evaluating S-SAIDI. To determine whether our sample is large enough to capture the range of

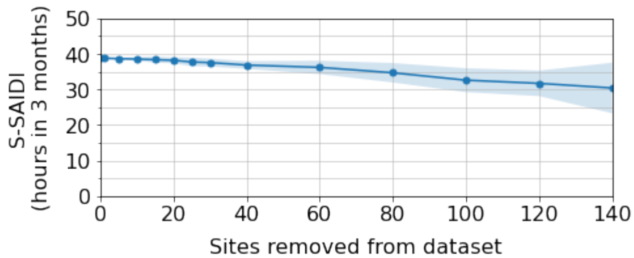

Figure 15: Calculated S-SAIDI ± one standard deviation as sites are removed from the dataset between June and August 2019. To evaluate whether PowerWatch covers a sufficient sample of the grid to compute a representative S-SAIDI, we remove sites from the dataset in 30 rounds and observe the effect on S-SAIDI. We see that as sites are removed, standard deviation of S-SAIDI remains relatively low, and the mean value of S-SAIDI drops slightly.

grid performance in Accra and estimate S-SAIDI, we conduct a similar dropout study to Section 6.2.3. We perform thirty rounds of dropout, randomly selecting sets of sites to remove and observing the effect of each site removal on S-SAIDI.

Had the distribution of outage durations changed significantly as we dropped out sites, we would have seen S-SAIDI vary widely, suggesting that we may be over- or undersampling from parts of the grid with diffiernt interruption experiences. However, as seen in Figure 15, that is not the case. As we drop out sites, we see a downward trend in the mean S-SAIDI, which reflects that the overall distribution of outage durations is asymmetric, with a tail containing a few long low-voltage outages; nevertheless, even as we drop out 60+ sites, we do not see S-SAIDI deviate significantly from our unaltered dataset (39 hours). This indicates that the PowerWatch deployment adequately sampled from the range of reliability present at our deployment sites.

As we remove sites, we also compare the distributions of outage durations between our full dataset and the subsample, using energy distance [67]. We used the implementation of the permutation test for equal distribution [24, 68] available from the *EUGENE* library [42]. We see that the distributions do not significantly differ, meaning the combined distribution does not change as a function of sites removed, further supporting that our sample is adequate.

7 DISCUSSION

Here, we consider some of the limitations of our system as it exists today and more general challenges for wide-area measurements of human-centric systems. We also consider some potential avenues for improving PowerWatch, as well as some new applications that could be enabled with the data made available by PowerWatch.

	Same City	Same Feeder	Same TX
Percent co-reporting	4.5%	11.22%	11.64%
Correlation of voltage first differences	0.04	0.11	0.14

Table 2: Co-reporting rates and voltage correlations scores of sensors under the same infrastructure. We identify sensors under the same infrastructure using maps available for a subset of the grid. We find that sensors under the same infrastructure experience higher rates of outage co-reporting. Similarly, a correlation on the first-differences of the reported voltage reported increases for sensors located under more local infrastructure. This provides evidence that electrical connections are discernible from our data stream, and that applications such as automated topology detection and subsequent root-cause analysis may be possible even without maps of the grid.

7.1 Subsampling the Grid

Our biggest challenge for estimating node placement and density requirements is the absence of ground truth: we cannot claim that we can observe every outage without monitoring every endpoint. One potential remedy could be to partner with utilities that have high smart meter penetration and richer understanding of ground truth, however, these utilities often have very different grid operating contexts. Another approach may be to explore dense deployment methodologies to verify that our expectations match reality.

7.2 Keeping Humans in the System

The PowerWatch methodology requires that participants install sensors on their premises to enable utility-independent measurements. However, this technique introduces noise and incurs cost to recruit and incentivize participants.

While these complications could be avoided by deploying sensors directly on utility infrastructure [14], there are also possible approaches to mitigate the impact of human behavior. For example, we observe that some participants switch off outlets, which appears to our system like a power outage–an event that cannot be filtered out with any existing accelerometer measurement. Adding a method to detect whether the outlet has been switched off, potentially using capacitive loading, time domain reflectometry, or voltage waveform analysis, could reject a switched-off outlet as a false positive. Other approaches may incentivize users to keep sensors powered, perhaps by providing them with data, such as outage notifications and maps, for their own use.

7.3 Beyond Outage Detection

We imagine the PowerWatch data stream could enable numerous applications above and beyond outage detection:

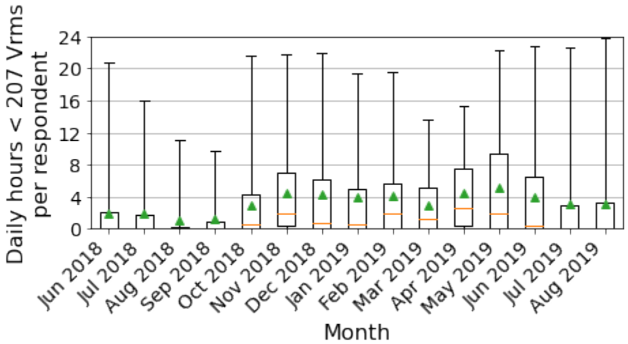

Figure 16: The number of hours respondents experience below the target voltage band (207 Vrms) per day. For each respondent the number of hours of undervoltage per day is calculated. Distributions of time undervoltage are shown per month including mean (green triangle), median (yellow line), and outliers. Measurement of time undervoltage and voltage instability can be used to inform investments to lessen load on existing infrastructure.

7.3.1 Topology Detection. Utilities may not have recorded the exact topology of the grid, particularly at the low-voltage tier [32, 52]. We have started to find electrical correlations between sensors that may provide information about the underlying configuration of the grid (shown in Table 2).

7.3.2 Root Cause Detection. Utilities may not know which pieces of infrastructure are most likely to fail. By associating PowerWatch sensors with maps of the underlying infrastructure, the system could report on the stability of specific equipment. Further, the system could be extended to model and predict the performance of this equipment.

7.3.3 Operational Insights. Utilities may not know how to localize power failures. The PowerWatch system could provide operationally-useful, real-time updates to the utility. PowerWatch may also be able to directly integrate with utility outage management systems, perhaps bootstrapping data transfer by accessing the same interface into utility systems used by the call center when logging customer calls. Further, PowerWatch detects that 18% of voltages sensed are outside the desired range. Undervoltage and overvoltage analysis could help guide equipment upgrades and replacement (shown in Figure 16).

7.3.4 Becoming Ground Truth. Finally, we imagine deploying PowerWatch with higher-resolution sensors, such as alongside micro-PMUs, or with side-channel measurements of the grid, such as satellite nightlights [32, 63] or internet scanning [58, 64]. With PowerWatch as ground truth, the data that these other sensors produce may be more more accurately transformed or interpreted to improve their power.

8 CONCLUSIONS

Our experiences with PowerWatch show that a network of inexpensive, sparsely-deployed sensors placed at imprecisely-selected measurement sites allow us to observe a substantially greater number of low-voltage outages than an operating utility. This provides empirical evidence to explore discrepancies between utility-reported and customer-reported outage rates. PowerWatch matches utility reported rates of high- and medium-voltage outages, validating baseline system performance, but PowerWatch achieves this parity at a fraction of the cost, creating a financially-viable path toward high- and medium-voltage monitoring for the most resource-constrained utilities. The agile and utility-independent measurement methodology also frees regulators and independent evaluators from reliance on the very utilities they are tasked with auditing. While future work will no doubt lead to optimized sensors, algorithms, and deployment methodologies, the data collected using PowerWatch already represents a significant advance in the region and PowerWatch has received additional funding for redeployment at three times the current scale. The data stream that the system generates will be used by both the United States and Ghanaian Governments in their efforts to improve energy reliability in Accra and elsewhere.

ACKNOWLEDGMENTS

This work was supported in part by the Development Impact Lab (USAID Cooperative Agreement AID-OAA-A-13-00002), part of the USAID Higher Education Solutions Network; in part by the Department for International Development Energy and Economic Growth initiative (DFID EEG Project A0534A); and in part by the CONIX Research Center, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

We would like to thank the anonymous reviewers and our shepherd for their guidance. This work would not have been possible without the creativity and support of Catherine Wolfram, Susanna Berkouwer, Steven Puller, and Karen Notsund. We are grateful to Benjamin Jantzen for his guidance on our statistical methods. We also thank our many collaborators including Jeff Garnett, Hana Scheetz Freymiller, and the rest of our contacts at the Millennium Challenge Corporation; Kofi Marfo, Sharon Abena Parku, and the rest of the monitoring and evaluation team at the Millennium Development Authority; Lane Powell, Pat Pannuto, Pierre Biscaye, Kobina Aidoo, and Geetika Pandya; and staff from the Electricity Company of Ghana, SMEC Ghana, nLine Inc., and ESB International. Finally, we would like to extend a special thank you to our field manager Kwame Abrokwah and the field officers in Ghana who have put in many hours working to implement and support our deployment.

REFERENCES

[1] 2001. IEEE Guide for Electric Power Distribution Reliability Indices. *Std 1366* (2001).

[2] 2018. Sustainable Energy for All: Monitoring, Evaluation, and Learning Framework. (2018). https://www.seforall.org/sites/default/files/SEforALL_MEL_0.PDF

[3] 2020. Electricity Company of Ghana. http://www.ecgonline.info

[4] 2020. SurveyCTO. https://www.surveycto.com/

[5] Amevi Acakpovi, Ruhiya Abubakar, Nana Yaw Asabere, and Issah B Majeed. 2019. Barriers and Prospects of Smart Grid Adoption in Ghana. *Procedia Manufacturing* 35 (2019).

[6] Charles Ackah. 2015. Electricity insecurity and its Impact on Micro and Small Businesses in Ghana. In *Institute of Statistical, Social and Economic Research (ISSER) workshop,"Electricity Insecurity and Its Impact on the Economy of Ghana"*, Vol. 14.

[7] Kobina Aidoo and Ryan C Briggs. 2019. Underpowered: Rolling blackouts in Africa disproportionately hurt the poor. *African Studies Review* 62, 3 (2019).

[8] Michael P Andersen, Sam Kumar, Connor Brooks, Alexandra von Meier, and David E Culler. 2015. DISTIL: Design and implementation of a scalable synchrophasor data processing system. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*.

[9] Jean Arlet. 2017. Electricity Tariffs, Power Outages and Firm Performance: A Comparative Analysis. In *Proceedings of the DECRG Kuala Lumpur Seminar Series*, Vol. 23.

[10] W Banuenumah, F Sekyere, and KA Dotche. 2017. Field survey of smart metering implementation using a simple random method: A case study of new juaben municipality in ghana. In *IEEE PES PowerAfrica*.

[11] Nipun Batra, Amarjeet Singh, and Kamin Whitehouse. 2016. Exploring The Value of Energy Disaggregation Through Actionable Feedback. In *International Workshop on Non-Intrusive Load Monitoring*.

[12] Derya Birant and Alp Kut. 2007. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering* 60 (2007).

[13] Brian Bowers. 1988. Social benefits of electricity. *IEE Proceedings A (Physical Science, Measurement and Instrumentation, Management and Education, Reviews)* 135 (1988).

[14] Maxim Buevich, Dan Schnitzer, Tristan Escalada, Arthur Jacquiau-Chamski, and Anthony Rowe. 2014. Fine-grained remote monitoring, control and pre-paid electrical service in rural microgrids. In *Intl. conference on Information processing in sensor networks*. IEEE Press.

[15] Bradford Campbell, Ye-sheng Kuo, and Prabal Dutta. 2018. From Energy Audits to Monitoring Megawatt Loads: A Flexible and Deployable Power Metering System. In *IEEE/ACM Conference on Internet-of-Things Design and Implementation (IoTDI)*.

[16] Ariel A Casarin and Luciana Nicollier. 2008. Prepaid meters in electricity. A cost-benefit analysis. *IAE Business School, Austral University* (2008).

[17] Octav Chipara, Chenyang Lu, Thomas C Bailey, and Gruia-Catalin Roman. 2010. Reliable clinical monitoring using wireless sensor networks: experiences in a step-down hospital unit. In *Proceedings of the ACM conference on embedded networked sensor systems*.

[18] Beh Boon Chun, Mohd Zubir Mat Jafri, and Lim Hwee San. 2011. Mangrove mapping in Penang Island by using artificial neural network technique. In *IEEE Conference on Open Systems*.

[19] Caterina Clerici, Marisa Schwartz Taylor, and Kevin Taylor. 2016. Dumsor: The Electricity Outages Leaving Ghana in the Dark. https://interactive.aljazeera.com/aje/2016/ghana-electricity-outage-dumsor/index.html

[20] IPMVP Committee. 2001. *International Performance Measurement and Verification Protocol: Concepts and options for determining energy and water savings, Volume I.* Technical Report. National Renewable Energy Lab., Golden, CO (US).

[21] Stephen Dawson-Haggerty, Steven Lanzisera, Jay Taneja, Richard Brown, and David Culler. 2012. @ scale: Insights from a large, long-lived appliance energy WSN. In *Conference on Information Processing in Sensor Networks*. ACM.

[22] Samuel DeBruin, Branden Ghena, Ye-Sheng Kuo, and Prabal Dutta. 2015. Powerblade: A low-profile, true-power, plug-through energy meter. In *ACM Conference on Embedded Networked Sensor Systems*.

[23] Prabal Dutta, Jonathan Hui, Jaein Jeong, Sukun Kim, Cory Sharp, Jay Taneja, Gilman Tolle, Kamin Whitehouse, and David Culler. 2006. Trio: enabling sustainable and scalable outdoor wireless sensor network deployments. In *International conference on Information processing in sensor networks*. ACM.

[24] Bradley Efron and Robert J Tibshirani. 1993. Permutation tests. In *An introduction to the bootstrap*. Springer.

[25] Electricity Company of Ghana. 2019. Prepayment Metering. http://www.ecgonline.info/index.php/customer-service/meter-information/prepayment-metering

[26] Energy Commission Ghana. 2019. National Energy Statistics 2000-2019. (2019). http://www.energycom.gov.gh/files/2020%20ENERGY%20STATISTICS-revised.pdf

[27] Joseph H Eto, Kristina H LaCommare, Peter Larsen, Annika Todd, and Emily Fisher. 2012. Distribution-level electricity reliability: Temporal trends using statistical analysis. *Energy policy* 49 (2012).

[28] James H Faghmous, Muhammed Uluyol, Luke Styles, Matthew Le, Varun Mithal, Shyam Boriah, and Vipin Kumar. 2013. Multiple hypothesis object tracking for unsupervised self-learning: An ocean eddy tracking application. In *AAAI Conf. on Artificial Intelligence*. Citeseer.

[29] Christoph Flath, David Nicolay, Tobias Conte, Clemens van Dinther, and Lilia Filipova-Neumann. 2012. Cluster analysis of smart metering data. *Business & Information Systems Engineering* 4 (2012).

[30] Vivien Foster, Samantha Witte, Sudeshna Ghosh Banerjee, and Alejandro Moreno. 2017. *Charting the diffusion of power sector reforms across the developing world.* The World Bank.

[31] Internal Monetary Fund. 2020. World economic outlook database. *World Econ Finance Survey* (2020).

[32] Dimitry Gershenson, Brandon Rohrer, and Anna Lerner. 2019. A new predictive model for more accurate electrical grid mapping.

[33] Ghana Statistical Service. 2014. Ghana Living Standards Survey Round 6 (GLSS6). (2014).

[34] GhanaWeb. 2019. Dumsor hits parts of Ghana. (2019). https://www.ghanaweb.com/GhanaHomePage/NewsArchive/Dumsor-hits-parts-of-Ghana-894685

[35] Christopher D Gore. 2017. *Electricity in Africa: The politics of transformation in Uganda.* Boydell & Brewer.

[36] Christopher Guo, Craig A Bond, and Anu Narayanan. 2015. *The adoption of new smart-grid technologies: incentives, outcomes, and opportunities.* Rand Corporation.

[37] Stephen Haben, Colin Singleton, and Peter Grindrod. 2015. Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE transactions on smart grid* 7 (2015).

[38] John Haigh. 2013. *Probability models.* Springer Science & Business Media.

[39] Mamudu Hamidu and Alfred Sarbah. 2016. An Assessment of the Electrical Energy Needs of Beauty Saloon Industry in Ghana. *American Journal of Management Science and Engineering* 1 (2016).

[40] Andreas Hasler, Igor Talzi, Jan Beutel, Christian Tschudin, and Stephan Gruber. 2008. Wireless sensor networks in permafrost research-concept, requirements, implementation and challenges. In *Proc. 9th Int'l Conf. on Permafrost*, Vol. 1. Citeseer.

[41] Jian He, Lin Cheng, and Yuan-zhang Sun. 2007. Transformer real-time reliability model based on operating conditions. *Journal of Zhejiang University-SCIENCE A* 8 (2007).

[42] Benjamin Jantzen. 2021. The EUGENE Project: A toolbox for detecting, comparing, and characterizing dynamical kinds. https://github.com/jantzen/eugene/.
[43] Norman Johnson and Samuel Kotz. 1978. Urn models and their applications: an approach to modern discrete probability theory. *The Mathematical Gazette* 62 (1978).
[44] Noah Klugman, Joshua Adkins, Susanna Berkouwer, Kwame Abrokwah, Ivan Bobashev, Pat Pannuto, Matthew Podolsky, Aldo Suseno, Revati Thatte, Catherine Wolfram, Jay Taneja, and Prabal Dutta. 2019. Hardware, apps, and surveys at scale: insights from measuring grid reliability in Accra, Ghana. In *Proceedings of the Conference on Computing & Sustainable Societies*. ACM.
[45] Noah Klugman, Javier Rosa, Pat Pannuto, Matthew Podolsky, William Huang, and Prabal Dutta. 2014. Grid watch: Mapping blackouts with smart phones. In *Workshop on Mobile Computing Systems and Applications*. ACM.
[46] Kenneth Lee, Eric Brewer, Carson Christiano, Francis Meyo, Edward Miguel, Matthew Podolsky, Javier Rosa, and Catherine Wolfram. 2016. Electrification for "under grid" households in rural Kenya. *Development Engineering* 1 (2016).
[47] Björn Lohrmann and Odej Kao. 2011. Processing smart meter data streams in the cloud. In *IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*.
[48] Alan Mainwaring, David Culler, Joseph Polastre, Robert Szewczyk, and John Anderson. 2002. Wireless sensor networks for habitat monitoring. In *ACM international workshop on Wireless sensor networks and applications*.
[49] D McCollum, L Gomez Echeverri, K Riahi, and S Parkinson. 2017. Sdg7: Ensure access to affordable, reliable, sustainable and modern energy for all. (2017).
[50] Millennium Challenge Corporation. 2008. Guidance on Common Indicators. 3 (2008). https://www.mcc.gov/resources/doc-pdf/guidance-on-common-indicators
[51] Millennium Challenge Corporation. 2017. Policy for Monitoring and Evaluation. https://www.mcc.gov/resources/doc/policy-for-monitoring-and-evaluation
[52] Millennium Development Authority. 2014. Power Distribution Feasibility Studies, Ghana Phase II Feasibility Assessment. (2014). https://www.mida.gov.gh/pages/view/90/Phase_II_Report_ECG_Final_112414-1.pdf
[53] Millennium Development Authority. 2018. Ghana Monitoring and Evaluation Plan. (2018). https://assets.mcc.gov/content/uploads/ghana-power-compact-me-plan.pdf
[54] Brian Min. 2019. Political Favoritism and the Targeting of Power Outages. (2019).
[55] Ramin Moghaddass and Jianhui Wang. 2017. A hierarchical framework for smart grid anomaly detection using large-scale smart meter data. *IEEE Transactions on Smart Grid* 9 (2017).
[56] Solomon Nunoo and Abel Kwesi Ofei. 2010. Distribution automation (DA) using supervisory control and data acquisition (SCADA) with advanced metering infrastructure (AMI). In *IEEE Conference on Innovative Technologies for an Efficient and Reliable Electricity Supply*.
[57] Gill Owen and Judith Ward. 2006. Smart meters: commercial, policy and regulatory drivers. *Sustainability First* (2006).
[58] Ramakrishna Padmanabhan, Aaron Schulman, Alberto Dainotti, Dave Levin, and Neil Spring. 2019. How to find correlated internet failures. In *International Conference on Passive and Active Network Measurement*. Springer.
[59] Particle.io. 2019. Electron Datasheet. https://docs.particle.io/datasheets/cellular/electron-datasheet/
[60] Pike Research. 2012. Southeast Asia Smart Meter Market Overview: Market Trends, Challenges, Future plans and Opportunities. https://www.engerati.com/sites/default/files/Andy%20Bae.pdf
[61] John Quayson-Dadzie. 2012. *Customer perception and acceptability on the use of prepaid metering system in Accra west region of electricity company of Ghana*. Ph.D. Dissertation.
[62] Ayodeji Olalekan Salau, Adekunle Olugbenga Ejidokun, Olakanmi Adewara, Oludayo Samuel Ajala, Ezekiel Aliyu, and Thomas Kokumo Yesufu. 2017. A GSM-based SMS power notification system for network operation centers. *Int. J. Sci. Eng. Res.(IJSER)* 8 (2017).
[63] Luca Schenato, Grazia Barchi, David Macii, Reza Arghandeh, Kameshwar Poolla, and Alexandra Von Meier. 2014. Bayesian linear state estimation using smart meters and pmus measurements in distribution grids. In *IEEE International Conference on Smart Grid Communications*.
[64] Aaron Schulman and Neil Spring. 2011. Pingin'in the rain. In *ACM conference on Internet measurement*. ACM, 19–28.
[65] Benjamin K Sovacool and Michael H Dworkin. 2014. *Global energy justice*. Cambridge University Press.
[66] NETL Modern Grid Strat. 2008. Adv. metering infrastructure. (2008).
[67] Gábor J Székely and Maria L Rizzo. 2013. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference* 143 (2013). https://www.sciencedirect.com/science/article/pii/S0378375813000633
[68] Gábor J Székely and Maria L et al. Rizzo. 2004. Testing for equal distributions in high dimension. *InterStat* 5 (2004).
[69] Prasad Tallapragada VSN, Maria Shkaratan, Ada Karina Izaguirre, Jaakko Helleranta, Saifur Rahman, and Sten Bergman. 2009. Monitoring performance of electric utilities: indicators and benchmarking in sub-Saharan Africa. (2009).
[70] Timescale. 2019. TimescaleDB documentation. https://docs.timescale.com/latest/main
[71] International Telecommunication Union. 2019. Mobile Cellular 2000 - 2018. https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2019/Mobile_cellular_2000-2018_Jun2019.xls
[72] Deepak Vasisht, Zerina Kapetanovic, Jongho Won, Xinxin Jin, Ranveer Chandra, Sudipta Sinha, Ashish Kapoor, Madhusudhan Sudarshan, and Sean Stratman. 2017. Farmbeats: An iot platform for data-driven agriculture. In *Symp. on Networked Systems Design and Impl.*
[73] Alexandra Von Meier, David Culler, Alex McEachern, and Reza Arghandeh. 2014. Micro-synchrophasors for distribution systems. In *ISGT 2014*. IEEE.
[74] Cheryl A Warren, Dan J Pearson, and Michael T Sheehan. 2003. A nationwide survey of recorded information used for calculating distribution reliability indices. *IEEE Trans. on Power Delivery* 18 (2003).
[75] Markus Weiss, Adrian Helfenstein, Friedemann Mattern, and Thorsten Staake. 2012. Leveraging smart meter data to recognize home appliances. In *IEEE Intl. Conference on Pervasive Computing and Communications*.
[76] Lulu Wen, Kaile Zhou, Shanlin Yang, and Lanlan Li. 2018. Compression of smart meter big data: A survey. *Renewable and Sustainable Energy Reviews* 91 (2018).
[77] World Bank Group. 2019. Is the utility measuring the SAIDI and SAIFI or any other measurements for service reliability? https://rise.worldbank.org/questions/utility-measuring-saidi-and-saifi-or-any-other-measurements-service-reliability-0
[78] World Bank Group. 2020. Ease of Doing Business Rankings. https://www.doingbusiness.org
[79] Rui Yao, Shaowei Huang, Kai Sun, Feng Liu, Xuemin Zhang, and Shengwei Mei. 2015. A multi-timescale quasi-dynamic model for simulation of cascading outages. *IEEE Transactions on Power Systems* 31 (2015).
[80] Matei Zaharia, Reynold S Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, and Michael J Franklin. 2016. Apache spark: a unified engine for big data processing. *Commun. ACM* 59 (2016).