Science of the Total Environment 780 (2021) 146486

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A contribution to drought resilience in East Africa through groundwater pump monitoring informed by in-situ instrumentation, remote sensing and ensemble machine learning

Evan Thomas [a,b,*], Daniel Wilson [b], Styvers Kathuni [b], Anna Libey [a], Pranav Chintalapati [a], Jeremy Coyle [a,b]

[a] Mortenson Center in Global Engineering, University of Colorado Boulder, United States of America
[b] SweetSense Inc., United States of America

HIGHLIGHTS

- Drought in the Horn of Africa threatens access to water for millions.
- Telemetry-connected sensors installed on 480 electrical groundwater pumps.
- Expert and machine learning systems designed to identify the functionality of pumps.
- Pump status sensitivity of 84% for the machine learner, 82% for expert classifier.
- In practice, could result in a 40% reduction in the relative risk of pump downtime.

GRAPHICAL ABSTRACT

ARTICLE INFO

Article history:
Received 18 December 2020
Received in revised form 11 March 2021
Accepted 11 March 2021
Available online 19 March 2021

Editor: Jay Gan

Keywords:
Water pump
Africa
Drought
Machine learning
Remote sensing

ABSTRACT

The prevalence of drought in the Horn of Africa has continued to threaten access to safe and affordable water for millions of people. In order to improve monitoring of water pump functionality, telemetry-connected sensors have been installed on 480 electrical groundwater pumps in arid regions of Kenya and Ethiopia, designed to improve monitoring and support operation and maintenance of these water supplies. In this paper, we describe the development and validation of two classification systems designed to identify the functionality and non-functionality of these electrical pumps, one an expert-informed conditional classifier and the other leveraging machine learning. Given a known relationship between surface water availability and groundwater pump use, the classifiers combine in-situ sensor data with remote sensing indicators for rainfall and surface water. Our validation indicates a overall pump status sensitivity (true positive rate) of 82% for the expert classifier and 84% for the machine learner. When the pump is being used, both classifiers have a 100% true positive rate performance. When a pump is not being used, the specificity (true negative rate) is about 50% for the expert classifier and over 65% for the machine learner. If these detection capabilities were integrated into a repair service, the typical uptime of pumps during drought periods in this region could potentially, if budget resources and institutional incentives for pump repairs were provided, result in a drought-period uptime improvement from 60% to nearly of 85% - a 40% reduction in the relative risk of pump downtime.

* Corresponding author.
E-mail addresses: evan.thomas@colorado.edu (E. Thomas), daniel.wilson@sweetsensors.com (D. Wilson), styvers.kathuni@sweetsensors.com (S. Kathuni), anna.libey@colorado.edu (A. Libey), pranav.chintalapati@colorado.edu (P. Chintalapati), jeremyrcoyle@gmail.com (J. Coyle).

https://doi.org/10.1016/j.scitotenv.2021.146486

1. Introduction

The prevalence of drought in the Horn of Africa has continued to threaten access to safe and affordable water for millions of people. Climate change has exacerbated water insecurity, increasing both the frequency and severity of droughts in this region, resulting in millions of people experiencing food insecurity (Funk et al., 2015a; Nicholson, 2014; Shabelle, 2011; Uhe et al., 2017). The economic and public health effects of droughts are further intensified by limited community capacity and institutional failures that affect the management of water infrastructure, leading to crop failures, displacement of people and disruption of migratory patterns (MacAllister et al., 2020; Knippenberg and Hoddinott, n.d.; Lolemtum et al., 2017; Nthambi and Ijioma, 2020; Bahru et al., 2019; Liou and Mulualem, 2019).

Historical drought support in the Horn of Africa has involved reactive response through emergency aid from international donors and multilateral organizations. However, proactive and preventive measures are estimated to save hundreds of millions of dollars when compared to emergency relief efforts (Venton, 2018; Godfrey and Hailemichael, 2017). Ensuring the availability of groundwater in strategic locations ahead of drought conditions can help prevent the need for emergency response while reducing some of the household and economic water stress (MacAllister et al., 2020; Thomas et al., 2020a; Macdonald et al., 2019; Tucker et al., 2014; Calow et al., 2010). Localized monitoring is imperative to ensuring prioritized and expedited actions are taken ahead of droughts in order to improve water security.

In a recent paper in this journal, we examined patterns in use of groundwater boreholes in northern Kenya and Afar Ethiopia, and compared these patterns to rainfall trends in the region. We observed in statistical models containing rainfall as a binary variable, a overall 23% increase in borehole runtime following weeks with no rainfall compared to weeks preceded by some rainfall. Further, a 1 mm increase in rainfall was associated with a 1% decrease in borehole use the following week. These trends reflect behavioral choices to use surface water sources when available, and do not, generally, reflect an intrinsic hydrologic relationship between rainfall and aquifer recharge (Thomas et al., 2019). Another recent study in this journal which investigated the influence of rainfall on handpump use in Kenya had similar observations, finding that wet season handpump use was 1/3rd lower than dry season use (Thomson et al., 2019).

Across sub-Saharan Africa, groundwater resources and recharge are considerable, and a significant source of future water security for the continent (MacDonald et al., 2021; Banks et al., 2020). In Afar, Ethiopia, the volcanic aquifers are recharged through annual runoff and therefore susceptible to drought conditions, though many are located 60-160 m below the surface and may not run dry even during periods of drought (Borgomeo et al., 2018; Kebede et al., 2008). In Kenya, several examinations suggest the groundwater is fossil in nature, located 100 m or more beneath the surface, and not as susceptible to drought as shallow aquifers. While the use of fossil aquifers pose water quality and sustainability challenges, estimates have suggested that there is enough water to supply all of Kenya for 70 years (Gramling, 2013). These pumps are powered either through the electrical grid, diesel generators, or solar power. The average yield of these pumps in both Ethiopia and Kenya are about 18–19 m3/h, with a standard deviation of 20–21 m3/h (Thomas et al., 2019).

The Drought Resilience Impact Platform (DRIP) integrates early detection and planning with proactive groundwater management to ensure water availability. Our previous work in Rwanda has shown that localized monitoring from sensor-equipped hand pumps enabled local service providers to increase functionality rates from 56% to 91% (Nagel et al., 2015). Furthermore, we determined that machine learning algorithms could be used to predict failures and achieve 99% uptime (Wilson et al., 2017). We have also examined the institutional frameworks supporting groundwater management as a means toward drought resilience (Turman-Bryant et al., 2019), and determined that for real-time data to effectively improve water delivery, this information must be integrated into local water management policies and practices. In a recent publication, we describe the motivation, stakeholders, theory of change, web-based decision support tool, and operations of DRIP (Thomas et al., 2020b). This work builds on complementary efforts in this sector to use mobile technologies to improve monitoring of rural water points (Thomson and Koehler, 2016; Thomson et al., 2012; Thomson, 2020; Koehler et al., 2015; Koehler, 2018).

In this paper, we build on the observed relationship between rainfall and borehole use, and describe the design and validation of several classification algorithms designed to identify a functional water pump (capable of delivering water on demand) from a non-functional pump (broken, distinct from a pump not being used deliberately). As the hydrologic context and water use practices in this region are complex, we demonstrate an approach to leverage statistical models instead of mechanistic relationships to achieve this capability.

The expert classifier relies on explainable and consistent logic statements establishing the relationship between measured parameters and the classification provided, while the machine learner derives a complex statistical model that, while potentially more accurate, has the disadvantage of not being immediately comprehensible to a typical user, a known limitation in machine learning (Roscher et al., 2020). We further examine the potential to use these classifiers to support condition-based maintenance. Condition-based maintenance has several advantages over time based maintenance, especially the ability to allocate limited maintenance resources where they are needed, instead of spreading maintenance resources evenly, including where they may not be needed (Ahmad and Kamaruddin, 2012).

2. Methods

Our operating context are electrical groundwater pumps in arid regions of Kenya and Ethiopia, and a motivation to ensure a higher rate of water pump functionality than the status-quo, recently observed during the 2016–2017 drought as 55% in Kenya (UNICEF, 2017) and 60% in Ethiopia (MacAllister et al., 2020).

At 480 of these pumps, we have installed electrical current sensors. These sensors, which measure the runtime of the pumps by logging the electrical current (amperage) delivered to the pump over time, are Pressac brand wireless current transformers (CTs). The CTs transmit their data to on-site cellular or satellite gateways, and then telemetry is relayed to a central database and web application. Fig. 1 illustrates a typical groundwater pump in northern Kenya (center) with an electrical current transformer being installed on the control panel (left) and the satellite or cellular telemetry gateway (right).

As of February 2021, 480 groundwater pumps are being monitored, with over 830 pump-years of data combined, since the first installation in January 2016. On average, each pump has 1.8 years of sensor collected hourly runtime data. In total, the water supplies of approximately 3 million people are being monitored on a daily basis.

The predictive algorithms described in this paper were developed by leveraging a combination of this in-situ and remotely sensed data, validated with manually-collected ground-truth. The flowchart in Fig. 2 describes the operating conditions of the groundwater pump system. Each pump has two primary real-world conditions - it is either “Functional” and capable of delivering water, or it is “Broken” and cannot deliver water without a repair or management intervention.

As a first-order approximation to distinguish functional versus broken conditions, our electrical runtime sensors indicate if the pump is switched on or off. However, by itself this approximation is insufficient to reflect the operating environment for these pumps. Namely, we have observed that many pumps may not be used, even if they are functional, during the rainy seasons in these regions where people are otherwise able to secure surface water for themselves and their livestock (Thomas et al., 2019).

Fig. 1. A typical groundwater borehole in Turkana, Kenya (center). Electrical monitoring sensors are installed at the pump controller (left), and wireless transmit data to a self-powered satellite or cellular transmitter (www.sweetsensors.com).

Therefore, a more sophisticated classification system is required to better distinguish between "Broken," a true-negative condition, and "Not-Running on Purpose," a true-positive condition. Indeed, considering this nuance in classifying functionality of groundwater pumps is consistent with recent literature that identifying the importance of including seasonality (Carter and Ross, 2016) and water point reliability (Bonsor et al., 2018) beyond the simple measure of active use when classifying water pump functionality. In the following sections, we describe the data collection and interpretation methods we applied to improve on this distinction between use and functionality in our context.

Our work is motivated toward supporting data-enabled responsive maintenance, and better still, preventive maintenance. Although it is useful to know when a pump is broken, it equally valuable to know that a pump is functional, with an aim toward effective resource allocation.

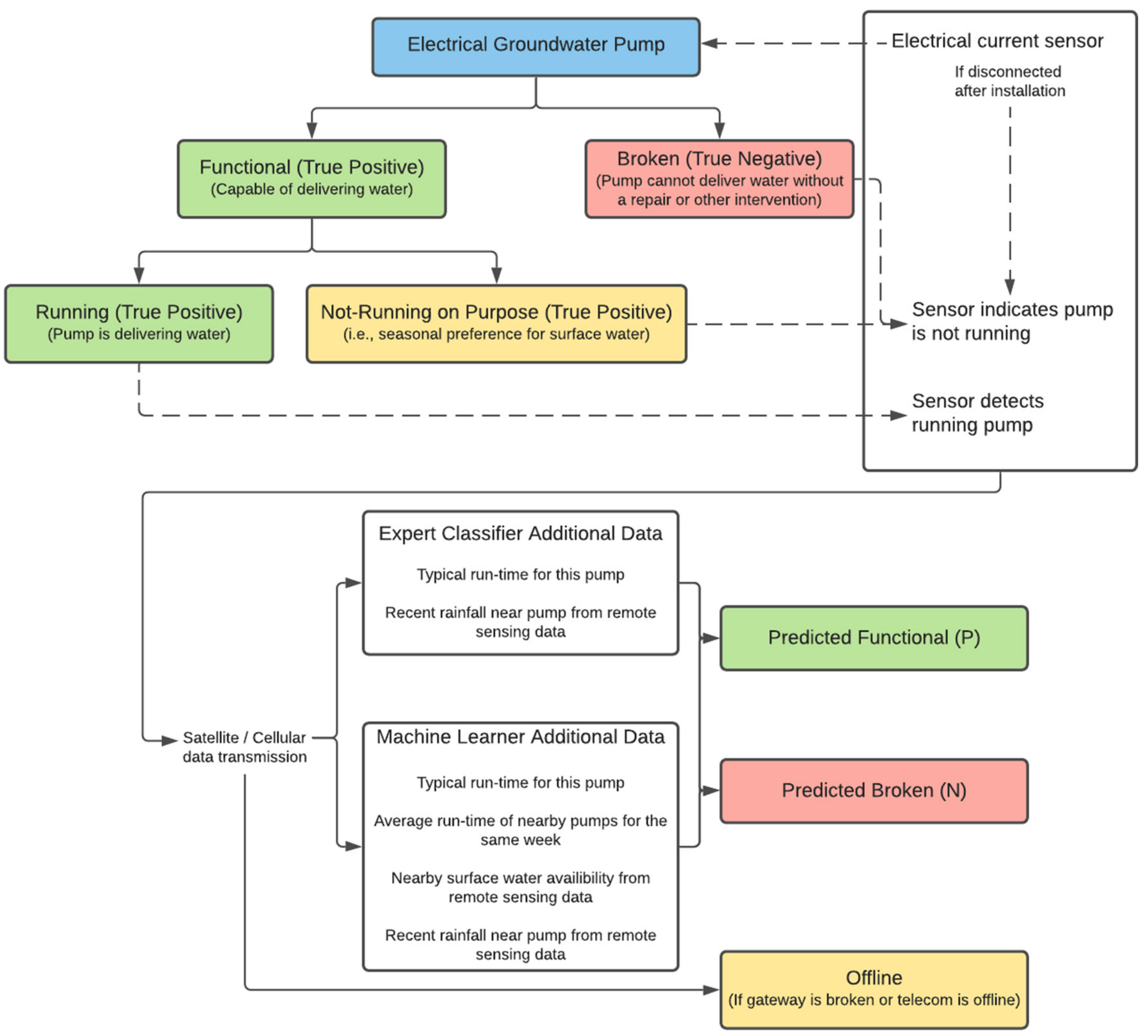

Fig. 2. Data system flowchart showing the operating conditions of the groundwater pump system including the three primary functional classification, the attached sensor instrumentation, the flow of data, and the predictions outputted by each of the classification systems.

2.1. Ground-truth data

In order to build and then validate our classification algorithms, we required a training set that reflects a ground-truth of pump conditions. Therefore, in collaboration with government and implementing partners we collected pump reports of water system status (functional, partly functional, or broken) using the mWater digital survey form platform (www.mwater.co). The partly functional status usually indicated that some repairs are required, but that the pump was capable of delivering water. Therefore, in our analysis we group partly functional and functional together to reflect conditions wherein water is delivered to users.

Pump reports were submitted during field visits and reviewed by program staff for data completeness. These reports provided details on ground-truth the status of both the pumps and the sensors. Pump reports were also used to identify false negatives (sensors reporting that a pump is not being used, when in reality it is) caused by accidental or deliberate disconnection of the sensor from the electrical power line. Sensor disconnections were identified via phone calls, site visits and documentation review. Sensor disconnections were validated based on interviewing of regional technicians who last serviced or inspected the scheme in question and recorded in the training data. An iterative process of pump report reviews, sensor data comparisons and verification through routine communication with field teams has led to a set of ground-truth training data for algorithm validation.

Table 1 describes the complete ground-truth data set, which includes a total of 1133 survey reports. About 73% of these surveys reflect a working water pump, about 15% indicate a broken water pump, and about 11% indicate that the sensor has been disconnected from the pump.

2.2. In-situ instrumentation and data cleaning

The principle of our pump monitoring system is to use pump motor operating current as a proxy for pump operation. A wireless current sensor is placed between the motor controller and pump. The presence of non-zero alternating current in the the conductors between the motor controller and the pump is used as a proxy for water pumping activity.

The current sensor used in this study is a Pressac brand split-core current transformer. This is an energy-harvesting sensor that uses energy scavenged from magnetic fields around current-carrying conductors to power a small microcontroller and 900 MHz radio. EnOcean is a proprietary ultra-low-power wireless harvesting and networking technology. If conductors are carrying enough current, the Pressac sensor is able to acquire a root mean square (RMS) current roughly once every 30 s and broadcast that data via the EnOcean protocol to a receiver.

The EnOcean receiver is a SweetSense-brand EnOcean gateway. The gateway receives the RMS current data from the Pressac sensor, compresses the data, caches it, and transmits it to the Internet once per day regardless of if any pump function has been observed, enabling a "heartbeat" from the sensor system to help identify broken sensor systems apart from non-used pumps. Depending on the local availability of cellular networks, the transmission is accomplished using either GSM cellular networks or the Iridium satellite network. We used Iridium's 9602 Short Burst Data Modem to transmit data. The 9602 SBD Modem can only transmit 340 bytes of data per transmission. Because of prohibitively expensive data rates, we limited transmissions to one per day. If a pump operated 24 h in a day, the Pressac sensor would transmit 2880 individual 13-bit floating point readings. Even when compressed, this is typically far too much data to fit into a single 340-byte SBD transmission. Therefore, we compress the data to a sampling rate of once every 40-minutes. When combined with additional meta and telemetry data, this constrains an average days-worth of data to one SBD transmission.

Table 1
Ground-truth training data collected through surveys. About 73% of these surveys reflect a working water pump, about 15% indicate a broken water pump, and about 11% indicate that the sensor has been disconnected from the pump.

Training data	Kenya	Ethiopia	Total
Total pump reports	333	800	1133
Functional water pump reports	221	610	831
Broken water pump reports	67	106	173
Sensor disconnection reports	45	84	129

There are four primary limitations to this approach to data collection from these pumps:

1. To fit data into SBD packets, data were sampled on a 40-minute basis, for a total of 36 possible observation of water pump runtime per day. In each 40-minute period, if runtime was observed, this status was applied to that window. This approach yields a possible measurement error in reported runtime of at least plus 40 min for each block of real-world pump runtime.
2. Electrical runtime does not explicitly reflect functionality or water demand, as pumps may be functional but deliberately not being used, often because surface water alternatives are available. Relatedly, given the variability in pump size, runtime does not equate to yield as smaller pumps will produce less water in the same period of time. Further, we do not consider population density in our model, despite the real-world variability in water demand based on population. Therefore, the simple measure of runtime is an imperfect proxy for yield and demand. This limitation motivates the contributions of this paper - we developed the classification systems described herein to enable more accurate classification of pump functionality and use given these limitations.
3. Occasionally, we observe only single data points in a day, corresponding to local midnight when our sensors transmit their data at a pump that otherwise had no operation that day and often had no observed operations within days. These data are unlikely to be real, given known periodic electrical noise in our system. Therefore, in our analysis we consider any reports of one data point corresponding to local midnight to be zero usage for that pump-day.
4. A challenging aspect of working with energy-harvesting sensors is that broken or disconnected sensors and a lack of energy to harvest result in the same sensor behavior. For example, if the Pressac sensor does not transmit data for one month, it can be difficult to determine if the lack of transmission is due to (1) a lack of current in the wire for the sensor to harvest (ie the pump is off), or (2) a broken energy-harvesting sensor. We have observed many instances of either deliberate or accidental damage or removal of the Pressac clamp (see Table 1, which yields false-negatives (false indications of pump non-use, or implied non-functionality). In our analysis, we do not discard this data, which therefore contributes to a greater false-negative performance. However, improved management of the sensor network would likely improve system performance.

In all cases of both these known sources of measurement error, as well as unknown sources of error, the validity and utility of the measurement and classification systems is reflected as performance metrics compared to the ground truth data.

To associate the site report ground-truth data with the daily sensor readings, the site report data needed to be extended to adjacent days when the apparent state was unchanged. Furthermore, in order to allow a direct comparison of the performance of both the expert classifier and machine learner models (described below) the ground-truth data must be formatted using assumptions that incorporate some of the subjective rules applied by the expert classifier.

To accomplish this, the sensor data was segmented into "blocks" of use or no use. Pump days were categorized as either usage (at least two sensor readings of activity) or non usage (zero or one sensor readings of activities). Contiguous runs of greater than 7-days of usage or non-usage were treated as a "block" of unchanging state. Intermittent periods of non-use less than 7-days was smoothed through grouping with the adjacent period. 7-day smoothing periods was selected consistent with the 7-day window used by the expert classifier, described below.

Finally, contiguous blocks were truncated at plus or minus 30 days from the ground-truth site report and a new block started, to minimize the risk of disparate conditions being grouped together. 30-days was chosen as an approximation of the Nyquist rate for sampling at twice the frequency of the signal measured (Candes and Wakin, 2008). In this case, the signal is rainfall seasonality and the responding change in borehole pump use. We truncate at 30 days plus or minus the report date to minimize the change that a block captures both rainy and dry seasons (in other words, periods in which the pump may be more or less likely to be in demand, regardless of function).

Any site report dated within a block was then assumed to reflect the true status of the pump for the entirety of the block. This approach is illustrated in Fig. 4 wherein ground-truth reports are shown as vertical bars (a verified broken pump in red, a verified working pump in green) and the associated blocks of sensor data affiliated with these site reports.

2.3. Remotely sensed data

Our work has previously used remotely sensed rainfall estimates from the Monthly Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) data (Funk et al., 2015b) to show that pump usage is inversely related to local rainfall, and that groundwater demand increases when absence of rainfall limits the availability of surface water alternatives (Thomas et al., 2019).

Further, when exploring available data, we observed a statistically significant correlation between surface water availability estimates derived from satellite observations, accessible through the Famine Early Warning System Surface Water Point Viewer (Senay et al., 2013) and borehole runtime. As these data are for specific water points that were surveyed in 2013 and do not capture our entire region of interest or any new water points constructed after 2013, we collect the "scaled depth" reported parameter and apply a inverse distance weighting algorithm (Thomson et al., 2008) to impute a normalized estimate of surface water availability at the local pump site. This normalized estimate is conceptualized as a percent from 0 to 100 of potential surface water retention proximate to the borehole. This estimate was available only for the Kenya data.

Therefore, in the predictive classifiers described in this paper, we incorporate remotely estimated data for:

1. Daily rainfall estimates using the CHIRPS system. To get rainfall data as soon as possible, we used the "preliminary" readings dataset. In this dataset, new daily rainfall data is available every five days with a two day lag. Each data point corresponds to a 0.05° area, or roughly 31 km^2.
2. Normalized daily surface water availability derived from an interpolation of the FEWS Water Point Viewer (Kenya only).

Similar to the discussion above regarding sensor measurement error, the validity and utility of these selections is reflected in the performance characteristics described in Results and Discussion.

2.4. Expert classification system

The sensor network leveraged for this work was first deployed in 2016. To make use of the data collected to support water pump operation and maintenance, we initially deployed an operational version of the machine learning (ML) based classification system described in this paper. However, while this approach has the advantage of continuous improvement as new data was incorporated, it was a "black box", and explaining the rules applied to classify a given pump was impossible. The term "black box" is commonly used when describing the limitations of machine learning. Because a machine learning statistical model is by definition non-mechanistic and deterministic, and instead relies on the relative weight of inputted predictors in determining an outcome, a machine learning based model cannot be reverse-engineered into clear logic statements. This approach has the well known limitation of reducing the confidence of regional level water managers using in the system as it was not easily understood or explained (Rudin, 2019).

An "expert classification" system was then developed to predict pump status using mapped out logic statements that could be more easily followed (Fig. 3). As implied in the term, such an expert classifier relies on domain knowledge expertise which, while informed, is still subjective and expert decisions may not be a consensus view (Valizadegan et al., 2013). In this case, we selected aggregation and threshold criteria based on discussions between and among users of our system, including local water system managers. The decisions made were subjective and non-systematic, however they reflect our best judgement on the balance between utility and interpretability of our method.

Daily pump status classification options are "Normal Use" when the preceding 7-day average runtime for a pump is consistent with the historical norm for the site; "Low Use" when the preceding 7-day average runtime is below the 20th percentile of use for the site; "No Use" when there is no observed pumping in the preceding 7 days, and "Seasonal Disuse" when the preceding 7-day period with no observed pumping coincides with CHIRPS rainfall data showing more than 10 mm within a 5 km radius of the site (the resolution of the CHIRPS data) during the previous week (i.e. 7–14 days prior to day of the status). There is also a "Repair" status applied when users complete electronic site reports indicating that a site is under repair, and an "Offline" status when sensors do not report data for more than 7 days. The pump status classifications occur on rolling 7-day increments to reduce over-sensitivity to daily variation in pump use. Classifications are made on a site-to-site basis as local characteristics largely determine pump use.

In all cases, we chose a 7-day averaging period to reflect the typical 7-day work week in both Ethiopia and Kenya, allowing an update to each pump classification to reflect a comprehensible weekly period. We selected 10 mm of rainfall as a common threshold for "moderate rainfall" using CHIRPS data (Bai et al., 2018).

We selected a rainfall period of 7–14 days prior to the current pump status classification consistent with our previous findings that borehole use increased in weeks following periods of no rainfall (Thomas et al., 2019). This is only one of many possible aggregations and lag periods possible. We selected one that was consistent with previous findings and easily explainable to users.

2.5. Machine learning supported prediction system

We sought to improve classification performance relative to the expert system by employing machine learning (ML) to better differentiate between the two "No Use" conditions: "Seasonal Disuse" and "Broken". To distinguish between "Use" and "No Use", we used the same rule employed in the expert classifier, as this was observed to match the ground truth data excellently. The ML was therefore only trained on blocks of time where no use occurred.

Supervised ML is a framework where a set of computerized models that can be thought of as black boxes are trained to data, in an attempt to generate predictions that closely match outcomes in a training dataset (Breiman, 2001a). These models can then be used to generate predictions for new observations. In this case, the outcome is borehole status "seasonal disuse" vs "broken", which is predicted from a number of "features" (covariates) which are described below. Features are other variables that are measured and which we believe may plausibly be predictive of the outcome. This is consistent with standard machine learning practices, where a large set of features thought to be plausible predictive of the outcome are provided to the ML algorithms, which often have an element of "feature selection" in which features that are actually predictive of the outcome are selected from among a large set

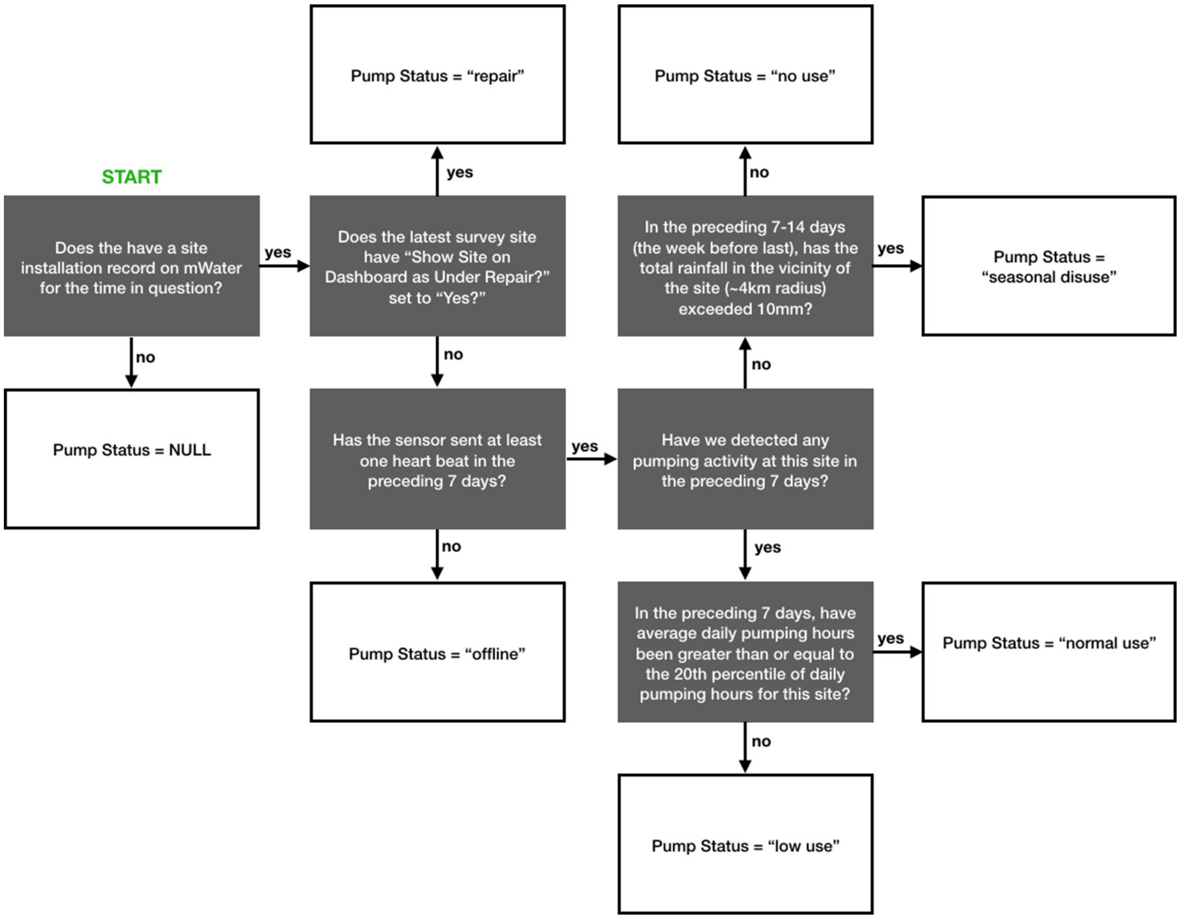

Fig. 3. Sensor data expert classification system flowchart.

of potential features. Most ML algorithms have some method of "feature selection" in which features that are not predictive of the outcome are ignored (Friedman, 2001). In the Results section we describe the outcome of the weighting of each of our selected features.

We hypothesized that ML would improve performance over the expert classifier system, as it is "trained" to the data, and therefore seeks to generate predictions that best match the training data. In contrast, the expert system was generated by human experts, and optimized not for agreement with the training set but for parsimoniousness (as few rules as possible), and face validity (the rules make sense to human experts).

We used ensemble ML, which combines predictions from a library of underlying candidate ML models ("learners"). We then used cross-validation to evaluate the performance of these learners and combine them. This approach allowed us to identify an optimal convex combination (weighted average) of an ensemble of candidate prediction algorithms (i.e. model stacking/Super Learning) (Zou et al., 2007; van der Laan et al., 2007). Our library consisted of 4 XGBoost learners (Chen and Guestrin, 2016) with different hyperparameter selections, LASSO, Ridge Regression (Friedman et al., 2010), Random Forests (Breiman, 2001b), and a null model. For each block, this yields a predicted probability that a pump is "broken". In practice, a threshold is then selected based on operational considerations to establish a binary classification of "broken" vs "no use". The threshold selected in our analysis is described in the Results section.

2.6. Features

In addition to determining an outcome for each pump-day in a block with a site report, we defined a number of features derived from both in-situ and remotely sensed data, as well as invariant attributes of the pump sites (such as administrative and geographic location). To the extent that these features capture relevant physical properties of the in-situ environment, our model reflects real-world relationships. During the model design, we explored data we collected and that which is publicly available to input into the model and examined the performance improvements associated with each feature selected. These features include the following:

1. Block Length - The length of a block of "use" or "no use" in days, ranging from 1 to 30. This continuous data feature was hypothesized to be relevant based on the assumption that long periods of no-use during periods of low surface water availability plausibly could indicate a broken pump.
2. Country - Kenya or Ethiopia. There are known differences between the two contexts in which we have sensor-instrumented pumps, including management approaches, budget constraints, hydrology and meteorology. By including the country location for each pump, we allow the classifier to benefit from any aggregate real-world average differences between these regions.
3. Rainfall - Derived from CHIRPS and justified as a possibly relevant feature as described above. Aggregated at 1 and 2 week intervals as separate features.
4. Surface Water - Derived from FEWS Water Point Viewer and justified as a possibly relevant feature as described above.
5. Relative Usage - Sensor-measured u of an electrical pump compared to its neighbors, using two definitions:

   - Geographic Neighbor - Usage at a pump compared to geographically nearby pumps. We hypothesize that nearby pumps may be used similarly to one another, given geographic and population density similarities.

- Correlated Neighbor - Usage at a pump compared to pumps similar in usage behavior. We hypothesized that pumps that behave similar to one another, regardless of if they are nearby geographically, may reflect other non-defined characteristics, such as urban/rural, population densities, migratory routes, political boundaries, and other features we do not provide the learner.

These two definitions yielded similar results, therefore in this paper we present correlated neighbor only.

2.7. Example data

Fig. 4 presents an example data set from one pump, to illustrate the features used by the expert classification and machine learning algorithms. In this example, the top panel shows sensor-measured pump use per day. While some pump use is observed during the period, the three highlighted blocks that indicate a ground-truth report all show zero runtime. This is a good example of the challenge inherent in our approach - how to distingush between these periods. The second panel shows CHIRPS detected rainfall in the previous 7 days. The third panel shows a comparison of the current pump of interest's behavior relative to statistically correlated neighbors. In this case, this pump is running less often than similar pumps during this period. In the final panel is an estimate of surface water availability derived from an interpolation of the FEWS Water Point Viewer. The vertical lines are ground-truth reports shown as a verified broken pump in red (true negative) and a verified working pump in green (true positive). The data are colored showing the functional / non-functional prediction concurrent with these site reports.

3. Results

3.1. Model validation and performance

We present the performance of both the expert classifier and the machine learning algorithm, validated against the ground-truth pump reports. First, the receiver operator characteristic (ROC) is presented

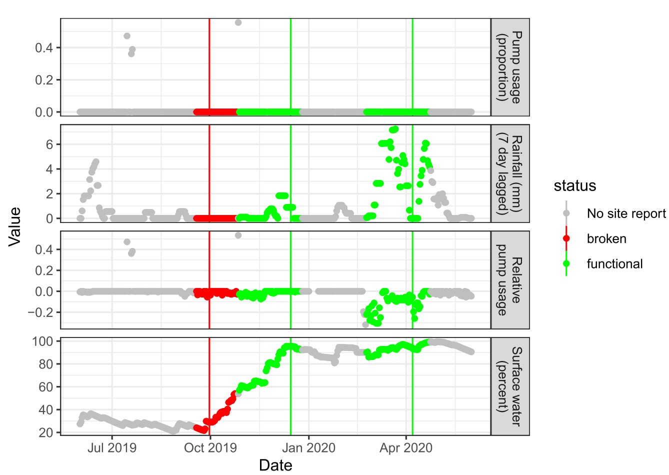

Fig. 4. Example sensor and remote sensing data with ground-truth pump reports and machine-learning predicted status. In this example, the top panel shows sensor-detected usage as a proportion of a 24 h day (i.e., 0.5 is 12 h of runtime). The second panel shows total rainfall in the previous 7-days, as indicated using the CHIRPS 5 km resolution remotely sensed rainfall product. The third panel shows a metric of the relative runtime of this pump compared to its neighbors. A negative value indicates lower than average runtime. The bottom panel is a measure of surface water depth, estimated as an interpolation of the FEWS Water Point Viewer. The colored sections correspond to ground-truth pump reports collected. In red, a pump report indicates the pump is actually broken. In green, the pump report indicates the pump is actually functional. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

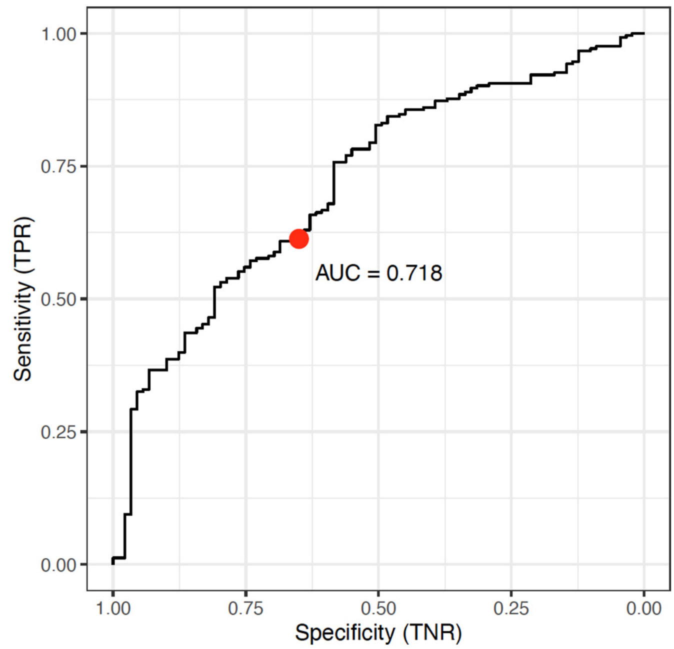

Fig. 5. The receiver operator characteristic (ROC) for the machine learning algorithm. The ROC represents the range of possible trade-offs between the classifiers' true positive (truly working pump classified as working, the sensitivity or true positive rate) and true negative (truly broken pump classified as broken, the specificity or true negative rate) rates when choosing a threshold to operationalize the classifier. The threshold of 0.72 selected in our analysis is shown as the red dot, while the total area under the curve (AUC) is specified as 0.718. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for the machine learning algorithm in Fig. 5. A ROC curve is a plot that illustrates the performance of a binary classifer system (in this case, a working or broken pump) based on a trade-off between specificity and sensitivity. Operationally, a threshold is chosen as a point on the curve meeting desired design criteria. A rule of thumb is that the most upper-left point on the curve is a threshold that optimizes sensitivity and specificity for a generic case (Hanley and McNeil, 1982). In this case, the ROC represents the range of possible trade-offs between the classifiers' true positive (truly working pump classified as working, the sensitivity or true positive rate) and true negative (truly broken pump classified as broken, the specificity or true negative rate) rates when choosing a threshold to operationalize the classifier.

Fig. 6 shows the relative importance of variables for the predictive performance of the model. Using a squared error loss, each variable is plotted against the increase in risk (decrease in predictive performance) associated with scrambling the values of that variable (Breiman, 2001b). This produces a ranking of variables, from most to least important. We note that Block Length is the most important variable weighed in the model. As a continuous variable between 1 and 30 days (with longer blocks truncated as previously described), the intuitive interpretation is simply that the longer a pump is not running, the more likely it is to be truly broken. Truncated blocks do not affect this interpretation, but instead provide additional data for the algorithm to re-estimate a pump's condition based on all available data.

In order to effectively compare these two classifiers performance in identifying functional and broken pumps, we must select a classification threshold for the machine learner that generates the same number of broken predictions as the expert classifier. A classification threshold of 0.72 was used to yield the same number of predicted broken periods as the expert classifier (149 periods). Using this, we compare the performance of the two classifier approaches.

Table 2 presents performance characteristics for the expert classifier and machine learning algorithm. We present three performance specifications including a.) All of the available data corresponding to

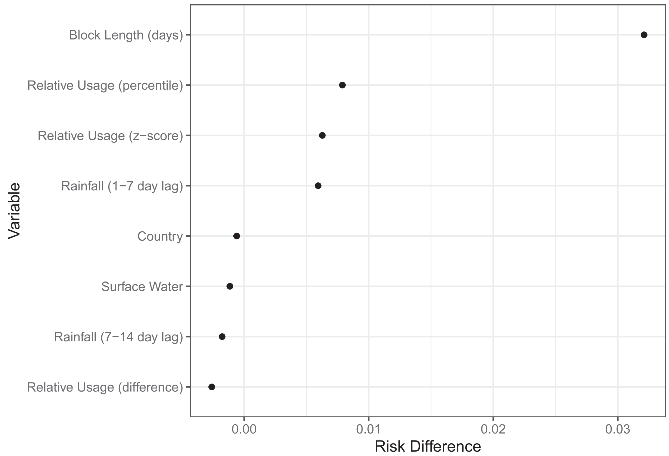

Fig. 6. The relative importance of each variable included in the machine learning model. The risk difference indicates the change in the mean squared error to the performance of the classifier. Conceptually, this plot can be interpreted as indicating that Block Length is approximately three times as important to the model's performance as Relative Usage.

ground-truth observations for both running and non-running pumps, b.) the performance when applied only to periods of time where some runtime is observed by the sensors, and c.) the performance when applied only to periods when no runtime is observed.

The machine learning algorithm out-performs the expert classifer across almost every metric. Our validation indicates an overall functioning pump detection sensitivity (true positive rate) of 82% for the expert classifier and 84% for the machine learner. When a pump is being used, both the expert classifier and the machine learner have a sensitivity of identifying these as functional pumps (true positive rate) 100% of the time.

When a pump is not being used, sensitivity tTrue Positive rate) drops to about 50% for the expert classifier and about 65% for the machine learner. These results indicate that for a pump that is not running, the expert classifier is no better than chance at identifying the difference between the broken and not-working-on-purpose conditions, while the machine learner exceeds this performance.

3.2. Implications for pump functionality

We then apply both the expert classifier and machine learning algorithm both to our training data (nearly 90 pump-years). Table 3 presents the impact on uptime for the training set if either the expert classifier or machine learning algorithm were used to trigger pump visits and, when appropriate, pump repairs. In all cases, we assume a 2-week lag between an identified potential failure and a visit or repair. This lag avoids the unreasonable implication that a pump identified as broken could immediately be returned to service. A 2-week period from identification to repair is approximately consistent with best practices observed of some of our stakeholder users.

Table 2
Classifier Performance for Expert Classifier and Machine Learner wherein a True Positive is a functional pump capable of delivering water regardless of actual current use, and a True Negative is a broken pump incapable of delivering water without a repair.

Performance metric	Expert classifier	Machine learner
All data (running and non-running pumps)		
True positive rate (sensitivity)	82.1%	84.5%
True negative rate (specificity)	47.8%	63.0%
Positive predictive value	90.9%	93.6%
Negative predictive value	29.5%	38.9%
Usage observed (running pumps)		
True positive rate (sensitivity)	100.0%	100.0%
True negative rate (specificity)	0.0%	0.0%
Positive predictive value	99.1%	99.1%
Negative predictive value	NA	NA
No usage observed		
True positive rate (sensitivity)	56.2%	62.1%
True negative rate (specificity)	49.4%	65.2%
Positive predictive value	75.0%	82.8%
Negative predictive value	29.5%	38.9%

Table 3
Modeled potential impact on pump functionality using training data.

	Units	Ethiopia	Kenya	Overall
Total observations	Pump-years	60.3	29.4	89.8
Uptime observed	%	88.2	77.6	84.6
Broken pump events observed	#/Pump/year	0.76	1.56	1.06
Expert classification algorithm				
Repairs attempted (negative prediction)	#/Pump/Year	1.35	2.28	1.66
Repairs made (true negative)	#/Pump/Year	0.40	0.68	0.49
Potential uptime (2 week repair lag)	%	92.9	84.7	90.1
Potential uptime change	%	5.3	9.1	6.5
Machine learning algorithm				
Repairs attempted (negative prediction)	#/Pump/Year	1.29	2.4	1.66
Repairs made (true negative)	#/Pump/Year	0.51	0.92	0.65
Potential uptime (2 week repair lag)	%	94.7	89.1	92.6
Potential uptime change	%	7.4	14.8	9.5

4. Discussion

4.1. Deployment

The expert classification system evaluated in this paper has been operationally deployed in both the Kenya and Ethiopia contexts since 2017. Our validation indicates that this system accurately classifies functioning pumps (including those currently not being used, but otherwise functional), with a sensitivity (true positive rate) of over 82% of the time. When a pump is being used, both statistical models accurately identify these as functional pumps (true positives) 100% of the time. When a pump is not being used, the expert classifier accurately identifies that such a pump is broken with a specificity of about 50%. The machine learning model improves on this performance with a overall functioning pump sensitivity (true positive rate) of nearly 85%, and a specificity when a pump is not running of over 65%. These performance differences highlight the improved capacity of the machine learner to identify the difference between a broken pump and a pump not running on purpose, despite the same raw data characteristics of no-runtime provided by the sensors, better addressing this nuance of pump functionality by incorporating seasonality and other features described previously.

4.2. Potential impact

The training set collected in this study indicated high uptime averages of over 88% in Ethiopia and over 77% in Kenya. This training set is likely biased toward higher functionality than the broader pump population, as functional pump reports were more readily collected with phone calls, in contrast to typically requiring site-visits to confirm the nature of non-functionality.

In contrast, recent reports reviewing borehole pump functionality during the 2016–2017 drought in East Africa indicated 55% in Kenya (UNICEF, 2017) and 60% in Ethiopia (MacAllister et al., 2020). These low levels of pump functionality are the underlying motivation for the work described in this paper.

Therefore, if we assume an average drought-period uptime of 60% in the region, repair efforts triggered through the sensor and machine learning system could detect and repair 62.1% of broken pumps (sensitivity), with a potential false-alarm rate (false negative rate) of 34.8% (1-specificity). While this may at first appear to be a nuisance high false alarm rate, in absence of such a sensor network indicating both running and non-running pumps, the alternative to visiting the occasional functional pump that does not require a repair is visiting all pumps on a regular schedule. Furthermore, we note that the status-quo approach wherein operators and community members report on pump breakdowns is presently effecting the 55% - 60% functionality rates reported. By visiting only those sites where the machine learner reports a breakdown, this could, in practice, result in a drought-period uptime of 84.8% - a 40% reduction in the relative risk of pump downtime.

Next, we discuss the potential cost and cost-effectiveness of such a repair model. While precise estimates are challenging to establish, for the purpose of illustration we estimate that each pump serves approximately 5000 people per week (Thomas et al., 2019; Thomas et al., 2020b), and costs on average \$4.5 per person. This estimate is based on the average per capita cost of water supply in comparable contexts in Ethiopia (\$3) and Kenya (\$6) (Libey et al., 2020). Therefore an illustrative annual cost per pump is at least \$20,000. If we assume that only 60% of demand for pump function is met during the dry and drought seasons (4 months per year), the effective cost of water supply per person increases to about \$6.

We estimate the annual cost per pump adding telemetry-connected sensors at \$250 (A cost of \$1000 per sensor over a 4 year lifetime), the cost of a false-alarm visit at \$500 (a rate of 1.66 visits per pump per year based on our model this adds \$830 per pump per year), and the cost of a major pump repair at \$1800 per year (a rate of 0.65 repairs per pump per year, this adds \$1170 per pump per year). These cost estimates are the median responses of community informant interviews we conducted in both our Ethiopia and Kenya context. We acknowledge that actual site visits and pump repair costs are highly variable and difficult to capture in this simplified illustrative example.

Using these simplified assumptions, a plausible revised cost per-pump cost of \$23,000 attributable to both false-alarm and pump-repair visits increases budgetary requirements by 15%. However, the increase in functionality during the drought season from 60% to nearly 85% has the potential to offset this cost by decreasing the effective annual cost per capita to about \$5.25, a marginal but perhaps meaningful cost effectiveness improvement of over 12%. We acknowledge that these are illustrative numbers only, with sensitive assumptions. However, in practice we observe that the increased costs of monitoring and repairs of rural water supplies should be readily offset by the cost-effectiveness of higher pump functionality, water access, and water security. Indeed, a 2018 study by USAID estimated that each \$1 invested in resilience in drought prone areas of Kenya, Ethiopia and Somalia results in \$3 in savings in averted losses and humanitarian need (USAID, 2018).

4.3. Limitations and future work

The classification systems we developed and present in this paper include various assumptions that impact the performance and subsequent utility of each model. As our work was motivated by improving the utility of the real-world deployment of this pump sensor network, in contrast to designing an optimal statistical model, we were functionally constrained in the number of additional features or sensitivity analyses we conducted for each assumption or selection. In the case of the machine learning model presented herein, we selected data features that were available to us ex post. The machine learning model's performance may be further improved through the identification and inclusion of additional data features especially those that may require additional in-situ data collection. These may include, for example, improved estimates on water yield or groundwater depth at each pump site, or improved estimates of population and livestock density.

A common question raised when weighing the potential benefit versus cost of an instrumentation system to support pump repairs is, simply, could a sensor be replaced with a pump operator or other community authority calling a service provider? We remind the reader that this approach effectively reflects the status-quo, and the intent of the sensor-triggered system is to increase transparency, accountability, responsiveness and ultimately improve water services. Therefore, in our indicative cost estimates, we assume that management and cost structures can be shifted toward responding regularly to the identified potential pump failures.

Work by ourselves and others has demonstrated significant improvements to water delivery when service providers have access to real-time information (MacAllister et al., 2020; Nagel et al., 2015; Thomson, 2020). However, there is a need for better understanding of the enabling and hindering factors that affect the utilization of real-time data and effective adoption of sensor-based technologies to improve water service delivery in low-income settings. In both our Ethiopia and Kenya project settings, adoption of these systems has been limited by the nearly non-existent budgetary allocations for borehole pump repairs. While these types of technologies have been demonstrated to support improved water pump functionality, it is also a clear prerequisite that direct budget and institutional incentives are required to make use of these tools (Thomas and Brown, 2020). In absence of direct budgetary support to act on pump breakdowns, we have also observed the value of the sensor-data network in supporting higher level decision making including through collaborations with the Kenya National Drought Management Authority, The Nairobi based Regional Centre for Mapping of Resources for Development, and the USAID and NASA supported Famine Early Warning Systems Network (Thomas et al., 2020b). However, it is clear that if any cost-effective is to be derived from these technologies on a local level, donors, governments and communities must allocate resources to enable direct action toward improved water system functionality.

Funding

The work described in this publication was made possible through support provided by the National Aeronautics and Space Administration, under the terms of Grant No. 80NSSC20K0150; the National Science Foundation under the terms of Award No. 1738321; the U.S. Agency for International Development, under the terms of Contract No. AID-663-TO-16-00001 and Cooperative Agreement No. AID-615-A-15-00008. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Agency for International Development.

CRediT authorship contribution statement

Evan Thomas: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft. **Daniel Wilson:** Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. **Styvers Kathuni:** Conceptualization, Investigation, Writing – original draft. **Anna Libey:** Data curation, Investigation, Writing – original draft. **Pranav Chintalapati:** Data curation, Investigation, Writing – original draft. **Jeremy Coyle:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Authors Thomas, Wilson, Kathuni are compensated employees of SweetSense Inc., the company providing the sensor products used to collect the data analyzed in this paper. The authors declare no other conflicts of interest.

References

Ahmad, R., Kamaruddin, S., 2012. An overview of time-based and condition-based maintenance in industrial application. Computers and Industrial Engineering.

Bahru, B.A., Bosch, C., Birner, R., Zeller, M., 2019. Drought and child undernutrition in Ethiopia: a longitudinal path analysis. A PLoS ONE 14 (7), e0220075.

Bai, L., Shi, C., Li, L., Yang, Y., Wu, J., 2018. Accuracy of CHIRPS satellite-rainfall products over mainland China. Remote Sens. 10 (3), 362.
Banks, E.W., Cook, P.G., Owor, M., Okullo, J., Kebede, S., Nedaw, D., Mleta, P., Fallas, H., Gooddy, D., John MacAllister, D., Mkandawire, T., Makuluni, P., Shaba, C.E., MacDonald, A.M., 2020. Environmental tracers to evaluate groundwater residence times and water quality risk in shallow unconfined aquifers in sub Saharan Africa. Journal of Hydrology 125753. https://doi.org/10.1016/j.jhydrol.2020.125753.
Bonsor, H., MacDonald, A., Casey, V., Carter, R., Wilson, P., 2018. The need for a standard approach to assessing the functionality of rural community water supplies. Hydrogeology Journal 26, 367–370.
Borgomeo, E., Vadheim, B., Woldeyes, F., Alamirew, T., Tamru, S., Charles, K., Kebede, S., Walker, O., 2018. The distributional and multi-sectoral impacts of rainfall shocks: evidence from computable general equilibrium modelling for the Awash Basin, Ethiopia. Ecological Economics 146, 621–632.
Breiman, L., 2001a. Statistical modeling: the two cultures. Stat. Sci. 16 (3), 199–231.
Breiman, L., 2001b. Random forests. Machine Learning. https://link.springer.com/article/10.1023/A:1010933404324.
Calow, R.C., MacDonald, A.M., Nicol, A.L., Robins, N.S., 2010. Ground water security and drought in Africa: linking availability, access, and demand. Ground Water 48 (2), 246–256.
Candes, E., Wakin, M., 2008. An introduction to compressive sampling. IEEE Signal Processing Magazine.
Carter, R.C., Ross, I., 2016. Beyond 'functionality' of handpumpsupplied rural water services in developing countries. Waterlines 35, 94–110.
Chen, T., Guestrin, C., 2016. XGBoost.
Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33 (1), 1–22.
Funk, C., Shukla, S., Hoell, A., Livneh, B., 2015a. Assessing the Contributions of East African and West Pacific Warming to the 2014 Boreal Spring East African Drought.
Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015b. The climate hazards infrared precipitation with stations–a new environmental record for monitoring extremes. Sci Data 2, 150066.
Godfrey, S., Hailemichael, G., 2017. Life cycle cost analysis of water supply infrastructure affected by low rainfall in Ethiopia. Journal of Water Sanitation and Hygiene for Development 7 (4), 601–610.
Gramling, C., 2013. Water resources. Kenyan find heralds new era in water prospecting. Science 341, 1327.
Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143 (1), 29–36.
Kebede, S., Travi, Y., Asrat, A., Alemayehu, T., Ayenew, T., Tessema, Z., 2008. Groundwater origin and flow along selected transects in Ethiopian rift volcanic aquifers. Hydrogeol. J. 16, 55–73.
Knippenberg, E., Hoddinott, J., 2017. Shocks, Social Protection, and Resilience: Evidence From Ethiopia Ethiopian Development Research Institute Working Paper. http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/131408/filename/131619.pdf.
Koehler, J., 2018. Exploring policy perceptions and responsibility of devolved decision-making for water service delivery in Kenya's 47 county governments. Geoforum 92, 68–80.
Koehler, J., Thomson, P., Hope, R., 2015. Pump-priming payments for sustainable water services in rural Africa. World Dev. 74, 397–411.
Libey, A., Adank, M., Thomas, E., 2020. Who pays for water? Comparing life cycle costs of water services among several low, medium and high-income utilities. World Dev. 136.
Liou, Y. A.; Mulualem, G. M. Spatio-temporal assessment of drought in Ethiopia and the impact of recent intense droughts. Remote Sensing 2019.
Lolemtum, J.T., Mugalavai, E.M., Obiri, J.A., 2017. Impact of drought on food security in West Pokot County, Kenya. International Journal of Scientific and Research Publications 7 (6), 742–750.
MacAllister, D. J.; MacDonald, A. M.; Kebede, S.; Godfrey, S.; Calow, R. Comparative performance of rural water supplies during drought. Nature Communications 2020.
Macdonald, A.M., Bell, R.A., Kebede, S., Azagegn, T., Yehualaeshet, T., Pichon, F., Young, M., McKenzie, A.A., Lapworth, D.J., Black, E., Calow, R.C., 2019. Groundwater and resilience to drought in the Ethiopian highlands. Environmental Research Letters 14, 095003.
MacDonald, A.M., Lark, R.M., Taylor, R.G., Abiye, T., Fallas, H.C., Favreau, G., Goni, I.B., Kebede, S., Scanlon, B., Sorensen, J.P.R., Tijani, M., Upton, K.A., West, C., 2021. Mapping groundwater recharge in Africa from ground observations and implications for water security. Environ. Res. Lett. 16, 034012.
Nagel, C., Beach, J., Iribagiza, C., Thomas, E.A., 2015. Evaluating cellular instrumentation on rural Handpumps to improve service delivery-a longitudinal study in rural Rwanda. Environ. Sci. Technol. 49, 14292–14300.
Nicholson, S.E., 2014. A detailed look at the recent drought situation in the Greater Horn of Africa. J. Arid Environ. 103, 71–79.
Nthambi, M., Ijioma, U.D., 2020. African Handbook of Climate Change Adaptation.
Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020. Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access.
Rudin, C., 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.
Senay, G.B., Velpuri, N.M., Alemu, H., Md Pervez, S., Asante, K.O., Kariuki, G., Taa, A., Angerer, J., 2013. Establishing an operational waterhole monitoring system using satellite data and hydrologic modelling: application in the pastoral regions of East Africa. Pastoralism 3 (20).
Shabelle, L., 2011. Drought-related food insecurity: a focus on the Horn of Africa. Drought Emergency. Food and Agricultural Organization.
Thomas, E., Brown, J., 2020. Using feedback to improve accountability in global environmental health and engineering. Environmental Science & Technology *0*.
Thomas, E., Jordan, E., Linden, K., Mogesse, B., Hailu, T., Jirma, H., Thomson, P., Koehler, J., Collins, G., 2020a. Reducing drought emergencies in the Horn of Africa. Sci. Total Environ. 727, 138772.
Thomas, E., Kathuni, S., Wilson, D., Muragijimana, C., Sharpe, T., Kaberia, D., Macharia, D., Kebede, A., Birhane, P., 2020b. The drought resilience impact platform (DRIP) - improving water security through actionable water management insights. Frontiers in Climate 727, 138772.
Thomas, E.A., Needoba, J., Kaberia, D., Butterworth, J., Adams, E.C., Oduor, P., Macharia, D., Mitheu, F., Mugo, R., Nagel, C., 2019. Quantifying increased groundwater demand from prolonged drought in the east African Rift Valley. Sci. Total Environ. 666, 1265–1272.
Thomson, H., Petticrew, M., Morrison, D., 2008. Geoscience Australia.
Thomson, P., 2020. Remote Monitoring of Rural Water Systems: A Pathway to Improved Performance and Sustainability?
Thomson, P., Koehler, J., 2016. Performance-oriented monitoring for the water SDG – challenges. Tensions and Opportunities. Aquatic Procedia 6, 87–95.
Thomson, P., Hope, R.A., Foster, T., 2012. Is silence golden? of mobiles, monitoring,and rural water supplies. Waterlines 31 (4).
Thomson, P., Bradley, D., Katilu, A., Katuva, J., Lanzoni, M., Koehler, J., Hope, R., 2019. Rainfall and groundwater use in rural Kenya. Sci. Total Environ. 649, 727–730.
Tucker, J., MacDonald, A., Coulter, L., Calow, R.C., 2014. Household water use, poverty and seasonality: wealth effects, labour constraints, and minimal consumption in Ethiopia. Water Resources and Rural Development 3, 27–47.
Turman-Bryant, N., Nagel, C., Stover, L., Muragijimana, C., Thomas, E.A., 2019. Improved drought resilience through continuous water service monitoring and specialized institutions—a longitudinal analysis of water service delivery across motorized boreholes in northern Kenya. Sustainability 11.
Uhe, P., Philip, S., Kew, S., Shah, K., Kimutai, J., Mwangi, E., van Oldenborgh, G., Singh, R., Arrighi, J., Jjemba, E., Cullen, H., Otto, F., 2017. Attributing drives of the 2016 Kenyan drought. Int. J. Climatol. (38), e554–e568.
UNICEF, 2017. Kenya Humanitarian Situation Report.
USAID, 2018. Economics of Resilience to Drought in Ethiopia, Kenya and Somalia.
Valizadegan, H.; Nguyen, Q.; Hauskrecht, M. Learning classification models from multiple experts. Journal of Biomedical Informatics 2013.
van der Laan, M.J., Polley, E.C., Hubbard, A.E., 2007. Super learner. Stat. Appl. Genet. Mol. Biol. 6 (25). https://doi.org/10.2202/1544-6115.1309.
Venton, C., 2018. Economics of Resilience to Drought - Kenya Analysis.
Wilson, D.L., Coyle, J.R., Thomas, E.A., 2017. Ensemble machine learning and forecasting can achieve 99% uptime for rural handpumps. PLoS One 12, e0188808.
Zou, H., Hastie, T., Tibshirani, R., 2007. On the "degrees of freedom" of the lasso. Ann. Stat. 35 (5), 2173–2192.