Toilet Alarms: A Novel Application of Latrine Sensors and Machine Learning for Optimizing Sanitation Services in Informal Settlements

Nick Turman-Bryant[a], Taylor Sharpe[b], Corey Nagel[c], Lauren Stover[d], Evan A. Thomas[c,*]

[a] *Department of Systems Science, Portland State University, Portland, Oregon, USA*
[b] *Department of Mechanical Engineering, Portland State University, Portland, Oregon, USA*
[c] *School of Public Health, Oregon Health and Science University, Portland State University, Portland, Oregon, USA*
[d] *Operations Research, Sanergy, Nairobi, Kenya*

Abstract

The cost-effectiveness and reliability of waste collection services in informal settlements can be difficult to optimize given the geospatial and temporal variability of latrine use. Daily servicing to avoid overflow events is inefficient, but dynamic scheduling of latrine servicing could reduce costs by providing just-in-time servicing for latrines. This study used cellular-connected motion sensors and machine learning to dynamically predict when daily latrine servicing could be skipped with a low risk of overflow. Sensors monitored daily latrine activity, and enumerators collected solid and liquid waste weight data. Given the complex relationship between latrine use and the need for servicing, an ensemble machine learning algorithm (Super Learner) was used to estimate waste weights and predict overflow events to facilitate dynamic scheduling. Accuracy of waste weight predictions based on sensor and historical weight data was adequate for estimating latrine fill levels (mean error of 20% and 22% for solid and liquid wastes), but there was greater accuracy in predicting overflow events (area under the receiver operating characteristic curve of 0.90). Although our simulations indicate that dynamic scheduling could substantially reduce costs for lower use

*Corresponding author
Email address: ethomas@colorado.edu (Evan A. Thomas)
URL: https://www.colorado.edu/mcedc/ (Evan A. Thomas)

latrines, we found that cost reduction was more modest for higher use latrines and that there was a significant gap between the simulated and implemented results.
Keywords: `sanitation, passive latrine use monitors (PLUMs), machine learning, information and communication technologies (ICTs), Super Learner`

1. Introduction

Globally, at least 2.3 billion people do not have access to improved sanitation facilities, and 4.5 billion people do not have access to safely managed sanitation services (UNICEF / WHO, 2017). While much attention has been focused on latrines for rural populations and campaigns to end open defecation (UNICEF / WHO, 2017; Robiarto et al., 2014; Trémolet, 2011; Coffey et al., 2014), the need for improved and safely managed sanitation facilities is acute in dense informal settlements in rapidly urbanizing areas (Bohnert et al., 2016; Brown et al., 2015). This need has three principal drivers: the high population density of informal settlements, the lack of institutional sanitation providers, and the challenge of safely transporting fecal waste out of the settlement (Paterson et al., 2007; Mara, 2012).

Today, more than half of humanity lives in a city. In low income countries the trend toward urban migration is particularly strong, with 31% of the population residing in urban areas and 4.2% of the population migrating to cities each year (United Nations Department of Economic and Social Affairs, 2015). However, urban growth and infrastructure development has often not been able to keep pace with the rapid influx of individuals and families, resulting in the formation of informal settlements and squatter's communities that lack basic water, sanitation, or electrical services (United Nations, 2015). The lack of sanitation services in informal settlements is particularly problematic, as fecal deposition in high traffic environments combined with increased residential density can greatly increase the risk of enteric infections (Kimani-Murage et al., 2014; Bhagwan et al., 2008). For example, children in Nairobi's informal settlements have a prevalence of diarrhea (20.2%) that is comparable to prevalences in rural Kenya (21.7%) but much greater than the rate reported for Nairobi at large (14.8%) (African Population and Health Research Center, 2014).

Attempts to provide reliable and appropriate sanitation services in informal settlements are often limited by the lack of legal protections, property ownership, resistance from governing authorities, and minimal water and sewage infrastructure (Bohnert et al., 2016). Given the lack of support from governments, sanitation solutions in informal settlements often depend on non-profits or social enterprises that rely on donations or revenue generating models to sustain services (Auerbach, 2016).

One of the key factors influencing the cost-effectiveness and reliability of service provision in informal settlements is the ability to optimize waste collection from latrines with variable use patterns that are spatially dispersed within an informal settlement. Optimization of latrine servicing typically implies a trade-off between increased collection efficiency and increased risk of latrine overflow events. Daily servicing effectively avoids the risk of latrine overflow, but inefficient servicing of latrines (i.e., servicing latrines before they are full) may not be cost-effective. On the other hand, less frequent servicing increases the likelihood of a latrine overflow event, which can be damaging to the operator's reputation, result in decreased demand or willingness-to-pay for services, as well as increase the risk of exposure to fecal contamination. Ideally, latrines would be serviced with the highest efficiency possible, but to do so requires real-or near-time monitoring of latrine fill levels (i.e., the fullness of the solid and liquid waste receptacles). In previous studies motion detector sensors (passive latrine use monitors - PLUMs) have been used to monitor latrine activity and compared against self-reported latrine use or observed latrine use (Delea et al., 2017; Bohnert et al., 2016; Sinha et al., 2016; O'Reilly et al., 2015). However, there are no known studies that attempt to estimate the accumulated solid or liquid waste detected using a latrine sensor.

Partnering with Sanergy Inc., an established sanitation service provider for informal settlements in Nairobi, Kenya, researchers from Portland State University and Sweet Sense investigated how latrine sensors could be used to estimate waste fill levels and improve servicing efficiency for forty latrines in Nairobi, Kenya. In particular, we evaluated (1) how accurately we could estimate solid and liquid waste weights based on motion sensor data, (2) how accurately we could predict a latrine overflow event to create a dynamic schedule for latrine servicing, and (3) how cost-effective sensor-enabled servicing would be compared to daily servicing or servicing based on data from on-site weighing. In order to answer these questions we developed three models to simulate the predictive performance and cost-effectiveness of dynamic scheduling in relation to Sanergy's existing static schedule. We also present the results from a dynamic schedule that was implemented over three months and compare its performance to the existing and simulated scheduling scenarios.

2. Materials and Methods

For this study a convenience sample of forty latrines was selected for installing the motion sensors. These forty latrines were chosen because they were clustered along a service route that was close to the central office and had reliable waste collector personnel. Forty-one latrines from a nearby route were selected as the comparison group to estimate outcome variables at baseline and after the intervention (see Table 1). General characteristics of each latrine were obtained from Sanergy's existing records (i.e., type of latrine, responsible waste collectors and field officers, and collection schedule).

In addition, three enumerators were employed to manually weigh and record daily on-site solid and liquid waste weights each time a latrine was serviced in the intervention and comparison groups. Weight measurements were recorded using the following procedure: (1) enumerators accompanied waste collectors each morning to each of the latrines designated for servicing; (2) at each latrine waste collectors removed the solid and liquid waste cartridges and weighed each cartridge using a hanging scale (see TOC image); (3) weights were manually recorded by the enumerators using a mobile application that did not rely on cellular network connectivity; (4) weight measurements were uploaded to the survey server each afternoon when enumerators returned to the main office; (5) an automated algorithm compiled weight records from the survey, subtracted the weight of the empty solid and liquid waste cartridges, and compared the list of latrines serviced against the list of latrines scheduled for servicing to account for missing data or discrepancies. Enumerators were also responsible for installing, trouble-shooting, and swapping out sensors when batteries were running low or sensors were not reporting. Sensors were installed in October, 2016, and three months of baseline weight and sensor data were collected before the intervention period from January through March, 2017. During the baseline period, all latrines were scheduled for servicing according to Sanergy's static schedule, whereas during the intervention period latrines with sensors were serviced using a dynamic schedule (both schedules described in further detail below). The purpose of the experiment was to see whether collection efficiency improved in the latrines with sensors during the intervention period when weight and sensor data were used to generate a dynamic servicing schedule.

The sensor unit was equipped with a passive infrared motion sensor that logged movement in the latrine throughout the day and transmitted the data each evening via a GSM radio to Sweet Sense servers. After all the sensors had called in, an automated algorithm was executed to compile all the weight and motion sensor data and run the machine learning algorithm to determine which latrines could be skipped the next day. During the intervention period,

Figure 1: Motion sensor installed in one of the latrines.

waste collectors were notified via text message each morning which latrines to skip. The sensor unit was also equipped with an RFID reader that logged activity from the waste collectors. Waste collectors were instructed to swipe their "Collected" or "Not Able to Collect" tags depending on the action taken. The "Not Able to Collect" tag was reserved for instances when the facility had overflowed or required cleaning beyond the waste collector's responsibility, but there were no instances when the "Not Able to Collect" tag was used. The latrine operator was also given an RFID tag to request assistance, and RFID scans from latrine operators were immediately transmitted to Sweet Sense servers and triggered a Salesforce push notification for Sanergy staff to check-in with the latrine operator. Finally, sensor data were uploaded to the Sweet Sense dashboard to display the daily collection schedule, the log of Salesforce push notifications and waste collector scans, and the approximate number of uses for each latrine.

Table 1: Sample Characteristics

	sensor	no sensor	p-value
number of latrines	40	41	
number of observations	4870	4797	
collections per latrine: median (IQR)	141 (32)	133 (21)	0.331
solid waste container sizes	31 with 45L 9 with 40L	41 with 40 L	
high use latrines: number (%)	21 (52%)	11 (27%)	
low use latrines: number (%)	19 (47%)	30 (73%)	
solid waste fill level: median (IQR)	0.52 (0.23)	0.43 (0.24)	<0.001
liquid waste fill level: median (IQR)	0.41 (0.20)	0.34 (0.20)	<0.001

In order to measure changes in the efficiency of latrine servicing over the course of the intervention period, the average solid waste fill level and capacity savings were selected as the main outcome variables. Waste fill level as a percent was defined as follows:

$$\text{Fill Level} = \frac{\frac{Waste\,Weight}{Waste\,Density}}{Cartridge\,Capacity} \tag{1}$$

Waste weights were determined by weighing solid and liquid waste cartridges on-site at the time of servicing, and the cartridge weight was subtracted from the waste weight using an automated algorithm. While the density of the solid waste varied based on the amount of sawdust and toilet paper used, a conservative density of 0.721 kilograms per liter was used to convert solid waste weight to solid waste volume based on the average weight recorded for full cartridges (average density for human feces without consumables can vary from 1.06 to 1.09 g/ml, Penn et al., 2018). The solid waste volume was then divided by the cartridge capacity, which varied between 40 L and 45 L, to determine the latrine fill level (see Equation 1). Given that solid waste generally filled faster than liquid waste, the average solid waste fill level was selected as the primary outcome variable for measuring changes in servicing efficiency. Capacity savings were defined as the number of latrine servicing events that could be avoided due to dynamic scheduling.

2.1. Predictive Models

We initially assumed that estimates of latrine fill levels based on motion sensor data would be sufficient for predicting when latrines could be skipped. However, while we were able to predict waste fill levels with sufficient accuracy (mean absolute percent error of 20% and 22% for solid waste and liquid waste, respectively), we found that the motion sensor data on their own were not sufficient to predict when a latrince could be skipped while minimizing the risk of an overflow event. Figure 2 attempts to characterize the complex chain of factors that make latrine servicing predictions difficult. First, waste weights did not always accurately reflect waste volumes because of the variable amount of consumables that were used each day (i.e., the amount of sawdust and toilet paper present in the solid waste cartridge) and the different cartridge volumes in each latrine. Second, the need to be serviced depended not only on the estimated fill level from the first day's latrine activity, but also on the anticipated waste that would be added the next day if the latrine were skipped. Also, conversations with latrine operators revealed that full cartridge capacity was not always desirable due to increased odor and complaints from customers. Finally, even when it was determined that a latrine needed to be serviced, there was no guarantee that the waste collector would service the latrine. Sometimes waste collectors were not able to access latrines, and sometimes waste collectors used their own judgment based on a visual inspection of the fill level and their experience with the route to determine whether the latrine needed servicing. Waste collectors also indicated that they were more likely to service some latrines based on the preferences of the operator, often creating a tension between Sanergy's desire for more efficient servicing and the operators' desires for more frequent servicing. Within the Sanergy business model, waste collectors were directly contracted by Sanergy while latrine operators were franchisees, creating a tiered management structure that often complicated incentives and intervention implementation.

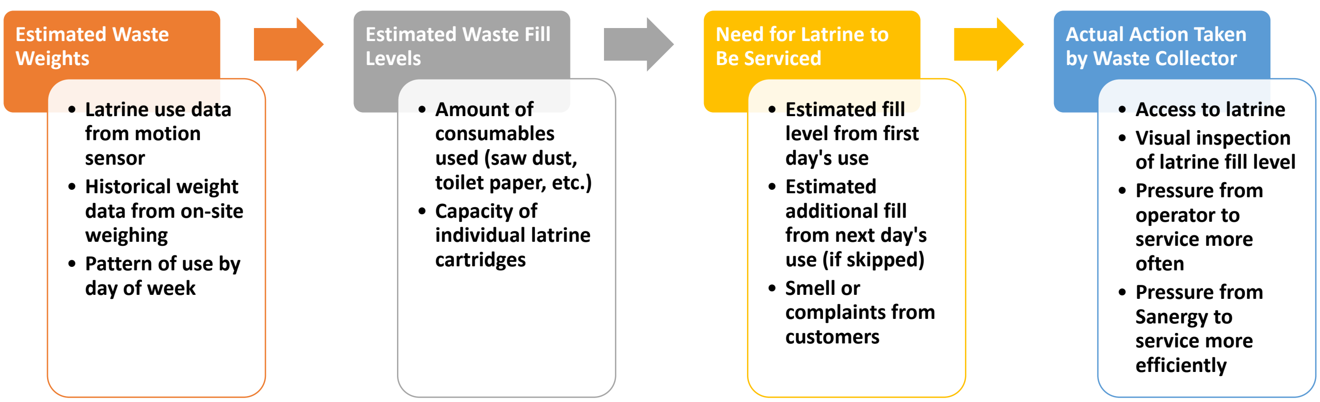

Figure 2: Chain of factors contributing to a latrine's need to be serviced.

Given the complex relationship between latrine use and servicing demand, we established that a simple linear correlation between motion sensor data and estimated fill levels would be insufficient for accurately predicting the need for servicing. Instead we used a machine learning algorithm (Super Learner, Polley et al., 2016) to predict when latrines would need to be serviced based on a variety of features that were identified using the available data (see Figure 3). We developed four models to compare the accuracy and cost-effectiveness of different scheduling scenarios. The first model represented Sanergy's business-as-usual static schedule, and the three simulated models represented the performance of dynamic scheduling using different data sources. In addition, we present in Table 2 the results from the actual dynamic schedule that was used during the intervention period and an additional simulated scenario that applies dynamic scheduling to lower-use latrines.

For the first model (Static Schedule) we used Sanergy's existing servicing schedule where thirty-six latrines were serviced daily and four latrines had reduced servicing schedules (i.e., four latrines were only serviced on Sundays, Mondays, Wednesdays, and Fridays based on waste collector recommendations). A dichotomous outcome variable was created to model whether a latrine would have overflowed had it been skipped based on weight data from consecutive days (i.e., if the estimated volumes from two consecutive days exceeded the cartridge capacity, then the outcome variable was classified as one; otherwise it was classified as zero).

In the second model (Sensor Only), we used sensor data and the Super Learner algorithm to predict when latrine servicing could be skipped. The predictor variables for this model included the latrine ID, the day of the week, and the normalized number of clicks from the motion sensor in the latrine. In addition, we used the number of clicks to create features that approximated the number of latrine uses and the number of edges associated with latrine use based on the methodology described in Clasen et al. (2012). This scenario was used to simulate the performance and cost-effectiveness of dynamic scheduling without the daily enumeration of weight data and servicing events.

For the third model (Weight Only), we used the record of daily solid and liquid waste measurements to predict when latrine servicing could be skipped. We first used Super Learner to predict the solid and liquid waste weights based on historical weight data (i.e., the latrine ID, the day of the week, and previous weight data collected from that latrine). Given the variability of latrine fill levels throughout the week, we created several features that improved the model's performance in predicting latrine waste weights, including: the average weight for each day of the week, the average weight for the previous seven days, the average weight for the previous three days, the weight from the previous day, and the first quartile, third quartile, median, and average overall weights for each latrine. The weight predictions from the first layer of the algorithm were then incorporated as a feature in the second layer of the algorithm that was used to predict the probability of an overflow event if skipped. This scenario was used to simulate the performance of dynamic scheduling with on-site weighing but without the capital and operating expenses associated with the sensors.

Finally, the fourth model (Sensor+Weight) combined sensor and weight data to predict waste weights and then used the full set of features to predict the need for servicing. To be explicit, in the first layer of the model all the features previously described (the latrine ID; the day of the week; the number of clicks; the estimated number of uses; the estimated number of edges; the average weight for each day of the week; the average weight for the previous seven days; the average weight for the previous three days; the weight from the previous day; the first quartile, third quartile, median, and average overall weights for each latrine; the number of RFID swipes; and the container size for solid and liquid wastes), were used to estimate the volume of solid and liquid waste in each latrine at the end of the day. This estimated waste volume was then combined with all the previously mentioned features to predict the probability of an overflow event if the latrine were skipped.

Predictions from the fourth model were used for dynamic scheduling during the implementation period, and we describe below the additional safeguards that were incorporated to prevent overflows. Finally, the relative importance of each of the features used in the three prediction models is shown in Figure 3.

2.2. Evaluation of Prediction Models

All models were evaluated using R (R Development Core Team, 2011), including the ROCR (Sing et al., 2009) and SuperLearner (Polley et al., 2016) packages. Super Learner is an ensemble learner that employs a variety of screen-

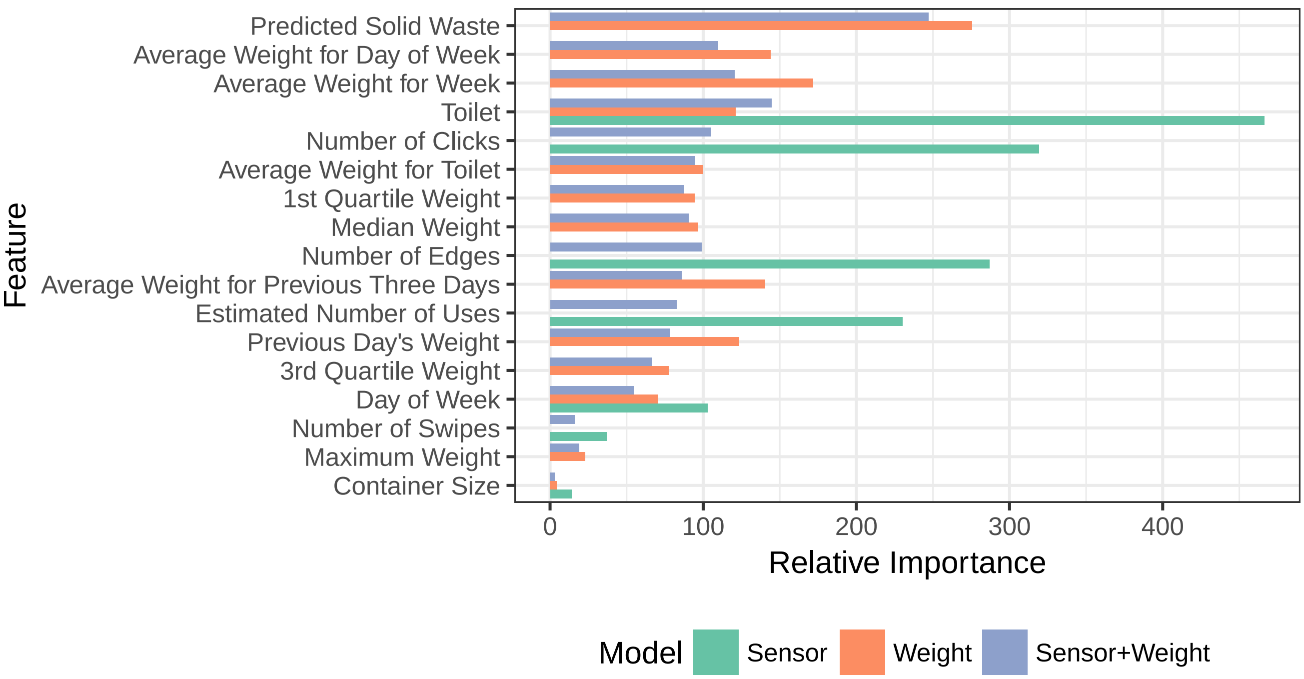

Figure 3: Relative importance of features used in the learner for predicting the probability of an overflow event for solid waste. The relative importance represented above is based on the mean decrease in Gini impurity from the randomForest learner. Gini impurity refers to the improvements in data classification that are contributed by each feature (Archer & Kimes, 2008).

ing and prediction algorithms to improve the accuracy of prediction (Polley & van der Laan, 2010). It has been used in recent studies to predict the failure of rural handpumps (Wilson et al., 2017) as well as to predict virological failure for HIV-positive patients on antiretroviral therapy (Petersen et al., 2015).

Several learners used to predict continuous and binomial outcomes were incorporated, including (ordered by weighting): Lasso regression (Tibshirani, 1996), multivariate adaptive regression splines (Hastie & Tibshirani, 1987; Milborrow, 2018), and random forests (Friedman, 2001). In order to evaluate the performance of each prediction model, the data were randomly split into training and testing sets based on each latrine site (70:30). To determine the relative weights associated with each learner's prediction in the ensemble, the algorithm performed ten-fold cross validation using the training data. The algorithm's predictive performance was then evaluated using the test data, where the mean absolute percent error (MAPE) was used to evaluate continuous outcomes and the area under the receiver operating characteristic (AUROC) curve, accuracy, sensitivity, and specificity were used to evaluate classification performance. The AUROC was selected as the primary metric for model comparison because it captures the overall accuracy of the model in predicting outcomes, regardless of the threshold chosen (see below), where an AUROC equal to one indicates perfect classification.

In order to make the performance of each model more tangible, we also present the predicted number of skips, the possible overflow events, the capacity savings, and the estimated costs and savings associated with each model in Table 2. The first band of results highlights the predictive performance of each model in classifying overflow events in the test data using only the training data (70% of randomly selected observations grouped by latrine). The second band of results presents the performance of the Actual Schedule during the implementation period and the simulated performances of each model for the same period. It is important to note that, while the simulated models were limited to the training data to evaluate classification performance (the first band of results), each model was trained on all available data when comparing performance during the implementation period (the second band of results). As a result, the simulated models had access to more data when generating the schedule for the implementation period compared to the Actual Schedule, which was retrained each evening using newly collected data.

For the purpose of this investigation the number of true negatives (i.e., instances when the algorithm accurately predicted that a latrine would not overflow if service were skipped) represented the potential for cost-savings due to higher efficiency latrine servicing. Given that the algorithm output a probability of overflow ranging from zero to one, a threshold was selected that would provide the lowest number of false negatives (i.e., instances when the algorithm incorrectly predicted that a latrine could be skipped) while minimizing the number of false positives (i.e., instances when the algorithm incorrectly predicted that a latrine had to be serviced). We were unable to quantify the overall cost of a false negative or latrine overflow event, as it involved tangible costs (e.g., latrine servicing crew, cleaning supplies, lost revenue due to latrine being closed, etc.) as well as intangible costs (e.g., damage to reputation of Sanergy brand or latrine operator, exposure to fecal contamination, etc.). As a result, we chose a final threshold of 0.22 for solid wastes and 0.10 for liquid wastes that allowed for the fewest number of potential overflow events, where potential overflow events were defined as latrine fill levels that were between 1.00 and 1.10 capacity.

2.3. Cost Assumptions

Servicing costs for each scenario were estimated based on cost and logistics data provided by Sanergy. Given that the primary expense for latrine servicing is labor, and given the small sample size for this experiment, costs were simplified to a per servicing event estimate. Cost-savings are represented as the amount of time and labor that could be avoided if dynamic scheduling were adopted at scale for latrines with similar use patterns. Capacity savings were defined as the number of skips divided by the total number of servicing days. Expenses related to waste collector labor were based on the assumption of each collector receiving a monthly salary of USD $225 and servicing approximately fifteen latrines per day. The expense of consumables was based on an average cost of USD $0.08 for disposable bags, sanitary bags, water, cleaning, and incineration per service event. All cost assumptions were estimated in consultation with Sanergy and based on expenses at the time of writing.

3. Results

Over the course of six months 4,870 service events were recorded for the forty latrines with sensors. When merged with the sensor data, a total of 4,371 weight and sensor observations were available for training and testing the learner. As seen in Figure 4 and Table 2, overall classification performance of the Static Schedule was low (AUROC of 0.52), whereas classification performance increased dramatically with the additional information provided by sensors (0.87), historical weight data (0.89), and combined sensor and weight data (0.90). Figure 5 displays the sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) that were evaluated on the testing data that was not used in model fitting. In addition, Table 2 displays the simulated performance of each model during the intervention period from January through March, 2017, including the predicted number of skips, the number of possible overflows, the capacity savings due to decreased latrine servicing, and the estimated savings per month based on reduced costs for labor and consumables. In total, there were 2,272 servicing events recorded during the three-month intervention period for the latrines with sensors. There were 566 opportunities for skipping servicing, and the performance of each of these models in predicting these potential skips varied considerably. Sanergy's static schedule reflected approximately 2% of the possible skips, whereas the dynamic schedules using sensor and weight data were able to predict between and 12% and 13% of the possible skips.

3.1. Comparison Group

Over six months 4,797 service events were recorded for the forty-one latrines without sensors that served as a comparison group. As shown in Table 1, the latrines with sensors had a higher median fill level compared to the latrines without sensors (52% vs. 43%). Given that the majority of the latrines with sensors were high-use latrines, where high-use was defined as having a maximum fill level and a third-quartile fill level greater than 60% of the cartridge capacity, there was less room for improving efficiency in the latrines with sensors compared to the comparison group. That is, the fact that latrines had a median fill level of 52% meant that there were fewer opportunities for skipping the latrines with sensors compared to the latrines without sensors. Despite there only being a 9% difference in median fill levels between the two groups there was significantly

Table 2: Performance metrics for the four prediction models, the actual implementation results, and a prediction model using low-use latrines. Two comparisons are made in the following table. In the first band of results each model is evaluated based on its performance on the hold-out data. In the second band of results each model uses all available data to simulate its performance during the three-month implementation period to give more concrete examples of how each model would have performed if used to inform latrine servicing.

Model Performance	Static Schedule	Sensor Only	Weight Only	Sensor+ Weight	Actual Schedule[a]	Low-Use Latrines[b]
Performance on Test Data From Baseline and Intervention Periods						
sensitivity	100%	96.4%	97.3%	97.9%	99.2%	95.4%
specificity	4.50%	53.7%	61.2%	61.9%	6.23%	63.1%
positive predictive value	49.2%	65.9%	69.9%	70.5%	55.5%	50.3%
negative predictive value	100%	94.2%	96.0%	97.0%	86.7%	97.2%
accuracy (AUROC)	52.2%	86.6%	89.2%	89.5%	52.7%	90.5%
Performance on All Data During Three-Month Intervention Period						
predicted skips	46[c]	279[c]	274[c]	298[c]	75[d]	1142[e]
possible overflow events	0	47	17	18	10[f]	69
capacity savings[g]	2.0%	12%	13%	13%	3.3%	52%
waste collector labor[h]	$1100	$1000	$1000	$990	$1100	$530
total consumables[i]	$150	$140	$140	$140	$150	$73
total cost per quarter	$1300	$1100	$1100	$1100	$1300	$600
savings per month[j]	NA	$44	$43	$48	$5	$200

[a] Performance for Actual Schedule is based on the dynamic schedule from the implementation period.

[b] Performance of the weight only model on lower use latrines in the comparison group.

[c] Out of 566 possible skips.

[d] Represents the actual number of skips during the intervention period.

[e] Out of 1383 possible skips.

[f] Instances when a latrine was scheduled for a skip but waste collectors serviced the latrine based on visual inspection of fill-level; there were no reported overflow events during the baseline or intervention periods.

[g] Number skips divided by the total number of servicing days.

[h] USD per quarter based on Sanergy records, with the average waste collector servicing 15 latrines per day and receiving a monthly salary of USD $225.

[i] USD per quarter based on USD $0.08 for disposable bags, sanitary bags, water, cleaning, and incineration per service event.

[j] Saving compared to the static schedule.

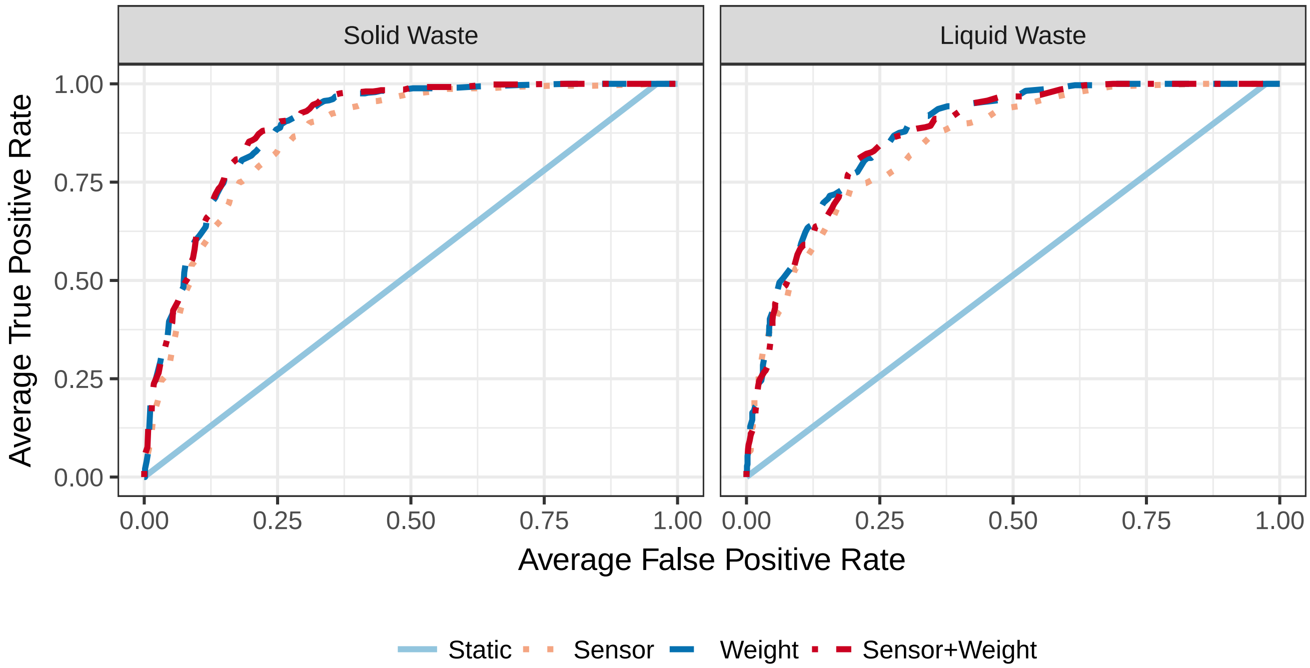

Figure 4: Area under the receiver operating characteristic (AUROC) curve for solid (left) and liquid (right) waste overflow predictions.

more opportunity for skipping in the comparison group. Using only weight data from the control group, the Super Learner algorithm was able to predict 1,142 skip events with a high degree of accuracy (AUROC of 0.91) and an estimated capacity savings of 52%. Given that we were not able to test dynamic scheduling in the comparison group, these simulated results represent the upper bound of potential capacity savings. As seen in Figure 6, average fill levels for latrines in both groups increased over the intervention period, which may reflect seasonal trends or general uplift due to Sanergy's efforts to improve servicing efficiency over the same period. Average solid waste fill levels increased from 49.8% to 55.0% for sensored latrines and from 43.0% to 44.6% for non-sensored latrines between the baseline and intervention periods. Similarly, average liquid waste fill levels increased from 40.7% to 43.9% for sensored latrines and from 36.1%

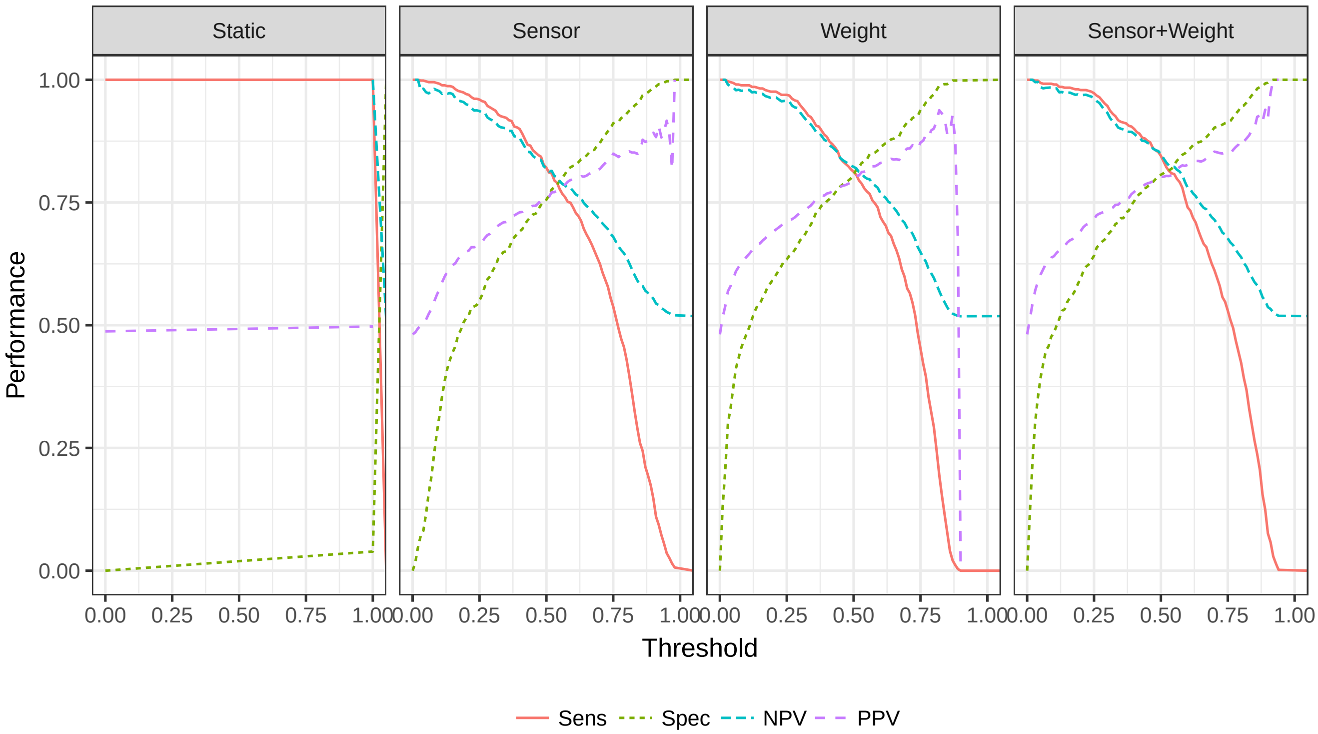

Figure 5: Sensitivity (Sens), specificity (Spec), negative predictive value (NPV), and positive predictive value (PPV) for solid waste overflow predictions over a range of probability thresholds.

to 38.6% for non-sensored latrines over the same periods.

4. Discussion

Using weight and sensor data from forty latrines in an informal settlement in Nairobi, we were able to demonstrate that a machine learning algorithm can predict with a high degree of accuracy when latrine servicing could be skipped (AUROC from 0.87 to 0.90 and capacity savings from 12% to 13%). These predictions were then used to create a dynamic latrine schedule that modestly increased solid waste collection efficiency between the baseline and intervention periods (see Figure 6). Although the machine learning algorithm was more ef-

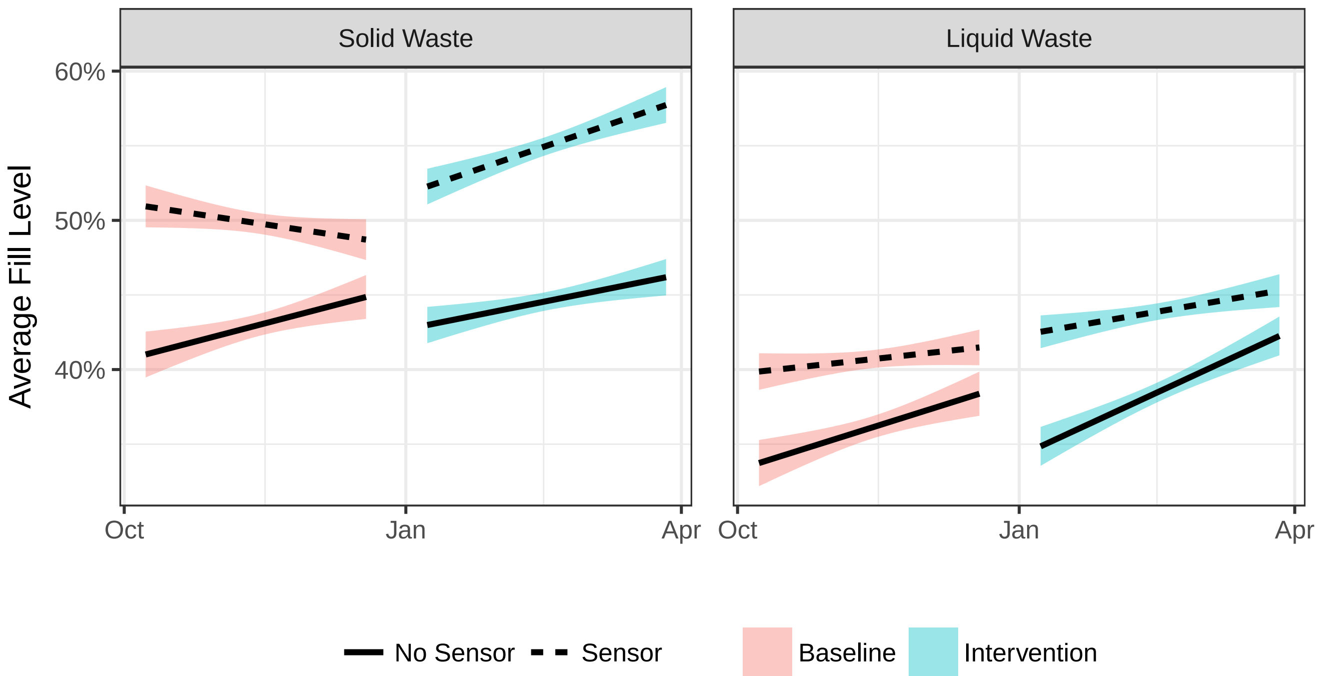

Figure 6: Average fill levels for the latrines with sensors (dashed line) and the latrines without sensors (solid line) for the baseline (pink) and intervention (blue) periods. The shaded regions represent the 90% confidence interval.

fective in identifying skip events compared to the Static Schedule (AUROC 0.52 and capacity savings of 2%), there was a significant gap between the simulated performance of the algorithm and the implemented results (AUROC 0.53 and capacity savings of 3%). It is important to note that the Sensor, Weight, and Sensor+Weight models were trained on more data than the Actual Schedule because the Actual Schedule was generated by retraining the model every day with the new data that was collected during the implementation period. In contrast, the Sensor, Weight, and Sensor+Weight models were trained on a random selection of 70% of the data (i.e., the training data) to evaluate their predictive performance on the test data (the 30% hold-out data). To simulate their scheduling performance during the implementation period, those three models were trained on all the data. However, we attribute most of the gap between simulated and actual performance to implementation challenges.

Implementation challenges were numerous. First, dynamic scheduling represented a significant deviation from the static schedules that waste collectors and field staff were accustomed to. Second, collecting accurate weight data was difficult given the relative inaccessibility of the latrines within the informal settlement and the challenge of weighing and recording waste weights while servicing latrines. In addition, waste collectors were accustomed to weighing waste cartridges at a central weighing station, a practice that was prone to error and mislabelled data. In order to facilitate more accurate weight measurements, a set of two on-site weighing machines were fabricated to enable waste collectors and enumerators to measure and record waste weights at the time of servicing. Even with this new system data entry was still subject to human error (e.g., inaccurate designations of latrines, entry error, or delayed uploading of records to the server). In addition, there were initially no records that were logged for latrines that were skipped, so it was impossible to distinguish between latrines that were skipped and data that were missing. This was corrected by creating a new mobile survey for waste records and an automated algorithm to check that events were logged for each latrine. However, even with these redundancy measures about 5% of expected entries were not accounted for each day. The majority of the missing data were from lower-use latrines in the comparison group, typically when a latrine was scheduled for servicing but no weight entry was recorded. This dynamic occurred more frequently with the low-use latrines in the comparison group because latrines with missing entries were automatically scheduled for servicing the next day as a fail-safe measure to prevent overflow. However, since some latrines were much lower use in the comparison group, waste collectors were more likely to skip those latrines multiple days regardless of the dynamic schedule's prescribed action for the day. The ability to generate dynamic schedules with multiple consecutive skip days was not explored in this investigation.

Because the dynamic schedule was new and required the approval and cooperation of latrine operators, the algorithm was initially tuned conservatively in order to minimize the risk of an overflow event. For example, even though solid wastes were the primary driver of service events, a probability of overflow for either solid or liquid wastes automatically designated a latrine for collection. In addition, if a latrine was skipped or there was a missed entry from the previous day, the latrine was automatically scheduled for collection. However, we eventually realized that waste collectors often skipped low-use latrines regardless of scheduling. Since missing data entries automatically designated a latrine for collection, lower-use latrines were often scheduled for collection even when waste collectors knew that they could be skipped. This combination of missing data and conservative scheduling resulted in a general distrust in the algorithm's predictions, prompting many waste collectors to service latrines according to their own intuition rather than the dynamic schedule.

However, it is important to note that the waste collector's intuition was correct more often than not. On at least ten occasions, the algorithm scheduled a latrine for skipping that clearly would have overflowed had the waste collector not serviced the latrine based on visual inspection. In this regard, the route selected for installing sensors was a safe choice because the waste collectors were reliable and the route was well-known and accessible by Sanergy staff. However, these very attributes also made the route less useful for the experiment, as the information being provided by the sensors and daily weights was unnecessary given the familiarity of the waste collectors and the daily servicing needed by most latrines. As a result, it was determined that collecting data from sensors or daily weights would be most useful on new routes where latrine patterns were still being established, on existing routes where latrine use was more variable, or on routes where latrines were used less frequently.

Although the accuracy of the algorithm may not be much better than that of a seasoned waste collector, there is an additional advantage that motion sensor data, weight data, or RFID scans can provide: the ability to track latrine servicing. Sanergy's capacity for reallocating waste collector labor depends on its ability to predict when latrines will need to be serviced while reliably tracking when latrines have been serviced. In this way service records provide a form of accountability for waste collectors, a quality assurance mechanism for honoring contracts with latrine operators, and a dataset for predicting future servicing. However, the high cost of hardware relative to the low cost of labor in Nairobi implies that cost savings would need to significantly increase for Sanergy to implement any changes at scale. Our simulations suggest that sensor and weight measurements could save between $43 and $200 per month for a route with approximately forty latrines depending on the frequency of use of the latrines. This cost savings represents the upper bound on all expenses related to latrine sensors (e.g., hardware, data transmission, operation and maintenance personnel, predictive analytics), weight records (e.g., enumerators, mobile devices, and predictive analytics), or RFID scanners. However, given the gap between simulation and implementation, these estimates may be optimistic.

There are additional considerations that may temper the cost savings associated with dynamic scheduling. First, 92% of the latrines with sensors and 54% of the latrines without sensors were co-located, meaning that latrines were being managed by the same operator in clusters of two or three. Co-located latrines were more likely to be skipped compared to standalone latrines, but the benefit of skipping a latrine is greatly diminished if waste collectors are already servicing a latrine in the same location. Second, this analysis was not able to quantify the potential cost associated with an overflow event. This cost would include additional labor and supplies for servicing an unsanitary latrine, but it would also include damage to the operator or Sanergy's reputation and reduced patronage. In addition, the current algorithm uses the latrine ID as a predictor variable to capture site-level variability and latrine-use trends. However, using the latrine ID as a predictor also makes the algorithm less portable given the need to collect baseline data from new latrines before making predictions on a new route. However, this baseline burn-in may be inevitable given that average weight trends were also significant predictors in the algorithm. Finally, this analysis was not able to take into consideration the additional administrative cost associated with reallocating waste collectors in a dynamic scheduling scenario. Given the geospatial distribution of latrines, the inability to remotely chart pathways through informal settlements, and challenges finding and accessing latrines for waste collection, it would be exceedingly difficult to dynamically redraw servicing routes for waste collectors on a regular basis.

In this study, sensors were able to monitor latrine activity, track latrine servicing, and facilitate communication between Sanergy staff and latrine operators. While RFID tags provided an important accountability mechanism for tracking servicing and motion sensor data provided rough estimates of latrine use, we found that motion sensor data did not significantly improve the algorithm's ability to generate a dynamic service schedule compared to weight data alone. With or without sensors, the high accuracy of predictions observed in this study could provide a promising application of machine learning for estimating waste weights and dynamically scheduling latrine servicing. Although we found that implementation lagged simulation significantly, we anticipate a much greater potential for servicing efficiency and cost savings when applied to lower use latrines.

The authors declare the following interests: Authors TS, CN, and ET were compensated employees of SweetSense Inc, the instrumentation provider, during the course of this study. Author LS was a compensated employee of Sanergy Inc. during the course of this study.

Acknowledgements

We appreciate our partnership with Sanergy Inc., and in particular the enumerators, waste collectors, field staff, and administrators that made this study possible. We also want to thank Jeremy Coyle for reviewing the code used in this analysis. This study was supported by the United Kingdom Department for International Development through the GSM Association, the Link Foundation, and the National Science Foundation IGERT Grant #0966376: "Sustaining Ecosystem Services to Support Rapidly Urbanizing Areas." Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

African Population and Health Research Center (2014). *Population and Health Dynamics in Nairobi's Informal Settlements: Report of the Nairobi Cross-Sectional Slums Survey (NCSS) 2012*. Technical Report April.

Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics and Data Analysis*, *52*, 2249–2260. doi:10.1016/j.csda.2007.08.015.

Auerbach, D. (2016). Sustainable Sanitation Provision in Urban Slums – The Sanergy Case Study. In E. A. Thomas (Ed.), *Broken Pumps and Promises: Incentivizing Impact in Environmental Health* chapter 14. (pp. 211–216). Springer.

Bhagwan, J. N., Still, D., Buckley, C., & Foxon, K. (2008). Challenges with up-scaling dry sanitation technologies. *Water science and technology: a journal of the International Association on Water Pollution Research*, *58*.

Bohnert, K., Chard, A. N., Mwaki, A., Kirby, A. E., Muga, R., Nagel, C. L., Thomas, E. A., & Freeman, M. C. (2016). Comparing Sanitation Delivery Modalities in Urban Informal Settlement Schools: A Randomized Trial in Nairobi , Kenya. *International Journal of Environmental Research and Public Health*, *13*, 1–14. doi:10.3390/ijerph13121189.

Brown, J., Cumming, O., Bartram, J., Cairncross, S., Ensink, J., Holcomb, D., Knee, J., Kolsky, P., Liang, K., Liang, S., Nala, R., Norman, G., Rheingans, R., Stewart, J., Zavale, O., Zuin, V., & Schmidt, W.-P. (2015). A controlled, before-and-after trial of an urban sanitation intervention to reduce enteric infections in children: research protocol for the Maputo Sanitation (MapSan) study, Mozambique. *BMJ open*, *5*, e008215. URL:

http://www.scopus.com/inward/record.url?eid=2-s2.0-84937242906&partnerID=tZOtx3y1. doi:10.1136/bmjopen-2015-008215.

Clasen, T., Fabini, D., Boisson, S., Taneja, J., Song, J., Aichinger, E., Bui, A., Dadashi, S., Schmidt, W. P., Burt, Z., & Nelson, K. L. (2012). Making sanitation count: Developing and testing a device for assessing latrine use in low-income settings. *Environmental Science and Technology*, *46*, 3295–3303. doi:10.1021/es2036702.

Coffey, D., Gupta, A., Spears, D., Khurana, N., Srivastav, N., Hathi, P., & Vyas, S. (2014). Revealed Preference for Open Defecation. *Economic and Political Weekly*, *49*, 43–55. URL: http://www.epw.in/special-articles/revealed-preference-open-defecation.html.

Delea, M. G., Nagel, C. L., Thomas, E. A., Halder, A. K., Amin, N., Shoab, A. K., Freeman, M. C., Unicomb, L., & Clasen, T. F. (2017). Comparison of respondent-reported and sensor-recorded latrine utilization measures in rural Bangladesh: a cross-sectional study. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, (pp. 1–8). URL: http://academic.oup.com/trstmh/article/doi/10.1093/trstmh/trx058/4590286. doi:10.1093/trstmh/trx058.

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*, 1189–1232. doi:10.1214/009053606000000795.

Hastie, T., & Tibshirani, R. (1987). Generalized Additive Models : Some Applications Generalized Additive Models : Some Applications. *Journal of the American Statistical Association*, *82*, 371–386.

Kimani-Murage, E. W., Fotso, J. C., Egondi, T., Abuya, B., Elungata, P., Ziraba, A. K., Kabiru, C. W., & Madise, N. (2014). Trends in childhood mortality in Kenya: The urban advantage has seemingly been wiped out. *Health and Place*, *29*, 95–103.

URL: http://dx.doi.org/10.1016/j.healthplace.2014.06.003. doi:10.1016/j.healthplace.2014.06.003.

Mara, D. (2012). Sanitation: What's the Real Problem? *IDS Bulletin*, *43*, 86–92. doi:10.1111/j.1759-5436.2012.00311.x.

Milborrow, S. (2018). Multivariate Adaptive Regression Splines. URL: https://cran.r-project.org/web/packages/earth/earth.pdf. doi:10.1214/aos/1176347963.

O'Reilly, K., Louis, E., Thomas, E., & Sinha, A. (2015). Combining sensor monitoring and ethnography to evaluate household latrine usage in rural India. *Journal of Water, Sanitation and Hygiene for Development*, *5*, 426–438. URL: http://washdev.iwaponline.com/cgi/doi/10.2166/washdev.2015.155. doi:10.2166/washdev.2015.155.

Paterson, C., Mara, D., & Curtis, T. (2007). Pro-poor sanitation technologies. *Geoforum*, *38*, 901–907. doi:10.1016/j.geoforum.2006.08.006.

Penn, R., Ward, B. J., Strande, L., & Maurer, M. (2018). Review of synthetic human faeces and faecal sludge for sanitation and wastewater research. *Water Research*, *132*, 222–240. URL: https://doi.org/10.1016/j.watres.2017.12.063. doi:10.1016/j.watres.2017.12.063.

Petersen, M. L., LeDell, E., Schwab, J., Sarovar, V., Gross, R., Reynolds, N., Haberer, J. E., Goggin, K., Golin, C., Arnsten, J., Rosen, M. I., Remien, R. H., Etoori, D., Wilson, I. B., Simoni, J. M., Erlen, J. A., van der Laan, M. J., Liu, H., & Bangsberg, D. R. (2015). Super Learner Analysis of Electronic Adherence Data Improves Viral Prediction and May Provide Strategies for Selective HIV RNA Monitoring. *Journal of acquired immune deficiency syndromes*, *69*, 109–18. URL: http://www.ncbi.nlm.nih.gov/pubmed/25942462. doi:10.1097/QAI.0000000000000548. arXiv:15334406.

Polley, E., LeDell, E., & van der Laan, M. (2016). Package 'SuperLearner': Super Learner Prediction. URL: https://cran.r-project.org/web/packages/SuperLearner/SuperLearner.pdf.

Polley, E. C., & van der Laan, M. J. (2010). Super learner in prediction. *U.C. Berkeley Division of Biostatistics Working Paper Series*, (pp. 1--19). URL: http://biostats.bepress.com/ucbbiostat/paper266/.

R Development Core Team (2011). R: a language and environment for statistical computing. URL: http://www.r-project.org/.

Robiarto, A., Sofyan, E., Setiawan, D., Malina, A., & Rand, E. C. (2014). *Scaling Up Indonesia's Rural Sanitation Mobile Monitoring System Nationally*. Technical Report December. URL: http://www.wsp.org/sites/wsp.org/files/publications/WSP-Indonesia-Mobile-Monitoring.pdf.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2009). ROCR: Visualizing the performance of scoring classifiers. URL: https://cran.r-project.org/web/packages/ROCR/ROCR.pdf.

Sinha, A., Nagel, C. L., Thomas, E., Schmidt, W. P., Torondel, B., Boisson, S., & Clasen, T. F. (2016). Assessing Latrine Use in Rural India: A Cross-Sectional Study Comparing Reported Use and Passive Latrine Use Monitors. *American Journal of Tropical Medicine and Hygiene*, . URL: http://www.ajtmh.org/cgi/doi/10.4269/ajtmh.16-0102. doi:10.4269/ajtmh.16-0102.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267--288.

Trémolet, S. (2011). *Scaling Up Rural Sanitation: Identifying the Potential for Results-Based Financing for Sanitation*. Technical Report November.

UNICEF / WHO (2017). *Progress on Drinking Water, Sanitation and Hygiene: 2017 update and SDG baselines*. Geneva. URL: http://apps.who.int/iris/bitstream/10665/258617/1/9789241512893-eng.pdf%0Ahttp://www.wipo. doi:10.1111/tmi.12329.

United Nations (2015). *The Millennium Development Goals Report*. Technical Report. arXiv:RePEc:adb:adbmdg:399.

United Nations Department of Economic and Social Affairs (2015). *World Urbanization Prospects: The 2014 Revision*. Technical Report. URL: http://www.demographic-research.org/volumes/vol12/9/. arXiv:arXiv:1011.1669v3.

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*, 1--13. doi:10.1371/journal.pone.0188808.