SmartSim: A Device-Accurate Smart Home Simulator for Energy Analytics

Dong Chen, David Irwin, and Prashant Shenoy
University of Massachusetts Amherst

Abstract—**Utilities have deployed tens of millions of smart meters, which record and transmit home energy usage at fine-grained intervals. These deployments are motivating researchers to develop new energy analytics that mine smart meter data to learn insights into home energy usage and behavior. Unfortunately, a significant barrier to evaluating energy analytics is the overhead of instrumenting homes to collect aggregate energy usage data and data from each device. As a result, researchers typically evaluate their analytics on only a small number of homes, and cannot rigorously vary a home's characteristics to determine what attributes of its energy usage affect accuracy. To address the problem, we develop SmartSim, a publicly-available device-accurate smart home energy trace generator. SmartSim generates energy usage traces for devices by combining a *device energy model*, which captures its pattern of energy usage when active, with a *device usage model*, which specifies its frequency, duration, and time of activity. SmartSim then generates aggregate energy data for a simulated home by combining the data from each device. We integrate SmartSim with NILM-TK, a publicly-available toolkit for Non-Intrusive Load Monitoring (NILM), and compare its synthetically generated traces with traces from a real home to show they yield similar quantitative and qualitative results for representative energy analytics.**

I. Introduction

Utilities are deploying "smart" electric meters, which record and transmit average home energy usage at fine-grained intervals, in large numbers, as part of a broad effort to transition to a smart electric grid. As of 2011, utilities had already installed nearly 40 million smart meters [1]. Researchers and companies are increasingly interested in mining smart meter data to develop new energy data analytics to learn insights into home energy usage and behavioral patterns. Such insights can be used to provide i) users more visibility into their energy usage at a low cost, e.g., without deploying any sensors or even requiring access to the home, and ii) third-party companies with private information they can use for marketing.

Researchers are actively developing such energy data analytics, e.g., [2], [3], [4], [5], and a number of well-established and startup companies, such as Bidgely, PlottWatt, OPower, Belkin, etc., have emerged to commercialize analytic techniques. One prominent example of energy data analytics and the focus of many of the companies above is Non-Intrusive Load Monitoring (NILM) or disaggregation, which analyzes smart meter data to infer (or disaggregate) the energy usage of individual electric devices. Researchers are also actively developing new ways to *prevent* such analytics to preserve consumer privacy [6], [7] and prevent third-parties from exploiting consumer data for profit. Unfortunately, there are a variety of logistical problems that make rigorously evaluating energy data analytics, such as NILM, or ways to prevent them challenging. Most importantly, there are no *complete* sub-metered datasets of home energy consumption that are publicly-available. By "complete," we mean energy usage data over an extended period, e.g., months to years, at a per-second resolution for an entire home *and all of its devices.* Of course, a complete sub-metered dataset that includes accurate ground truth data from each device is necessary for properly evaluating the accuracy of any energy data analytics technique, or any technique for preventing it.

While some public datasets exist, the most complete sub-metered datasets that include per-second resolution [5], [8], of which there are few, typically include only energy data from each circuit in a home, even though many devices may connect to a single circuit. Further, these public datasets with second-level data generally include only a few homes, e.g., 3-6, over a short period of time due to the expense of deploying such instrumentation. While existing utility smart meters record average power data every few minutes at most, we focus on second-level granularity in this paper for many reasons. First, second-level resolution is the finest granularity that commodity in-panel and plug-level energy sensors, such as the TED, eGauge, and WeMo, use, and there are indications that the next generation of utility smart meters will operate at second-level resolution [9]. Next, second-level resolution is capable of yielding highly accurate results that compromise privacy, as the probability of two devices starting at the same second is low [8]. Finally, second-level resolution is the granularity that many current researchers are targeting [3].

Given the lack of complete public datasets at per-second resolution, researchers must either deploy their own sensor systems in selected homes to gather data for evaluation, limit their evaluation to a few large appliances (connected to dedicated circuits), or use coarser-data, e.g., minute-level or worse [10], where energy data analytics are not as effective. Each of these options is undesirable. The former requires a significant investment in time and money to gather data, while the latter two severely restrict the range of analytics suitable for evaluation. Many NILM algorithms and energy analytics techniques already exist that focus on identifying a small number of the largest home appliances, which are generally the easiest to detect. We expect future NILM and energy analytics research to focus on other aspects of energy usage, such as detecting smaller appliances, tracking device energy usage in real-time [4], or using energy usage to infer home occupancy [11], [12]. However, even if a complete large-scale dataset were available, it would restrict researchers to evaluating analytics on the homes within the dataset, which may not be representative, as homes vary widely by region and climate. For example, while the Pecan Street dataset provides per-minute resolution data for a large number of homes in Texas [13], the energy usage of these homes likely differs from those in the Northeast U.S., Europe, etc.

Ultimately, no single dataset, no matter how large, can enable researchers to rigorously evaluate the data characteristics that affect NILM accuracy, or other similar types of energy data analytics. Prior work [3] has evaluated popular NILM algorithms against a variety of public datasets and found widely disparate results across multiple accuracy metrics. The disparate results are due to the energy usage characteristics of the homes in each dataset. To date, there has been little work on evaluating what characteristics of a home's energy usage (or its set of devices) affect NILM accuracy. To perform such an evaluation, researchers must be able to vary the characteristics of home energy use, e.g., the type and number of devices, usage patterns, etc., in a controlled fashion and evaluate the accuracy. Such an evaluation would be highly useful given the plethora of NILM algorithms now available, enabling users to select the best one for home energy usage characteristics. Such an analysis might lead to improvements in existing algorithms or the development of new algorithms by identifying weak points in existing algorithms.

In this paper, we address the problem by developing SmartSim, a device-accurate smart home energy trace generator. SmartSim enables researchers to generate complete datasets for homes that include second-level energy data for the entire simulated home and each of its simulated devices. SmartSim generates second-level energy usage traces for devices by combining a device energy model, which specifies the device's energy usage pattern when on, with a device usage model, which specifies the device's frequency, duration, and time of operation. SmartSim leverages an empirical modeling methodology from prior work [14] to generate device energy models, and can generate custom device usage patterns based on well-known distributions or derive usage patterns from real datasets. In the latter case, we observe device usage patterns can be inferred using coarse energy data, e.g., every minute or five minutes without requiring access to high resolution second-level data. For the device energy model, SmartSim relies on a library of device models we created based on real trace data. To the best of our knowledge, SmartSim is the first device-accurate energy data simulator intended to support the development and study of new energy data analytics.

We are publicly releasing SmartSim as a tool for researchers.[1] SmartSim integrates with the recently released NILM Toolkit (NILM-TK) [3] in that it generates trace files in NILM-TK format, which is based on the HDF5 binary file format. NILM-TK includes canonical implementations of well-known NILM algorithms, including Hart's original combinatorial optimization (CO) algorithm [15] and one based on applications of Factorial Hidden Markov Models (FHMM) [5], [16]. NILM-TK also includes support for the wide variety of accuracy metrics used when evaluating NILM algorithms.

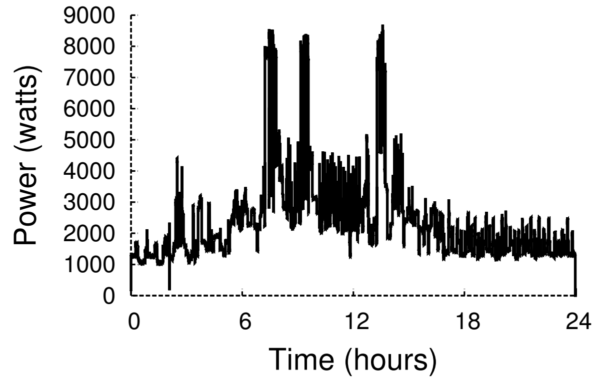


Fig. 1. Modern homes operate complex loads, resulting in a highly stochastic power profile that complicates energy data analytics.

In this paper, we describe the SmartSim tool, its design methodology, its basic data pipeline, as well as different potential use cases and areas for improvement and further work. We then use the tool to generate energy traces for a simulated home, leverage NILM-TK to analyze the data using the CO and FHMM NILM algorithms, and compare the results to the results of the algorithms for both the REDD dataset and the Smart* dataset, which are the two publicly-available datasets with second-level resolution that are closest to being complete. Our results indicate that SmartSim yields quantitatively similar results for these NILM algorithms as when using real datasets, while producing energy traces that also are qualitatively similar.

II. Background and Motivation

The number, diversity, and complexity of electrical devices in homes continues to increase. A typical home may now have as many as 100 loads. Many of these loads may be low power, e.g., alarm clocks, nightlights, phone chargers, etc., but they contribute to noise that complicates energy data analytics. In addition, devices are increasingly diverse: many homes now operate a variety of networked electronic devices in addition to devices with simple motors, e.g., refrigerators, and heating elements, e.g., ovens, which are also generally compute-driven. These devices are complex in that their power usage is programmatically regulated by internal computers. As a result, these devices often do not exhibit clear edges due to distinct power state transitions, but instead vary their power continuously or operate numerous distinct internal loads.

To illustrate, Figure 1 shows the complexity of a home's per-second power usage over one day (a), as well as the power usage and real-reactive scatterplot. The data is from Home A of the public Smart* dataset [8]. Many energy data analytics, including many previously proposed NILM algorithms, do not account for the complexity of power usage in modern homes. The techniques generally assume the presence of only a few simple devices [17]. The primary reason for this is the difficulty in collecting both aggregate power data from a

[1] See https://github.com/sustainablecomputinglab/smartsim/

home, as well as data from each of its individual loads. Along with the deployment of smart meters, there has been an effort to publicly release datasets. While useful, the availability of public datasets only enables researchers to compare the performance of different algorithms on common datasets. However, public datasets *do not* enable researchers to rigorously evaluate the energy data characteristics that affect accuracy.

There exist a wide variety of analytics techniques proposed and evaluated on various datasets. For example, recent work identifies 18 different NILM algorithms from prior work [2]. However, researchers have generally not evaluated these techniques to understand what characteristics of energy data and devices affect their performance. Recent work comparing different NILM algorithms on many public datasets indicates widely different performance among algorithms [3]. *The only way to improve upon these existing algorithms (and other energy data analytics) is to better understand the characteristics of energy data that affect their performance, and then focus on optimizing for those particular characteristics.* Unfortunately, public datasets, while useful, do not provide the means to alter a home's energy data characteristics in a controlled fashion to understand the impact on accuracy, or to understand performance on homes that have different characteristics from those in the dataset.

III. SmartSim Design

SmartSim generates per-second traces of average power usage for simulated homes that are *device accurate* in that the simulated home trace is the sum of individual devices in the home. SmartSim's goal is to enable researchers to control the energy data characteristics of simulated homes to evaluate what aspects of a home's energy use affect the performance and accuracy of energy data analytic techniques. Thus, SmartSim relieves analytics researchers from the task of instrumenting multiple homes to gather energy data.SmartSim's design includes multiple phases that users configure to generate a simulated home trace. Figure 2 shows SmartSim's pipeline.

First, the user must choose the set of devices in the home. Users may choose devices from a library of device models that comes with SmartSim, or may provide their own device models. SmartSim models devices based on a well-defined modeling methodology proposed in prior work [14]. Thus, users may add new devices based on data they have collected themselves, e.g., by applying this methodology to construct a SmartSim-compliant model of a device. Users may also choose to alter the parameters of existing device models in the library, either to better model a new device or to evaluate how the change affects the performance of a particular energy analytic technique. We discuss device energy models further below. In addition to choosing specific devices, users may also select models for *background noise*, which encapsulate the aggregate power use of the large number of low-power loads that are often not metered or modeled in public datasets, but add to the complexity of real home energy data and significantly affect the performance of energy data analytics. Users may either model background noise using a well-known distribution, e.g., gaussian, or leverage empirical data to derive a model of background noise. Explicitly simulating background noise is important in determining the resilience of NILM algorithms and other energy data analytics to unmetered devices that lack training data, which is common in real deployments.

Since selecting a large number of devices for a simulated home may be cumbersome, SmartSim also provides template homes that are already pre-populated with a set of devices, such as a canonical apartment with a few devices and low background noise or a four bedroom home with many devices and high level of background noise. After selecting the set of devices in the home, the user next selects the usage pattern for each device. As we discuss below, the usage pattern dictates a specified distribution for duration, frequency, and timing of device use that SmartSim draws from when determining a device's operation. Users may select standard well-known distributions, e.g., gaussian, normal, etc., for each dimension of use, or empirically derive usage pattern distributions from known data. As above, SmartSim includes a library of device usage pattern distributions for different time intervals, e.g., day-long, week-long, etc.

Finally, after determining the set of devices, their device energy models and usage pattern distributions, and the background noise (contributed by unmetered devices), SmartSim generates random device-accurate traces that conform to the models and distributions.

A. Device Energy Modeling

We leverage an empirical modeling methodology from prior work to generate device energy models [14]. The methodology classifies the energy usage pattern of different devices based on how they consume AC power. AC devices are either resistive, inductive, capacitive, or non-linear based on how they their current waveform aligns with the phase of the sinusoidal AC voltage waveform. Resistive, inductive, and non-linear loads exhibit a small set of common characteristics that can be used to model their fine-grained energy usage over time [14]. These characteristics are complete in that the power usage of every load is composed of them, as described below.

- **On-off Models** are simple devices that exhibit one or more fixed power states, and largely include low-power resistive devices, such as incandescent lighting.
- **On-off Growth/Decay Models** exhibit a steep rise in power when activated and a smooth growth or decay from the initial peak to a stable state. Growth and decay are logarithmic or exponential functions, respectively, where the growth/decay parameter is device-specific. High-power resistive loads and inductive loads, including AC motors and compressors, follow an on-off growth/decay model.
- **Stable Min-Max Models** exhibit a stable maximum or minimum state with frequent large deviations, e.g., every few seconds, from the stable state. The magnitude and frequency of the deviations are random variables with distributions specific to each device. Non-linear loads that

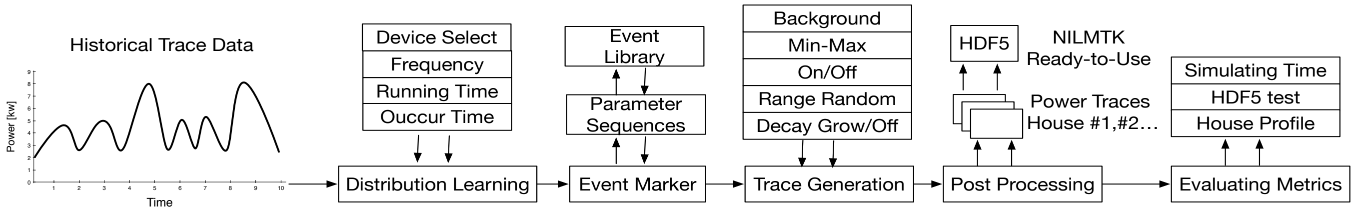


Fig. 2. Depiction of SmartSim's data generation pipeline from device modeling to trace generation.

are programmable, e.g., computers, or include thermostatic controllers often follow stable min-max models.

- **Random Range Models** draw power randomly within a fixed bounded range when active. These devices are defined by a bounded range where power usage is a random walk between the bounds. High-power non-linear loads, such as microwaves, follow a random range model.

In addition to the basic device usage models above, SmartSim also defines compound models for cyclic and composite devices. Cyclic devices operate at regular intervals based on a timer or a thermostat, while composite devices include multiple sub-loads, which may each follow different basic models above. For example, dryers include a resistive heating element and an inductive motor with different basic models. SmartSim includes an initial library of device energy models where each device is a composition of one or more of the basic models above. Each device model defines model type and the its device-specific parameters.

In addition to the individual device energy models, SmartSim also includes support for inserting sensor error into traces. Most sensors exhibit $\geq$1% error, which may influence energy data analytics. By default, SmartSim injects a 1% error in power values uniformly distributed around the correct value, such that the average power is equal over the trace period. We choose 1% because this is the rated error for the sensors we use. In addition to sensor errors, SmartSim is also capable of inserting data dropouts to emulate the imperfect data real systems often gather; dropouts are periods where the sensor is offline and no data is collected. Evaluating the effect of data dropouts on energy data analytics is important in practice, since real systems routinely experience significant dropouts, as discussed in prior work [3].

B. Device Usage Modeling

While the modeling methodology above determines how a particular device uses energy when on, it does not define a device's usage pattern: that is, the frequency, timing, and duration of a device's use over time. In some cases, device usage modeling is not necessary, since some devices operate in the background without user intervention. For example, the usage pattern of refrigerators, air conditioners, heaters, etc., are covered by the cyclical model above with cycles that are relatively static due to the use of timers or vary slightly due to the environment. However, the majority of devices in a home are interactive such that their use is based on the pattern of user activity in the home. Interactive devices may be used in a variety of different ways. For example, consider the usage pattern of a television, which users could turn on frequently for a short period of time and then turn off, or turn on infrequently for long periods of time before turning off.

The usage patterns of interactive devices are likely to impact the accuracy of energy data analytics, since the more on/off transitions that occur within a home the more stochasticity in the data, and the likelihood that two or more power state transitions will simultaneously occur. In addition, energy data analytics are often able to easily identify repeated cyclical patterns in data, e.g., by analyzing data the frequency domain. For each interactive device, SmartSim defines a distribution for its frequency, duration, and timing. In our initial library, we empirically derive these distributions from the Smart* dataset. That is, we compute histograms that specify the probability of a specific time-of-use (within each hour), frequency of use (number of times used in a day), and duration of use.

When generating a device's energy usage pattern for a given day, SmartSim first determines the number of times the device is used in a day by drawing a number from the frequency distribution. SmartSim then determines the duration of use by drawing from the duration distribution. While the frequency, duration, and time-of-use may be dependent for some devices, we leave more sophisticated device-specific modeling of usage patterns to future work. SmartSim is currently able to generate day-long and week-long traces. For week-long traces, the distributions above are distinct for each day, enabling them to capture weekly patterns, i.e., certain devices are used differently on specific days of the week.

IV. IMPLEMENTATION

We implement SmartSim in 845 lines-of-code (LOC) in `python` and integrate it with the NILM-TK toolkit to outputs trace in NILM-TK format, such that users can directly run NILM-TK algorithms on the data. NILM-TK implements the two most prominent NILM algorithms based on the Combinatorial Optimization (CO) algorithm proposed by Hart [15] and an algorithm based on Factorial Hidden Markov Models (FHMM) proposed more recently [5], [16]. Our initial library includes 25 different devices, usage patterns, and background noise modeled after data and usage patterns from the Smart* dataset [8]. SmartSim also includes tools to derive and insert new models. For example, SmartSim includes a tool that is able to extract the histogram of frequency, duration, and time-of-use from existing time-series power data. While our current library enables a wide range of experiments on energy analytics, as part of recent work on automated model derivation, we are also expanding the set of devices in our library.

V. EVALUATION

SmartSim's goal is to generate device-accurate home energy traces that are qualitatively and quantitatively similar to traces of real energy data. Our qualitative comparison is primarily visual, as depicted in Figures 3-5, and demonstrates that SmartSim's traces appear similar to traces of real data for a variety of devices. Figures 3-5 demonstrates real data and simulated data from SmartSim are qualitatively similar for a variety of different devices. Note that the simulated data here is generated by SmartSim dynamically from a parameterized device model and not from pre-existing data.

Since there is no metric to quantify the realism of a simulated energy trace, our quantitative evaluation instead compares the results of existing NILM algorithms using both real datasets, e.g., REDD [5] and Smart* [8], and SmartSim's simulated datasets to demonstrate that they yield comparable results. In particular, we use the NILM-TK implementations of the Combinatorial Optimization (CO) algorithm [15] from Hart's original work on NILM and the more recent approach based on Factorial Hidden Markov Models (FHMM) [5], [16].

We use SmartSim to construct a house that mimics Home A from the Smart* dataset [8] in that it uses models of the same devices and mimics its usage patterns. Home A is 1700 square foot with four full-time occupants. The home has a total of eight rooms including its basement. The main level has a living room, bedroom, kitchen, and bathroom, while the second story has two bedrooms and a bathroom. Major appliances include an electric dryer and washing machine, heat recovery ventilation (HRV) unit, dishwasher, refrigerator, and freezer. Home A has 35 wall switches, which primarily control room and closet lighting; switches also control an exhaust fan in each bathroom and the garbage disposal. Home A can represent the most regular residential houses.

Importantly, note that the device-accurate traces SmartSim generates differ from the real data, since they are randomly generated based on the frequency, duration, and time-of-use distributions for each device. Thus, we expect the results of the NILM algorithms to differ somewhat. Intuitively, if SmartSim's models are too simple, then the NILM algorithms should be able to perform better, i.e., have higher accuracy, on the simulated data. In contrast, if the results are similar in magnitude on SmartSim, Smart*, and REDD, then the simulated data has similar complexity.

Before presenting our results, we first discuss NILM evaluation metrics. While there are many evaluation metrics used in prior work (see [3], [2] for a survey), we focus on the Normalized Error in Assigned Power (NEP), the F1-score, and the Matthews Correlation Coefficient (MCC). The NEP metric is the error between the device's actual and inferred power usage, normalized by its total energy usage. If $\tilde{p}_i(t)$ denotes device p_i's actual power usage at time t and $p_i(t)$ denotes its inferred power usage, then we define NEP as shown below.

Dataset Name	*F1*	*NEP*	*MCC*	*Traning Time*	*Disaggregate Time*
REDD (CO)	0.38	1.66	0.110	3.27	3.18
REDD (FHMM)	0.38	1.51	0.077	385.17	2.07
Smart* (CO)	0.80	3.81	0.083	9.84	2.86
Smart* (FHMM)	0.89	1.79	0.021	814.71	1.63
SmartSim (CO)	0.76	0.87	0.281	9.34	3.58
SmartSim (FHMM)	0.79	0.74	0.303	776.19	2.82
SmartSim+Noise(CO)	0.54	1.99	0.033	10.01	3.05
SmartSim+Noise (FHMM)	0.63	1.01	0.042	934.93	2.36

TABLE I
CO AND FHMM RESULT COMPARISON ON REDD, SMART* AND SIMULATOR.

$$\delta = \frac{\sum_{t=1}^{T} |\tilde{p}_i(t) - p_i(t)|}{\sum_{t=1}^{T} \tilde{p}_i(t)} \tag{1}$$

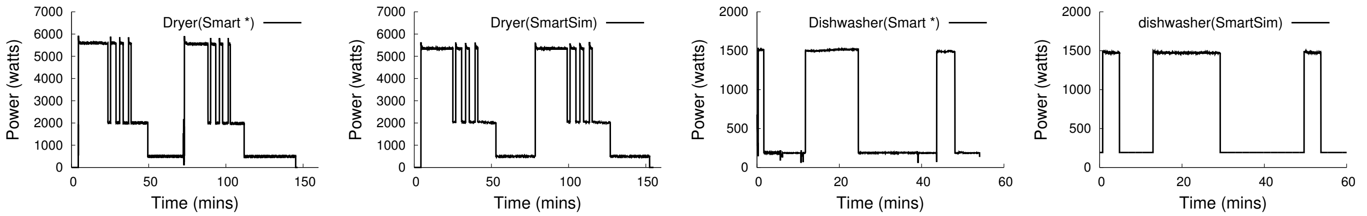
NEP is a measure of the reading-to-reading error in a NILM algorithm's inference for a given device. NEP may be computed for each device, or over an entire home by summing all device's inferred power usage and comparing it with the home's actual power usage. While there is no upper bound on NEP, a value of one indicates that errors are equal to the device's energy usage. In general, a NEP near one is not good, since simply inferring a device's energy usage to be zero at each time t results a value of one.

The F1-score and MCC metrics apply only to binary classifiers and thus assume each device is either on or off at any time t. Thus, at any time t, the state of an inferred device may be either a True Positive (TP), a True Negative (TN), a False Positive (FP), or a False Negative (FN). The fraction of time each state occurs defines a confusion matrix. The F1-score is then the harmonic mean of precision (or TP/(TP+FP)) and recall (or TP/(TP+FN)), where values range from 1.0 to 0.0 with 1.0 being the best. However, precision and recall only quantify the benefit of returning positive results, and not the benefit of correctly inferring negative results (or true negatives). A more balanced single measure of a binary classifier's overall performance is the MCC, as shown below, which considers all possible outcomes of a binary classifier. MCC values are in the range -1.0 to 1.0, with 1.0 is perfect inference, 0.0 is random inference, and -1.0 indicates detection is always wrong.

$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{2}$$

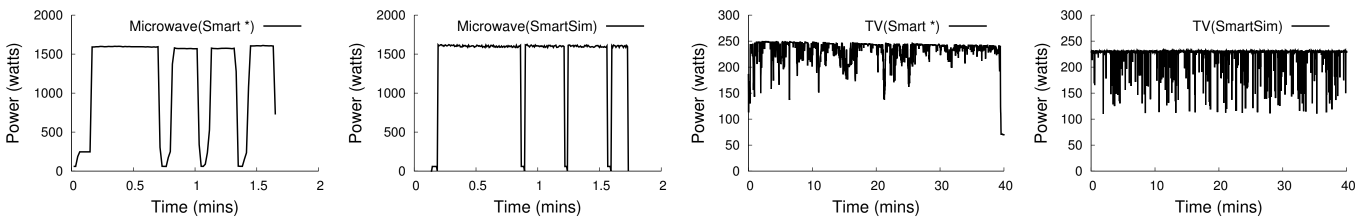
Table I shows the results of running the CO and FHMM NILM algorithms on the REDD dataset, Smart* dataset, and data generated from SmartSim (both with and without background noise). For CO, we disaggregate every device, while for FHMM we disaggregate only the top five devices in terms of aggregate energy usage, as is typical when using the FHMM approach [3]. FHMM is memory-intensive with running time that is exponential in the number of devices, which makes disaggregating a large number of devices intractable.

In this case, the F1, MCC, and NEP are the average over all the devices for CO and the average over only the top five



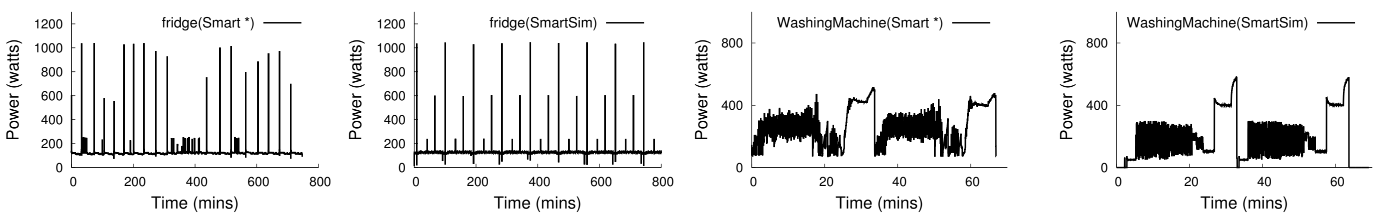
(a) Dryer in Smart* (b) Dryer in SmartSim (c) Dishwasher in Smart* (d) Dishwasher in SmartSim

Fig. 3. Examples of real and simulated device data.



(a) Microwave in Smart* (b) Microwave in SmartSim (c) TV in Smart* (d) TV in SmartSim

Fig. 4. Examples of real and simulated device data.



(a) Refrigerator in Smart* (b) Refrigerator in SmartSim (c) Washing Machine in Smart* (d) Washing Machine in SmartSim

Fig. 5. Examples of real and simulated device data. Note that the characteristics of SmartSim's generated data, including running time, usage pattern, etc., are not fixed, but randomly generated based on device models.

devices for FHMM. The magnitude of the results on the REDD and Smart* datasets are similar as in prior work [3], including the training and disaggregation times. The results also indicate the need for better understanding NILM algorithms, as there is little consistency when using the F1, NEP, and MCC metrics. For example, REDD using CO has a much worse F1-score than Smart* using CO, but a much better NEP, while the MCCs are similar. The accuracy results from SmartSim are similar in magnitude to those from Smart* in terms of the F1-score and MCC, although the NEP is slightly better using the simulated data. In particular, the MCC for Smart* using both CO and FHMM is near that of SmartSim with background noise added. Interestingly, while both NILM algorithms on the Smart* and SmartSim datasets report good F1-scores above 0.5 and nearly 0.9 in one case, their MCCs are near zero, which indicates random detection. This result indicates the F1-score, while common, may not be an adequate metric for evaluating NILM accuracy, as it does not consider the algorithms' high inaccuracy in predicting negative results, i.e., true negatives.

Of course, averaging the F1-score and MCC over many devices may mask large discrepancies between devices that influence the average. As a result, we also break down the MCC and F1-score for the top five devices using the FHMM algorithm from recent work [5], [16], [3]. Tables II and III shows the results, which demonstrate that the MCC and F1-score for both Smart* and SmartSim's data are similar in magnitude for most of the devices.

Device Name	*Smart* (FHMM)*	*SmartSim (FHMM)*
Dryer	0	0
LivingRoomOutlets	0	0
FurnaceHRV	-0.006	0.063
BedroomOutlets	-0.146	0
DiningRoomOutlets	-0.159	0

TABLE II
DEVICE-LEVEL FHMM RESULT COMPARISON ON SMART* AND SIMULATOR USING MCC.

VI. Related Work

There is significant prior work on energy data analytics, including NILM [2]. In many cases, the authors do not even evaluate their algorithms on real data due to the difficulty in collecting ground truth data from many devices in a home. In cases where prior work does evaluate the techniques on

Device Name	*Smart* (FHMM)*	*SmartSim (FHMM)*
Dryer	0.418	0.118
LivingRoomOutlets	0.999	0.971
FurnaceHRV	0.674	0.490
BedroomOutlets	0.963	0.304
DiningRoomOutlets	0.999	0.284

TABLE III
DEVICE-LEVEL FHMM RESULT COMPARISON ON SMART* AND SIMULATOR USING F1-SCORE.

real data, the authors do not study what characteristics of the data affect performance and accuracy, but rather simply report the results. Recent work has shown widely disparate results for different algorithms on different homes using different metrics. SmartSim's goal is to address these problems by enabling researchers to study energy data analytic performance by precisely controlling the data characteristics. To the best of our knowledge, SmartSim is the first device-accurate energy data simulator intended to support the development and study of new energy data analytics. While there has been some prior work on energy data modeling, it has been in the context of grid modeling for demand-response [18] and home modeling for distributed battery deployments [19].

VII. CONCLUSION

Research on energy data analytics is a highly active area due to the proliferation of smart meters. However, researchers currently do not have the tools to rigorously evaluate their techniques. Publicly-available datasets, of which there are few, are incomplete and do not enable researchers to control the energy usage characteristics of a home to understand its effect on performance. SmartSim addresses the problem by enabling researchers to generate device-accurate energy data traces for simulated homes. In this paper, we integrate SmartSim with NILM-TK, a publicly-available NILM toolkit, and compare its synthetically generated traces with those from a real home to show they yield similar results for representative analytics.

Acknowledgements. This research is supported by NSF grants IIP-1534080, CNS-1405826, CNS-1253063, CNS-1505422, and the Massachusetts Department of Energy Resources.

REFERENCES

[1] "U.S. Energy Information Administration, Frequently Asked Questions, How Many Smart Meters are Installed in the U.S. and who has them?" http://www.eia.gov/tools/faqs/faq.cfm?id=108&t=3, 2011.
[2] K. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is Disaggregation the Holy Grail of Energy Efficiency? The Case of Electricity," *Energy Policy*, vol. 52, no. 1, January 2013.
[3] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "NILMTK: An Open Source Toolkit for Non-Intrusive Load Monitoring," in *e-Energy*, June 2014.
[4] S. Barker, S. Kalra, D. Irwin, and P. Shenoy, "PowerPlay: Creating Virtual Power Meters through Online Load Tracking in Smart Homes," in *BuildSys*, November 2014.
[5] J. Kolter and M. Johnson, "REDD: A Public Data Set for Energy Disaggregation Research," in *SustKDD*, August 2011.
[6] S. McLaughlin, P. McDaniel, and W. Aiello, "Protecting Consumer Privacy from Electric Load Monitoring," in *CCS*, October 2011.
[7] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private Memoirs of a Smart Meter," in *BuildSys*, November 2010.
[8] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes," in *SustKDD*, August 2012.
[9] "Smart Meter Implementation Programme," https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/42737/1480-design-requirement-annex.pdf.
[10] J. Kolter and A. Ng, "Energy Disaggregation via Discriminative Sparse Coding," in *NIPS*, December 2010.
[11] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy Detection from Electricity Consumption Data," in *BuildSys*, November 2013.
[12] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, and P. Shenoy, "Non-Intrusive Occupancy Monitoring using Smart Meters," in *BuildSys*, November 2013.
[13] "Pecan St." http://www.pecanstreet.org/.
[14] S. Barker, S. Kalra, D. Irwin, and P. Shenoy, "Empirical Characterization and Modeling of Electrical Loads in Smart Homes," in *IGCC*, June 2013.
[15] G. Hart, "Nonintrusive Appliance Load Monitoring," *IEEE*, vol. 80, no. 12, December 1992.
[16] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised Disaggregation of Low Frequency Power Measurements," in *SDM*, April 2011.
[17] G. Hart, "Residential Energy Monitoring and Computerized Surveillance via Utility Power Flows," *IEEE Technology and Society Magazine*, vol. 8, no. 2, June 1989.
[18] O. Ardakanian, S. Keshav, and C. Rosenberg, "Markovian Models for Home Electricity Consumption," in *GreenNets*, August 2011.
[19] J. Venkatesh, B. Aksanli, T. Rosing, J. Junqua, and P. Morin, "HomeSim: Comprehensive, Smart, Residential Energy Simulation and Scheduling," in *IGCC*, June 2013.