Nadezda Buneeva, Andreas Reinhardt: AMBAL: Realistic Load Signature Generation for Load Disaggregation Performance Evaluation. In: Proceedings of the 8th IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 455-460, 2017.

AMBAL: Realistic Load Signature Generation for Load Disaggregation Performance Evaluation

Nadezda Buneeva
Technische Universität Clausthal
Clausthal-Zellerfeld, Germany
nadezda.buneeva@tu-clausthal.de

Andreas Reinhardt
Technische Universität Clausthal
Clausthal-Zellerfeld, Germany
reinhardt@ieee.org

Abstract*—Well annotated power consumption traces are a crucial prerequisite for the development and analysis of load disaggregation algorithms. Due to the high efforts required to collect such traces in the real world, their synthetic generation has emerged as a viable alternative. However, many current models for the synthetic trace generation simply combine statistical information about household occupancy with the energy consumptions of the most frequently performed user activities. While this may suffice for high-level analyses (i.e., considering groups of households or entire cities), such models do not reflect the actual diversity of consumption signatures in real data. We overcome this limitation in this paper by presenting a system design to model appliance power consumption at a user-definable accuracy. Our Automated Model Builder for Appliance Loads (AMBAL) allows to derive models from real device power consumption data collected by means of smart plugs. These models are represented by sequences of parametrized signatures; each model's complexity is kept minimized for its desired level of accuracy. We evaluate the accuracy of AMBAL's models for device traces with consumption patterns of different complexity, taken from existing appliance-level data sets. Moreover, a synthetic appliance trace generator is presented which allows to recombine appliance models in an effort to simulate user activities in homes with a definable complexity. The generated data is valuable for the development of data analysis algorithms (e.g., Non-Intrusive Load Monitoring), and we integrate it with the NILMTK framework to demonstrate that a similar disaggregation performance is achieved for actual and generated traces.

I. Introduction

Electrical consumption data has become widely available thanks to the deployment of smart metering infrastructures, and their analysis has gained significant research interest. Detailed information about electrical energy usage is an enabler for techniques to reduce and optimize energy use [1], predict future demand [2], or control peak consumption [3]. Instead of operating on aggregate traces, however, most such techniques exploit features of the power consumption characteristics of individual appliances. To gain access to these data, they internally rely on a combination of two components. Firstly, collected aggregate consumption data undergoes disaggregation into the contributions of individual appliances (also referred to as Non-Intrusive Appliance Load Monitoring, or NIALM). Secondly, the resulting appliance-level traces are being analyzed for specific characteristics that allow for the realization of aforementioned services. Accurate household-level power consumption models represent an important prerequisite for the development of load analytics techniques, as they allow for testing and improving algorithms without necessitating data collection campaigns. Consequently, approaches to derive such models have been extensively investigated lately [4, 5]. It needs to be noted, however, that a commonality among many solutions is their approach of modeling appliances as binary entities that consume constant power when switched on, and none when inactive. Characteristic power consumption fluctuations during appliance activity are not part of the models, and can consequently not be exploited by data analytics.

One approach to circumvent the limitations of such binary appliance model approximations is to use appliance-level data, e.g., from power consumption data sets such as REDD [6], Smart* [7], Tracebase [8], ECO [9], or AMPds [10]. However, too low sampling rates and intermittent sampling in some of these data sets cause the loss of useful information and make their processing more complicated. An insufficient number of appliances being monitored or too short monitoring periods may result in trace collections that do not allow for generalization. Likewise, often only a small number of households are part of data collection campaigns, which may lead to the collection of too similar traces. At last, some of the data sets contain only aggregated data or lack annotations, which strongly limits their usability for the given purpose, as many NIALM algorithms need annotated appliance-level traces for their training phase as well as for their evaluation.

Since the collection of real-world data is a costly and time-consuming process, the generation of synthetic power consumption traces represents a viable alternative. AMBAL, the principal contribution of this paper, is a solution to create such accurate power consumption models of appliance loads. Its established models enable the quick generation of numerous traces for testing purposes, the option to conjointly generate comprehensive annotations, and the opportunity to emulate different user activities. Except for specifying the desired model accuracy, AMBAL requires no manual interactions to establish appliance power consumption models. In order to cater for its practical use, we present a trace generator that synthesizes created device models into aggregate traces and use them for the evaluation of disaggregation algorithms. A realistic use case of the tool is demonstrated by integrating it with NILMTK [11] and comparing its disaggregation performance for actual and synthetically generated traces.

II. Related Work

Scientists have investigated techniques to model electricity consumption in the residential sector in different ways [4]. It has been shown that *bottom-up* modeling approaches are particularly well suited to create fine-grained models of a household's demand for electrical energy. They are based on input data such as the power consumption and technical properties of appliances, energy consumption measurements of individual homes (e.g., from electricity bills), and consumption-related behavior. One advantage is that profiles derived for a single dwelling are often representative for groups of similar homes and can be aggregated to any extent. Bottom-up approaches can be categorized into *statistical random* models, *probabilistic empirical* models and *time-of-use* models [5].

The former two model types are characterized by the use of empirically determined consumption patterns as input values, e.g., extracted from nation-wide surveys. Statistical random models extract the appliance (de-)activation times from such data and replicate them with some added randomness to introduce variations on household consumption [12]. Probabilistic empirical models, in turn, primarily rely on the empirical collection of information about household loads and their variabilities, and use probabilistic procedures to synthesize these values into aggregate consumption traces. In contrast to the aforementioned two classes, time-of-use models target to model power usage in relation to user behavior, and are mostly derived with the help of residents who specify such information in different kinds of surveys. Within this class, many models have been proposed in literature (e.g., [13–16]), in which the authors have commonly used time-of-use data from surveys to provide probability distributions for different occupant activities and underlying appliance usage.

The modeling techniques described above focus on the simulation of building occupants' energy usage behavior. For the modeling of individual appliances, however, most of them use average consumption data (daily/monthly/yearly mean values). This leads to a limited suitability of the generated data for power consumption analyses, since real device power signatures are often much more complex than simplified binary models. Only a few approaches (e.g., [17]) adopt appliance-level consumption traces from the previously collected datasets and re-use these trace segments. While reproducing appliance load patterns properly, such approaches often emit batches of identical traces, and may thus cause overfitting issues when machine learning techniques are being used.

Opposed to the modeling of appliance power consumption by means of statistical mean values, Barker et al. introduce a device-accurate power load modeling approach in [18]. Five basic model types were derived (resistive, inductive, capacitive, non-linear and composite loads), each of which has a certain pattern of power usage. The model types allow for capturing the power consumption patterns of household loads with high accuracy, based on actual measured data. However, part of the modeling process (trace segmentation and model choice) relies on manual interactions and is thus very time-consuming for large input data sets. Iyengar et al. applied the aforementioned concepts and provided an approach to automate the model derivation from appliance power consumption traces in [19]. This technique was validated both against the models manually created through Barker's approach as well as against real traces; the derived models showed a 1–3% deviation from the base data for most simple loads, and up to 10% for complex signatures. Since the methods used in this approach cannot be adjusted to the needs of the model user, however, it is not possible to influence the trade-off between the model's accuracy and its size. At last, in terms of the recombination of appliance models, Chen et al. have provided a simulation framework to concatenate individual simulation models in [20], which shares its fundamental idea with the synthetic trace generator we present in Sec. V.

III. Load Models

The principles of the approach to automated appliance load modeling introduced in this paper are based on the seminal work of Barker et al. [18], according to which electrical loads in the residential sector can be divided into resistive, inductive, capacitive, non-linear, and composite loads. Each of these load types exhibits certain power consumption patterns which can be described by one of the following models:

1) ON/OFF model: This model captures the behavior of the devices which consume a constant power in their active state. It is primarily applicable to resistive loads and simply assumes a power consumption of P_{on} for the duration of the appliance's activity (t_{active}) and P_{off} at other times.

2) ON/OFF Decay/Growth model: Inductive appliances (such as refrigerators) can be modeled through the superposition of an ON/OFF model with an exponentially decaying component. Likewise, capacitive appliance types may use a growth model (e.g., a logarithmic function) instead of the decaying component.

3) Stable Min-Max model: This model is applicable to devices that exhibit a stable baseline power and repeatedly experience short positive or negative power deviations from this value. It is constructed similar to a regular ON/OFF model with a stable power value P_{on}, but features two additional parameters: The maximum deviation of spike values from the stable power, P_{spike}, and λ as a parameter to describe the temporal distribution of the spike presence.

4) Random Range model: A random range model is described by parameters P_{min} and P_{max}. Within this range, the power variations conform to a uniform distribution. Many non-linear loads for which the Stable Min-Max model is inapplicable (such as desktop PCs) can be modeled this way.

5) Compound model types: Compound models consist of a combination of basic models described previously, possibly extended by further parameters. They are required to capture the behavior of complex loads better. One example for a compound load is a refrigerator's recurrent operation, for which information about cycle times is needed. Composite compound models, in turn, consist of a sequential ordering and/or the superposition of basic models.

A. Accuracy metric for models

To assess the accuracy of load models, a metric that quantifies the discrepancy between a consumption model and the originating trace is required. For the evaluation of the models derived by the approach introduced in this paper, we use the *Mean Absolute Percentage Error* (MAPE) value. The MAPE is a standard statistical measure of deviation between traces and computed as a percentage, according to Eq. (1).

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{P_{\text{data}}(t) - P_{\text{model}}(t)}{mean(P_{\text{data}})}\right| \cdot 100\% \qquad (1)$$

P_{data}	power values from the actual measured data
P_{model}	power values computed using the model
$mean(P_{\text{data}})$	arithmetic mean of power values in segment
N	number of samples in the segment

A key characteristic of MAPE is its independence of a given scale, which makes it well-suited to achieve comparable accuracy values across multiple comparisons.

IV. AMBAL: Automated Load Modeling

AMBAL is our system for the creation of appliance power consumption models. It allows to derive appliance models with selectable accuracy levels (i.e., MAPE values) to the real measured data while keeping the model size minimal for the desired level of accuracy.

A. Overview

An overview of the main operational phases of AMBAL is shown in Fig. 1, and briefly summarized as follows.

- *Preprocessing*. In this phase input traces are prepared for the further analysis, e.g., by eliminating sampling gaps and re-sampling input traces to the same sampling rate.
- *Extraction of active segments*. In this step, continuous consumption traces are segmented into phases during which the device is actively used (i.e., the operating cycles of the device which should be modeled). Periods of inactivity are used to separate different active segments.
- *Segmentation*. This second segmentation step is used to identify points during an appliance's active state at which its consumption characteristics change. Such changes often occur in composite loads when internal components are switched on or off. An adaptation of the load model type is often required at such load change points in order to fit a model accurately.
- *Model fitting*. Every segment resulting from the previous phase is fitted into ON/OFF and ON/OFF Decay/Growth models (cf. Sec. III) in order to find the one with the lowest MAPE value (i.e., best fit) to initial data.

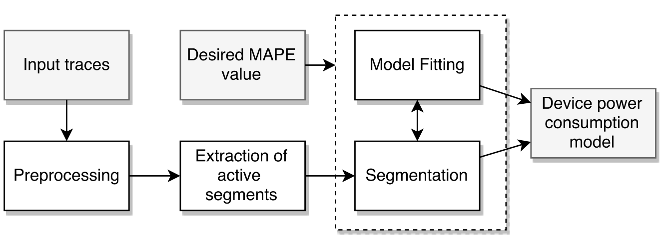
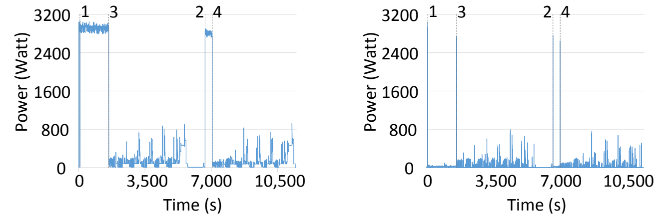


Fig. 1. Overview of the operational phases of AMBAL



(a) Power consumption trace of a washing machine with segmentation points as per Fig. 2b

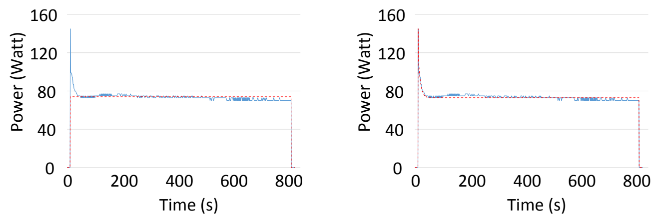
(b) Absolute difference values for the power consumption trace in Fig. 2a

Fig. 2. Trace segmentation when using absolute power difference values

The *segmentation* and *model fitting* phases are performed iteratively in order to reach the desired MAPE value of the device model. While the desired MAPE value is not reached for the extracted model, AMBAL identifies one more possible state change point in the input data and divides the segment at this point. It subsequently tries to fit the two resulting segments into the ON/OFF or the ON/OFF Decay/Growth basic models and re-computes the MAPE value to the input data. If an improvement to the MAPE has been made, the newly added segmentation point is maintained and the process is iteratively repeated. Should the MAPE value not decrease even when another segmentation point has been added or when the desired MAPE value is reached, the model fitting algorithm terminates. Details on trace segmentation and model fitting are provided in Sec. IV-B. As a penultimate step, AMBAL analyzes its extracted models for similarities which allow for their combination into aggregate models (cf. Sec. IV-C). At last, AMBAL checks if the resulting trace segments exhibit random fluctuations and fits them into the Stable Min-Max or Random Models in this case where meaningful (see Sec. IV-D).

B. Fitting models to parts of the active segments

While building load models for resistive loads is generally straightforward using ON/OFF models, compound appliance loads can seldom be approximated by fitting a single model to their entire activity period. However, there are points at which internal state changes (such as the activation of a component) lead to considerable changes of their power consumption. In order to derive accurate models for such appliances, AMBAL determines state changes based on the absolute power differences between subsequent measurements. To accomplish this technically, the absolute values of all step changes in power consumption are inserted into a list, sorted, and considered in descending order. An example for this segmentation is shown in Fig. 2a, which shows the power consumption trace of a washing machine. The absolute difference values for the power consumption trace are visualized in Fig. 2b, with the greatest four entries labeled and ranked by their magnitude. Transferring their locations into Fig. 2a again results in the segmentation points. The samples in-between these points are then approximated by individual models. Our segmentation approach is fundamentally different from prior work on automated appliance load model derivation [19], which uses a trace's *approximate entropy* to identify segmentation points.



(a) ON/OFF model; $P_{\text{on}} = 74W$, $P_{\text{off}} = 0W$, $t_{\text{active}} = 805s$. Resulting MAPE: 2.87%

(b) ON/OFF Decay model; $P_{\text{on}} = 73W$, $P_{\text{off}} = 0W$, $\lambda = 0.1$, $P_{\text{peak}} = 145W$, $t_{\text{active}} = 805s$. Resulting MAPE: 2.04%

Fig. 3. Fitting one segment of a refrigerator's consumption trace into an ON/OFF and an ON/OFF Decay model; the latter model is chosen by AMBAL due to its lower MAPE

Once AMBAL has identified a potential next segmentation point from the list of ordered power consumption differences, it tries to fit both the ON/OFF and the ON/OFF Decay/Growth model to the data contained in each of the two newly established segments. The optimum parameter values for the underlying mathematical functions are autonomously determined. The resulting parametrized models are then compared to each other, and the one with lower MAPE value to the real data is chosen. In case the MAPE still exceeds the user-specified target value, AMBAL iteratively selects additional segmentation points, as per the above description in Sec. IV-A.

An example for the AMBAL's modeling is presented in Fig. 3 for one working cycle of a refrigerator. In this case, AMBAL has identified an ON/OFF Decay model as the best fit in this case (Fig. 3b), since its MAPE value is better than for the simple ON/OFF Model (Fig. 3a). Since the choice of the model for segment is made based on the MAPE value, random models (Stable Min-Max and Random) cannot be considered for fitting in this step (the MAPE value would be higher for them than for ON/OFF models, even if they describe the initial data properly). Random models are therefore taken into account only in the last step of the algorithm (see Sec. IV-D).

C. Clustering and model aggregation

Many electrical appliances exhibit approximately the same consumption behavior when running in similar conditions and in the same mode. It can hence be meaningful to aggregate similar load models to reduce the number of models derived from input data and improve their resilience against outliers. AMBAL applies a two-step process to generalize its models by means of their aggregation.

Firstly, modeled segments are clustered based on their duration. AMBAL uses the DBSCAN clustering technique [21] for this purpose. An advantage of DBSCAN over other clustering algorithms is that the number of clusters does not need to be known in advance. Secondly, AMBAL builds an averaged version of the models contained in each cluster by computing the arithmetic mean of all segments stored in a cluster and running its modeling step (cf. Sec. IV-B) across the resulting data trace. This averaged model is then compared to the previously established individual models of the clustered segments. Only segments for which a MAPE difference of at most 1% between their individual models and the averaged model exists, are considered to be similar and are further described via the averaged model. Their individual models are discarded. Segments with a larger difference are retained as individual models, in order to avoid an excessive loss of precision.

D. Fitting of random models

After the main segmentation process is completed and the requested MAPE value is reached, the resulting segments are examined for presence of random fluctuations in the data. In order to decide whether a segment exhibits random behavior, the differences between actual trace data and data generated using the previously derived segment model are computed and analyzed. If the difference in power exceeds a threshold value P_{th} for more than a specified fraction of data values, the segment is assumed to contain random influences. In such case an appropriate random model (Stable Min-Max or Random) is used for modeling of this segment. The parameters P_{th} and the fraction of data points n used for the analysis of power difference values depend on the device operation duration and the appliance's power consumption; setting $P_{\text{th}} = 10W$ and $n = 10\%$ have led to good results in our experiments.

V. Synthetic Trace Generation

In order to investigate the quality of the traces synthetically generated from AMBAL's models, a tool to convert them into aggregate power consumption data is needed. A trace generator has been prototypically developed to this end. Through the simulation of user activities and different occupancy scenarios, it caters to the highly realistic simulation of device actuations. Moreover, it provides the opportunity to vary the number and types of devices simulated in the synthetic aggregate trace, and can thus generate traces of varying complexity for the analysis of load disaggregation algorithms. The generator outputs the traces for each device and an aggregated trace for the whole household which can be then used to evaluate the disaggregation performance. The aggregated daily power consumption trace is attained through superposition of individual AMBAL models, combined in accordance with underlying user activity models (following the work of Richardson et al. [22]).

A. Integration with NILMTK

The Toolkit for Non-Intrusive Load Monitoring (*NILMTK*) was initially presented by Batra et al. in [11]. This framework is designed to help researchers in the evaluation and testing of disaggregation algorithms. The toolkit provides a number of reference benchmark algorithms and a common set of accuracy metrics allowing to compare disaggregation approaches. NILMTK uses a *HDF5*-based file format [23] for input traces allowing to store the measurements along with the corresponding metadata. In order to enable using the data from synthetic trace generator introduced in this work, our synthetic trace generator was extended by a HDF5 converter in order to use generated synthetic consumption traces in conjunction with NILMTK.

VI. EVALUATION

After having introduced the AMBAL approach to automatically derive appliance models as well as the generator for synthetic traces, we assess the realistic nature of generated traces next. For our evaluation of AMBAL, we source input data from *ECO* [9] and *Tracebase* [8]. Both are open data sets comprising a collection of electrical appliance power traces. Their sampling rate of 1 Hz caters for a high data granularity without risking the inadvertent loss of short activity segments. From both data sets, we have sourced traces of different appliances for the duration of a single day, i.e., 86,400 seconds.

A. Model size vs. accuracy

The objective of our first evaluation is to provide an insight into the relation between a requested MAPE value and the corresponding model size for different device types. For this purpose, models for different devices from both data sets were derived and their complexity (in terms of the number of segments) was analyzed. The sizes of models generated from Tracebase data for different requested MAPE values are given in Table I; figures for ECO are specified in brackets. Intuitively, as 4%-MAPE models reproduce the signatures of most devices more accurately, they simultaneously present the highest requirements to storage (as, e.g., observed for the laptop computer). The tabulated values also confirm that many devices can be accurately modeled using 5 or less segments.

B. Accuracy improvements and algorithm termination

The AMBAL algorithm is designed to terminate as soon as the requested MAPE value is reached. However, as described earlier, this can lead to model overfitting for non-linear and composite loads if the requested MAPE value is too low. Fig. 4 visualizes the incremental improvements of the segmentation process (comparing the number of segments required to reach the corresponding MAPE) for desktop PC, LCD TV (both non-linear loads), washing machine and dishwasher (both composite loads). As can be seen from the figure, the first segments added to the model improve the MAPE value significantly. However, as the number of segments grows, improvements become less and less noticeable and most segments do not contribute much to the accuracy while making the model more complex.

TABLE I
MODEL SIZES FOR DIFFERENT MAPE VALUES FOR SELECTED DEVICES FROM THE TRACEBASE AND ECO (IN BRACKETS) DATA SETS

Appliance	Number of segments for a MAPE of...			
	10%	8%	6%	4%
Coffeemaker	5 (3)	6 (4)	8 (4)	10 (4)
Dishwasher	4 (7)	4 (7)	4 (7)	5 (7)
Freezer	1 (1)	1 (1)	1 (2)	3 (2)
Microwave	2 (2)	2 (2)	2 (3)	3 (3)
PC Desktop	1 (5)	1 (13)	8 (19)	143 (25)
PC Laptop	85 (2)	140 (4)	211 (19)	577 (371)
Refrigerator	1 (1)	1 (1)	2 (1)	2 (2)
TV	19 (2)	36 (2)	71 (2)	164 (2)
Washing machine	20 (9)	30 (9)	71 (15)	176 (77)
Kettle	1 (1)	1 (2)	1 (4)	1 (4)

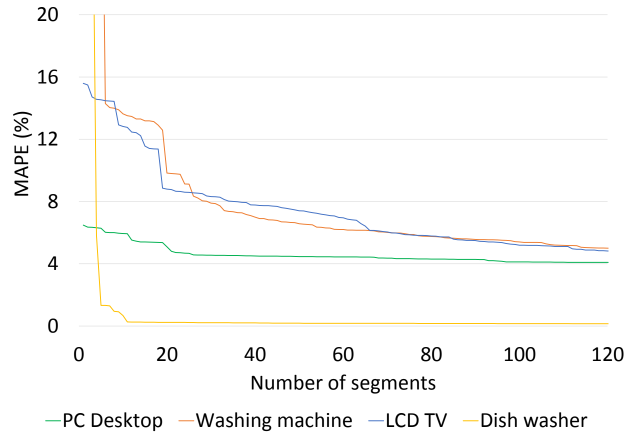


Fig. 4. Relation between the number of segments and corresponding MAPE value for different appliances

C. Usability of synthetically generated traces

In order to generate traces for their use in load disaggregation research, it is imperative for the synthetic data to reflect the features of actual consumption traces. To evaluate how closely the output traces of our synthetic appliance trace generator resemble actual load profiles, we utilize NILMTK's disaggregation performance as an indicator. The evaluation is benchmarked using the F1 score, a metric often used in information retrieval. It represents an average of the precision and recall and has a value range of (0,1]; in essence, the higher is its value, the better the disaggregation accuracy. Note that the F1 measure is scale-independent which allows for the comparison of the disaggregation accuracy across different device types. For the evaluation, we have used NILMTK v0.2; the latest version available at the time of writing. Disaggregation is performed using the Combinatorial Optimization disaggregation algorithm, one of the benchmark algorithms provided in NILMTK.

The following experiment aims at investigating whether the disaggregation performance differs when using real or synthetic aggregate data. We have thus generated aggregate traces containing activities of all appliances listed in Table II (Set A) as well as using two subsets of devices for those with large power consumption or long-lasting runtimes (Set L) and devices with small power demand or operated only for short periods of time (Set S). In both cases, user activity models were used to create realistic device actuation patterns. The real data used to test the system were the same from which appliance load signature models have been extracted with the help of AMBAL. Two disaggregation runs were performed, in both of which the disaggregation performance has been determined when NILMTK used the same data (real or generated) for training and testing.

As can be seen from Table II, almost all devices contained in set L can be well recognized in aggregate traces. This can be explained by their characteristic load signatures, long activity durations, and the average consumed power which distinguishes them from other appliances as well as from each other. In contrast, the disaggregation performance for devices

TABLE II
F1 SCORES OF THE ACHIEVED DISAGGREGATION PERFORMANCE

Device	real data			generated data		
	Set A	Set L	Set S	Set A	Set L	Set S
CD player	0.017	-	0.032	0.025	-	0.075
Iron	0.084	-	0.090	0.014	-	0.000
Vacuum cleaner	0.177	-	0.753	0.000	-	0.782
Desktop PC	0.753	0.726	-	0.848	0.923	-
Printer	0.004	-	0.061	0.016	-	0.495
TV	0.421	0.459	-	0.688	0.679	-
Microwave	0.000	-	0.000	0.000	-	0.000
Kettle	0.002	-	0.990	0.057	-	0.000
Dishwasher	0.286	0.785	-	0.238	0.469	-
Toaster	0.142	-	0.755	0.000	-	0.870
Cooking stove	0.331	0.522	-	0.214	0.311	-
Coffeemaker	0.036	-	0.151	0.000	-	0.009
Washing machine	0.174	0.294	-	0.097	0.296	-
Refrigerator	0.665	0.682	-	0.765	0.846	-

with short operational times and simple consumption patterns (set S) is, apart from few exceptions, often rather low. This effect is even more pronounced when set A (containing all 14 devices) is being disaggregated: The devices from set L dominate over other those only present in set S due to the aforementioned reasons, and even reduce the recognition rates of devices from set S. Differences between real consumption data and synthetically generated traces are, however, less noticable. If we compare the disaggregation F1 scores attained for real data with the scores for synthetic data in Table II to this end, it can be noted that similar trends can be observed for both real and generated data, pointing at the usability of synthetic traces for disaggregation evaluation. In fact, the properties of synthetic models even increase the disaggregation performance for some of them (e.g., toaster and refrigerator).

VII. CONCLUSIONS

We have introduced AMBAL as a solution for the automated modeling of appliance power consumptions. It works by analyzing appliance-level traces for the presence of characteristic consumption patterns and approximates them in the form of parameterized models with a definable level of accuracy. We have used AMBAL to model load signatures from two data sets commonly used for testing and evaluation of load analytics algorithms. Our evaluation results have shown that simple loads can be modeled with high accuracy at a small model complexity (1–5 model segments to achieve 4% MAPE value). To demonstrate a use case for the generated models, a trace generator has been implemented and used to synthesize aggregate consumption traces based on user activity models. For the evaluation of the data produced by the developed generator, we have compared the disaggregation performance of a benchmark algorithm implemented in NILMTK on real and synthetic data. Based on the high similarity of the results, we believe that synthetic data can be used to accelerate the evaluation and development of energy analytics algorithms.

REFERENCES

[1] J. Carroll, S. Lyons, and E. Denny, "Reducing Electricity Demand through Smart Metering: The Role of Improved Household Knowledge," Trinity Economics Papers (TEP) Working Paper No. 0313, 2013.

[2] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural Networks for Short-term Load Forecasting: A Review and Evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, 2001.

[3] J. Shen, C. Jiang, and B. Li, "Controllable Load Management Approaches in Smart Grids," *Energies*, vol. 8, 2015.

[4] L. G. Swan and V. I. Ugursal, "Modeling of End-use Energy Consumption in the Residential Sector: A Review of Modeling Techniques," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, 2009.

[5] A. Grandjean, J. Adnot, and G. Binet, "A Review and an Analysis of the Residential Electric Load Curve Models," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 9, 2012.

[6] J. Z. Kolter and M. J. Johnson, "REDD: A Public Data Set for Energy Disaggregation Research," in *Proceedings of Workshop on Data Mining Applications in Sustainability (SIGKDD)*, 2011.

[7] 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 *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, 2012.

[8] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, and R. Steinmetz, "On the Accuracy of Appliance Identification Based on Distributed Load Metering Data," in *Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT)*, 2012.

[9] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The ECO Data Set and the Performance of Non-intrusive Load Monitoring Algorithms," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (BuildSys)*, 2014.

[10] S. Makonin, B. Ellert, I. V. Bajic, and F. Popowich, "Electricity, Water, and Natural Gas Consumption of a Residential House in Canada from 2012 to 2014," *Scientific Data*, vol. 3, no. 160037, 2016.

[11] 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 *Proceedings of the 5th ACM International Conference on Future Energy Systems (eEnergy)*, 2014.

[12] R. Yao and K. Steemers, "A Method of Formulating Energy Load Profile for Domestic Buildings in the UK," *Energy and Buildings*, vol. 37, no. 6, 2005.

[13] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, "A Bottom-up Approach to Residential Load Modeling," *IEEE Transactions on Power Systems*, vol. 9, 1994.

[14] J. Widén and E. Wäckelgård, "A High-Resolution Stochastic Model of Domestic Activity Patterns and Electricity Demand," *Applied Energy*, vol. 87, no. 6, 2010.

[15] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic Electricity Use: A High-resolution Energy Demand Model," *Energy and Buildings*, vol. 42, no. 10, 2010.

[16] A. Ihbal, H.-S. Rajamani, R. A. Abd-Alhameed, and M. Jalboub, "The Generation of Electric Load Profiles in the UK Domestic Buildings Through Statistical Predictions," *Journal of Energy and Power Engineering*, vol. 6, no. 2, 2012.

[17] D. Fischer, A. Härtl, and B. Wille-Haussmann, "Model for Electric Load Profiles with High Time Resolution for German Households," *Energy and Buildings*, vol. 92, 2015.

[18] S. Barker, S. Kalra, D. Irwin, and P. Shenoy, "Empirical Characterization and Modeling of Electrical Loads in Smart Homes," in *Proceedings of the 4th International Green Computing Conference (IGCC)*, 2013.

[19] S. Iyengar, D. Irwin, and P. Shenoy, "Non-intrusive Model Derivation: Automated Modeling of Residential Electrical Loads," in *Proceedings of the 7th ACM International Conference on Future Energy Systems (eEnergy)*, 2016.

[20] D. Chen, D. Irwin, and P. Shenoy, "SmartSim: A Device-Accurate Smart Home Simulator for Energy Analytics," in *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2016.

[21] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*, 1996.

[22] I. Richardson, M. Thomson, and D. Infield, "A High-resolution Domestic Building Occupancy Model for Energy Demand Simulations," *Energy and Buildings*, vol. 40, no. 8, 2008.

[23] The HDF Group, "Hierarchical Data Format (HDF) Version 5," 2012.