Gemello: Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill

Nipun Batra
IIIT Delhi
nipunb@iiitd.ac.in

Amarjeet Singh
IIIT Delhi
amarjeet@iiitd.ac.in

Kamin Whitehouse
University of Virginia
whitehouse@virginia.edu

ABSTRACT

The first step to saving energy in the home is often to create an *energy breakdown*: the amount of energy used by each individual appliance in the home. Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each and every home. In this paper, we propose a more scalable solution called *Gemello* that estimates the energy breakdown for one home by matching it with similar homes for which the breakdown is already known. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. We evaluate this approach using 57 homes and results indicate that the accuracy of Gemello is comparable to or better than existing techniques that use sensing infrastructure in each home. The information required by Gemello is often publicly available and, as such, it can be immediately applied to many homes around the world.

1. INTRODUCTION

Buildings account for more than 30% of total energy usage around the world, of which up to 93% is due to residential buildings [41, 18, 42]. Some of this energy can be saved by producing an *energy breakdown*: the amount of energy used by each individual appliance in the home, akin to the itemised bills we get from grocery stores. With such a breakdown, utility companies can focus conservation programs on homes that have an especially inefficient appliance, such as a fridge, water heater, or air conditioner. Energy feedback can also help induce energy-saving behaviour in the occupants themselves [15, 4, 30, 43, 24].

Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each home and/or intensive manual training of the system. One approach, called non-intrusive load monitoring (NILM), tries to infer the energy breakdown based on the aggregate power trace from a single smart meter installed in the home. This approach is perhaps the most practical because smart meters are already being rolled out in millions of homes worldwide. However, current techniques require high resolution data (1 minute sampling frequency or higher) [39, 11, 4] while most smart meters today only support 15-minute or hourly sample rates to support time-of-use energy pricing. Even if smart meters had a higher sampling rate, most of the world does not yet have smart meters and many places do not even have plans to deploy them. Alternatives to NILM are more accurate but require specialized sensors to be installed inside the home [20], on each individual appliance [32, 26, 16], or on each circuit in the breaker box [37]. All of these solutions are limited by the need for instrumentation to be deployed in every home.

In this paper, we propose a more scalable solution called *Gemello* that produces an energy breakdown in homes without requiring new hardware to be installed in each home. Instead, Gemello estimates the energy breakdown for a home by matching it with similar homes that do have a hardware-based disaggregation solution. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. From an energy perspective, homes in the same geographic region are often very similar because they have similar construction methods, use the same heating fuels, and contain similar fridges, washing machines, and other appliances. Gemello exploits this fact to provide an energy breakdown for many homes in a region by instrumenting only a fraction of them.

Of course, no two homes are exactly identical and finding a perfect twin is unlikely. Therefore, Gemello uses a different set of matching homes to estimate the energy usage of each individual appliance. The key to success is the ability to define 'similarity' on a per-appliance basis. For example, homes with similar seasonal trends in their monthly energy bill are expected to have similar heating/cooling energy (referred as heating, ventilation, and air conditioning [HVAC] from now on); homes with similar square footage are expected to have similar lighting loads; and homes with a similar number of occupants are expected to have similar washing machine energy usage because that is driven by the amount of clothes worn each day. The energy usage of each appliance is predicted by a different set of features, and so Gemello finds a different set of homes to predict the energy usage of each appliance.

We evaluate this approach using 57 homes from the publicly available Dataport data set [38], in which the ground truth energy breakdown is measured by metering each appliance of the home individually for one year or more. Results

KDD '16, August 13 - 17, 2016, San Francisco, CA, USA

ISBN 978-1-4503-4232-2/16/08...$15.00

DOI: http://dx.doi.org/10.1145/2939672.2939735

show that the accuracy of Gemello is comparable to or better than established NILM techniques called FHMM and LBM, both of which require on-site, high-frequency power metering in each home [33, 45]. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. Our analysis shows that these results are robust with as few as 7 instrumented homes: the accuracy for HVAC loads is ≈69%, still 8% better than state-of-the-art approaches. The accuracy of Gemello becomes higher in homes with smart meters that provide power readings at 15-minute resolution.

The Gemello technique has potential for immediate impact because all of the information it requires is already available. Essentially all homes already have a standard power meter that provides total monthly energy usage for billing purposes, and in many regions this information can be downloaded online[1]. Household characteristics such as the size of the home and the number of people in the household are often publicly available and are already being used by companies to match homes for other types of eco-feedback[2]. Finally, many companies are already collecting submetering information from thousands of homes around the world[3]. By combining these three sources of information, Gemello could be used to provide an energy breakdown for homes around the world without the need to install new instrumentation in each location.

2. RELATED WORK

Various techniques for measuring appliance level energy consumption have been studied in the past. The simplest technique is to install appliance level sensors that monitor and report appliance energy consumption [26, 16]. Many commercial vendors are also selling appliance level sensors such as Kill-A-Watt[4] and Hobo plug load data logger[5]. Instead of directly monitoring the appliance of interest, a few recent systems have looked into *indirect* sensing techniques. Kim et al. [32] develop a system called Viridiscope that leverages the correlation amongst sensor streams, like using a vibration sensor on a fridge to tell if the compressor is running or not, and then using a model to determine fridge's power. Similarly, Clark et al. [13] develop a system called Deltaflow that employs energy harvesting sensors and performs computation on the activation of these sensors to determine appliance power draw. Jain et al. [25] install temperature sensors inside a home to estimate air conditioner energy usage. Gupta et al. [20], Chen et al. [12] and Gulati et al. [19] use the electromagnetic interference typically generated by electronic appliances to determine appliance usages. Previous work has also looked into using sensors deployed at household circuits (lesser in number than individual appliances) to infer per-appliance energy usage [37]. All these solutions are limited by the need for instrumentation to be deployed in every home. Previous research has shown that there are various challenges beyond cost of the system that prevent such systems from being scaled [23].

NILM is often considered the most practical approach towards generating energy breakdown. However, like other approaches mentioned above, it also requires instrumentation in each home. While the effort of deploying a smart meter may be less in comparison to other related approaches, most of the world still does not have smart meter infrastructure.

Since its inception in the early 1980s by George Hart [21], the field of NILM has seen various approaches based on different machine learning techniques leveraging different features of the power trace. However, many of these approaches require submetered data to learn a model of each appliance and these models have not been shown to generalize well across homes. Even if the models did generalise, NILM approaches require high frequency power metering with resolutions of 1-minute or higher. Very high frequency approaches (>10 kHz) [11] use features such as voltage-current trajectories to detect events in aggregate power time series. However, current smart meters do not collect data at such high rates because they are designed and deployed for the purposes of time-of-use pricing and there are currently no efforts to deploy devices suitable for energy disaggregation on a large scale. Therefore, these techniques, while promising, face real practical barriers before being used at scale.

Additionally, existing approaches for energy disaggregation [33, 39, 21, 45, 35] require a model of each appliance. The main differences between these techniques are how they are created and how they are used to infer the hidden states of the appliances based on the aggregate power readings. For example, some systems model appliances as finite state machines (FSMs). However, such approaches generally show poor accuracy on complex appliances such as washing machine and other electronics, as a FSM is a poor model for such appliances. Some systems assume the model is manually generated, learned from training data [21, 33], and in rare cases learned automatically [5]. In all cases, however, the accuracy of these models depends on how well the model approximates the true appliances in the home and it has not yet been demonstrated that these model-based approaches generalize well across homes. Only recently, researchers have started looking into *automatically* learning arbitrarily complex appliances using deep neural networks [29]. However, as claimed by the authors themselves, the work is just scratching the surface. These challenges are likely to impact the generalisability and applicability of existing approaches towards solving the energy breakdown problem.

3. APPROACH

The goal of Gemello is to predict the energy consumption of household appliances given easy to collect information such as a single aggregate energy reading per month and static household characteristics. The key intuition behind Gemello is that if per appliance energy consumption is available for a home then any other home that is 'similar' in nature will also have the same energy consumption for that appliance. This intuition leads to the following specific requirements for Gemello to work:

1. Small set of submetered homes (SH) for which per appliance energy consumption is directly measured.

2. Easy to collect information for all the homes in the

[1] http://www.greenbuttondata.org/
[2] https://opower.com/
[3] http://powerhousedynamics.com/
[4] http://www.p3international.com/products/p4400.html
[5] http://www.onsetcomp.com/products/data-loggers/ux120-018

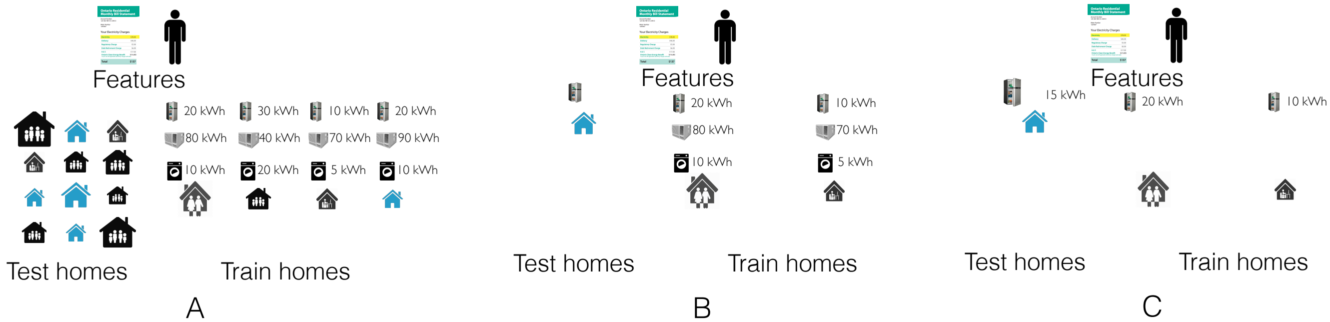

Figure 1: Overview of Gemello. A) We assume a small set of submetered homes (train homes) for which we have access to appliance level energy consumption. Both submetered homes and test homes (for which we need to predict per-appliance energy consumption) have access to monthly energy and static household features; B) We match a given test home to k closest train homes on monthly energy and static household features; C) We predict the appliance energy consumption of the test home as a function of the energy consumption of that appliance across the k closest train homes chosen in the previous step.

universal set U - This includes monthly energy consumption (j^{th} home represented by a vector $[EM_{j_1}, \dots, EM_{j_{12}}]$) which is anyway collected for billing purposes and static household information such as size and number of occupants (represented by a vector $[S_{j_1}, S_{j_2}, \dots]$).

For each target home i ($i \subset U$), we first derive the feature set (as discussed later) using vectors $[EM_{i_1}, \dots, EM_{i_{12}}]$ and $[S_{i_1}, S_{i_2}, \dots]$. Let these feature set be represented by $[I_{i_1}, I_{i_2}, \dots]$. We next derive the K nearest neighbours for the i^{th} home from amongst the submetered homes in SH by comparing the feature set $[I_{i_1}, I_{i_2}, \dots]$ with the feature sets derived for the remaining homes SH. Let this set of 'similar' homes be $[N_1, \dots, N_K]$. Monthly energy consumption for kth appliance (L_k) for ith home is then calculated as:

$$L_{k_i} = \frac{L_{k_1} + L_{k_2} + \cdots + L_{k_K}}{K}$$

where $L_{k_1}, \dots, L_{k_K}$ represent the monthly energy consumption for the same kth appliance from K identified 'neighbours' which are already sub-metered (and hence this information is directly measured). Figure 1 outlines the proposed Gemello algorithm.

We define 'similarity' differently for each electrical load. For example, homes having 'similar' monthly energy consumption are likely to have similar heating/cooling energy as the monthly energy consumption in many countries is often dominated by heating/cooling loads. Homes having 'similar' area are likely to have 'similar' lighting fixtures and thus may consume similar lighting energy. Homes having a 'similar' number of occupants may have 'similar' dish washer and washing machine energy usage, as clothes and dishes loads is likely to be proportional to the number of occupants. By defining 'similarity' differently for each load, we leverage the fact that no two homes are exactly identical in all respects, but that every home is likely to have a set of similar homes that can be leveraged for predicting energy consumption at the load level.

For an electrical utility that has no smart metering infrastructure or homes in their customer base that already have appliance level monitors already installed, it can chose these set of homes in a way that all the remaining (large number of) homes are 'similar' to one or more of these neighbourhood homes. Standard clustering techniques can be used on the feature set derived from already available information, as is used in Gemello for estimating K neighbours, to create cluster of homes from their customer base and then a small set of homes can be selected from each cluster for installing sub-metering infrastructure.

We develop further understanding of operation of Gemello by answering the questions below.

Q1. What features are derived from the monthly consumption and the static household information? Monthly energy consumption, $[EM_{i_1}, \dots, EM_{i_12}]$, is used to derive features such as seasonal trend, variance, range (difference between maximum and minimum) , ratio of minimum to maximum monthly energy consumption, skew, kurtosis and the 25^{th} and 75^{th} percentile. Static household features, $[S_{i_1}, S_{i_2}, \dots]$ include area, number of occupants and number of rooms. In total, we have 17 features from monthly data and 3 corresponding to static household properties. These are summarised in Table 1.

Q2. When higher resolution information, such as 15-minute AMI data is available, can it be leveraged in Gemello ?
Since 15-minute AMI data is very rich compared to monthly energy consumption, we divide the features derived from AMI data into the following three broad classes:

1. Seasonal and Trend - To derive these features, we first decompose the 15-minute AMI data time series using Seasonal Decomposition of Time Series by Loess (STL) [14] into seasonal and a trend 15-minute timeseries. We chose the period of 1 week for STL decomposition as homes show weekly patterns. Features derived in this class include max., energy and first 7 coefficients obtained from the Fourier transform of the seasonal component. These 7 coefficients correspond to periods of 1 through 7 days. This class will be broadly useful to estimate HVAC energy consumption since homes similar in the seasonal and the trend component are likely to be similar in HVAC energy consumption.

Feature category	Feature sub category	Features (f)	#f
Monthly energy	Raw monthly energy	12 month household energy aggregate	12
	Derived monthly energy	Variance over 12 month aggregate energy, Min. energy/Max. energy across 12 months, Max. energy-Min. energy across 12 months, (Max. energy - Min. energy)/Max. energy, Skew, Kurtosis, 25^{th}, and 75^{th} percentile of 12 months energy	9
Static household characteristics		Household area, # occupants, # rooms	3
15 min AMI data	Time series decomposition	Max. of seasonal and trend component, Energy of seasonal component across months of HVAC usage, Top-7 Fourier transform coefficients of seasonal component	14
	Weather correlation	Correlation between hourly aggregate energy and external temperature	1
	Usage patterns	Autocorrelation (lag 1 day), Fraction of energy used across 24 hours, Fraction of energy used across 7 days	32
	Step changes	Proportion of step changes in 3 bins (<500 Watts, >500 and <1000 Watts; >1000 Watts), Cluster centers of above 3 bins	6

Table 1: The features above are used to define similarity between homes. Gemello primarily uses features of the monthly energy bill and static household characteristics. However, it can also use features of 15-minute AMI power data if the home has a smart meter installed.

2. Temporal Patterns - Features derived in this class include fraction of energy usage across the 7 days of the week, energy usage across the 24 hours of the day and energy usage during the night time. Such features can be used to derive energy consumption of appliances showing temporal usage patterns. For example, certain working class populations may do their groceries, clothes and dishes on weekends and thus would have higher fridge, washing machine and dryer energy consumption on weekends. Further, fridge being a background load, running even in the night time when most other loads would be off, its energy can be easily derived by exploiting similarity in night hours energy usage.

3. Step changes - A step change is likely to be caused by an appliance changing its state. For example, washing machine turning ON could cause a step change of 500 Watts. We compute the fraction of step changes whose magnitude are less than 500 Watts, between 500 and 1000 Watts and greater than 1000 Watts. Step changes greater than 1000 Watts are caused by high power appliances such as HVAC, while step changes less than 500 Watts are caused by low power appliances such as fridge. On similar lines, we find find the cluster centres of the step changes in these three bins. The cluster centres will likely indicate the power of the most frequently cycling appliance (typically fridge).

4. Weather correlation - In addition to 15-minute AMI data, we also use 15-minute weather information as a feature. Homes having similar HVAC energy are likely to have similar response to weather and thus similar correlation between aggregate and temperature.

In total we derive 53 features from 15-minute AMI data. These are summarised in Table 1. Since these features are on different scales, we normalise them in the range 0 to 1 using standard scaling procedures[6].

Q3. From amongst all the features that we derive, how do we decide which features to use for estimating energy consumption of a particular appliance? Using the entire set of features may cause our model to overfit and give poor generalisation. Thus, we select the top-N features from the given feature set using standard feature selection algorithms.

Q4. Once we have selected the feature set, how do we select the neighbouring homes from where the average sub-metered information is available? Given a set of selected features, 'similarity' between homes can be found using distance functions. Standard distance functions such as Euclidean, Manhattan, etc. can be used. For each test home, we find its 'neigh bourhood' by finding its K nearest neighbours from the set of submetered homes, akin the well-known K nearest neighbours algorithm.

Q5. Once the neighbouring homes are selected, how do we estimate the monthly energy consumption for a particular appliance in the selected home? Having found K nearest neighbours for the test home, we predict the energy consumption of an appliance in the selected home as the average of energy consumption of that appliance across the K submetered homes. While we use the regular average, we can also use weighted average, which assigns higher weights to closer neighbours while making the prediction. We leave using weighted average for future work.

[6] https://github.com/robjhyndman/anomalous-acm

HVAC	Lights	Fridge	Dryer	Dish washer	Washing machine
31	12	21	32	26	16

Table 2: Number of homes in the dataset with ground truth values for each appliance

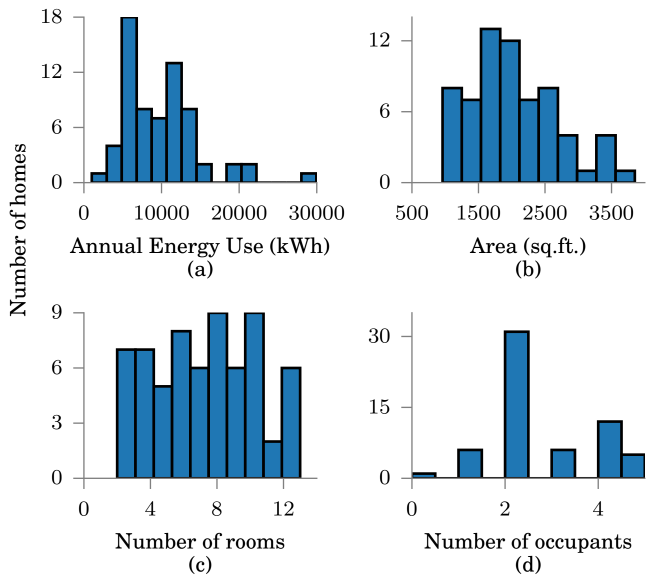

Figure 2: Our dataset has a wide diversity of homes, including very small homes/households as well as very large ones.

4. EVALUATION

4.1 Data set

We use the publicly available Dataport data set [38] for evaluating our approach. Dataport data set contains aggregate and appliance level power information for more than 600 homes in Texas, USA for up to 3 years. Submetered and household aggregate power data was collected at 1-minute resolution. 57 homes contain data for 1 year (2013) across 6 appliances of interest (fridge, HVAC, lights, dryer, washing machine and dish washer) and house metadata such as the number of occupants, area of the home and number of rooms. While the data set contains submetered data from a large number of appliances, only the chosen 6 appliances had data across significant number of homes[7]. Not all of these 57 homes monitored all these 6 appliances. Table 2 shows the number of homes for each of these 6 appliance types.

While many NILM data sets have been released in the past, none of the other data set satisfied the requirements of containing both aggregate and submetered power data for a large number of homes for at least one year. While the HES data set [46] contains data from a large number of homes for a large time, it does not contain aggregate power data. Given the high variability present in submetering across homes, we can't assume the aggregate to be the sum of submetered loads. The ECO data set [10] contains data from 6 homes for 8 months. But, there is no single load which is submetered across all these 6 homes. AMPds [36], iAWE [7], Blued [2] contain data only from a single home. REDD [34] contains data for a very short duration from 6 homes. While UK-DALE [28] contains data from 6 homes, only one of them has year long data.

We compare the attributes of our sample (57 homes) to the population (homes in the Texas and USA region). Figure 2a shows the distribution of annual energy usage across the 57 homes with a mean of 9923 kWh, the mean of Texas region is 13896 kWh [17]. Figure 2b shows the distribution of square footage area of households in our sample, with a mean of 2021 sq. ft. The mean for Texas region as reported in two different reports is 2393[8] and 1757 sq. ft [17]. The distribution of number of rooms in our sample is shown in Figure 2c having a mean of 7.3 rooms. While the statistics for average number of rooms per home in Texas are not available, a survey reports an average of ≈3 bedrooms per homes in the US[9]. Figure 2d shows the distribution of number of occupants per home with a mean of 2.6, the mean for Texas is 2.7.[10]

4.2 Baseline approaches

We compare the accuracy of our approach against the following two approaches:

4.2.1 Factorial Hidden Markov Model (FHMM)

FHMM is a well known NILM technique [33, 39] and commonly used as a benchmark by energy disaggregation researchers [8, 45, 34, 9, 35]. In this approach, each appliance is modelled as a hidden Markov model (HMM) containing 'n' states, where a state indicates mode of operation of the appliance. Many appliances can be modelled as 2 state appliances (ON or OFF). The HMM for each appliance is described by 3 parameters: 1) Initial probability (π)- containing the probability of each state at the initial time; 2) Transition matrix (A)- containing the probability of transition between states (e.g. ON to OFF); and 3) Emission matrix (B)- containing the power draw distribution of different states of the appliance. Energy disaggregation is done by performing a joint optimisation over the individual HMMs.

4.2.2 Latent Bayesian Melding (LBM)

In recent state-of-the-art work, Zhong et al. [45, 44] propose a method of melding models such as FHMM with prior information from population. Such prior information introduces constraints such as the expected number of cycles per day of an appliance, expected duration of usage per day and expected energy usage of an appliance per day. Their analysis shows upto 50% accuracy improvement over other approaches.

4.3 Evaluation metric

We define our metric based on prior work [33, 8]. Let the predicted and ground truth appliance energy for an appliance a in home h for month m be $Pred(h,a,m)$ and $GT(h,a,m)$ respectively. The absolute percentage error (PE) for the h,a,m triplet is given by:

[7]http://bit.ly/1QZKwG2

[8]http://1.usa.gov/1LRS08D

[9]http://1.usa.gov/1oaaa1K

[10]http://1.usa.gov/1oyImh3

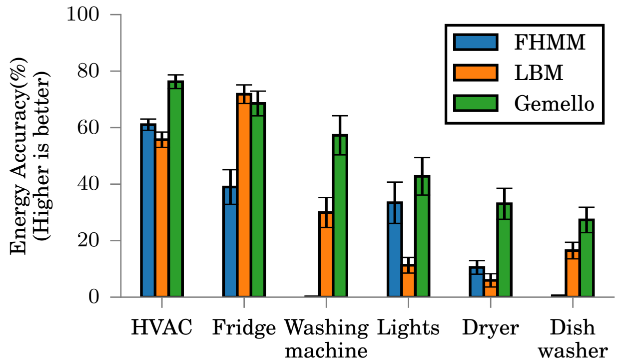

Figure 3: Gemello achieves Energy Accuracy that is comparable to or better than the baseline approaches across all appliances

$$\mathrm{PE(h,a,m)} = \frac{|\mathrm{Pred(h,a,m)} - \mathrm{GT(h,a,m)}|}{\mathrm{GT(h,a,m)}} \times 100\%$$

The percentage accuracy (PA) for h, a, m triplet is given by:

$$\mathrm{PA(h,a,m)} = \begin{cases} 100 - \mathrm{PE(h,a,m)}, & \text{if } \mathrm{PE(h,a,m)} \leq 100 \\ 0, & \text{otherwise} \end{cases}$$

The percentage accuracy ($PA(h,a)$) for an appliance a in home h is given by the mean of $PA(h,a,m)$ across all months.

$$PA(h,a) = \overline{PA(h,a,m)}$$

We now define our main metric Energy accuracy (%) for an appliance a to be the mean of $PA(h,a)$ across all homes.

$$\text{Energy accuracy}(a) = \overline{PA(h,a)}$$

In our results, we also calculate the standard error to show the variation in energy accuracy across homes. Higher 'Energy accuracy' indicates better disaggregation performance.

4.4 Experimental setup

Our experimental setup tries to replicate the following real world scenario- we have a small subset of submetered homes and a large number of homes without smart meters. However, our data set has the limitation that only a small number of homes (Table 2) containing aggregate and appliance energy for long duration (1 year) and household characteristics is available. Thus, we use the *leave-one-out* cross validation technique to evaluate our approach, where we assume the set of submetered homes to be all the homes except the test home. Since we also want to tune the parameters of our model, we do a nested cross-validation [22], where the inner loop of cross validation is used for model tuning. We tune our model on two parameters- the number of neighbours (K) and the top-N features. We vary K from 1 to 6 and N from 1 to max(number of features, 10). The top-N features are learnt using the ExtraTreeRegressor[11] using the default parameters provided in the Scikit-learn [40] implementation.

We perform this nested cross validation for each of the three feature sets- *monthly*, *static* and *monthly+static* features. We chose to perform nested cross validation on these three sets separately to ensure that we don't overfit when given *monthly+static* features. Of these three feature sets, the feature giving highest training cross-validation accuracy is chosen for reporting the accuracy on the test households [22].

We use data from 2013 for our evaluation. For all homes, we had aggregate and appliance monthly energy consumption for the 12 months. For the HVAC, the evaluation was done only on the months (May-October) in which it is typically used in Texas.

To compare our performance with FHMM, we train an FHMM on 15-minute appliance level data from all homes across the dataset using nilmtk [8, 27]. Each appliance is modelled as a 3 state HMM, as is commonly done in the literature [45, 8]. The FHMM is composed of an HMM for each home contains the 6 appliances of interest. It must be noted that the result of FHMM disaggregation is a 15-minute power signal for each appliance. We calculate the per month energy from this power signal.

To compare our performance with LBM, we use the implementation provided by the paper authors. The FHMM is trained on 15-minute appliance level data from the dataset. Population statistics such as appliance number of cycles, duration of usage are computed on the entire dataset. The LBM paper authors confirmed that the parameters and hyperparameters we learnt were reasonable. Like FHMM, LBM outputs a 15-minute power timeseries, from which we calculate the monthly energy consumption per appliance.

We have made our code public and provided easy to reproduce IPython notebooks on Github[12].

4.5 Results

Our main result in Figure 3 shows that Gemello gives comparable or better disaggregation performance across all loads in comparison to the 2 baseline approaches. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for latent bayesian melding based approach. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all.

We discussed the results of the benchmark algorithms with several energy disaggregation researchers. We found that there are several reasons why the performance of FHMM and LBM is poor in comparison to our approach. First, FHMM is known to work poorly when the 'unexplained' energy is high. Unexplained energy is the proportion of energy consumed by loads other than that in the model. While one may argue for the incorporation of additional loads, previous research has shown that FHMM performance degrades as number of appliances increases [31, 8]. It is also generally not feasible to create FHMMs with a large set of appliances as FHMMs have time complexity exponential in the number of appliances. Second, at 15-minute resolution, many appliance lose their distinctive electrical features. For example, a compressor cycle of a fridge may last than 15 minutes and may get confused with noise when data is sampled at 15-minutes interval. Third, FHMMs and in general signature based NILM approaches are prone to poor accuracy when

[11] http://bit.ly/20Qd50F

[12] https://github.com/nipunbatra/Gemello

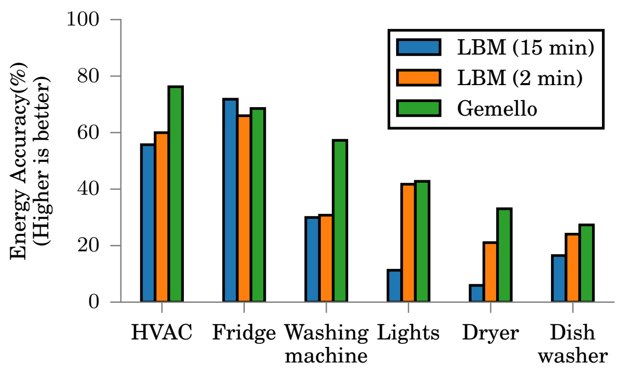

Figure 4: Gemello achieves higher accuracy with monthly sampling frequency than the state-of-the-art LBM algorithm achieves with 15-minute or even 2-minute sampling frequency.

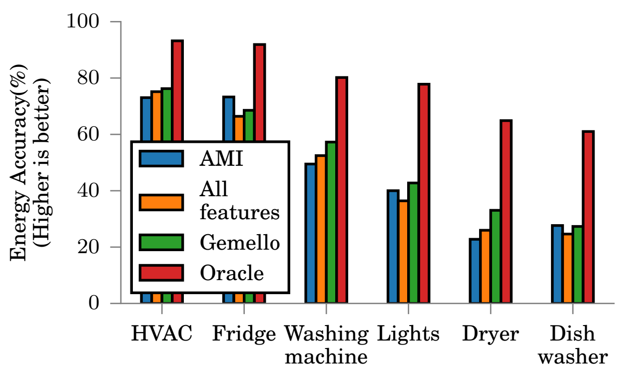

Figure 5: The performance of Gemello can be improved by up to 5% for some appliances if 15-minute AMI data is available. The Oracle illustrates the upper bound on accuracy that can be achieved with this data set if the neighbour selection algorithm were perfect.

appliances of similar signature are present. In our data set, the dish washer and the dryer had similar electrical signature. Fourth, FHMM is expected to do well for appliances such as fridge and HVAC which have dense usage. But, sparsely used appliances such as dryer and washing machine show poor accuracy. Fifth, appliances such as washing machine, dish washer also exhibit complex electrical signatures, such as spin cycle, dry cycle, rinse cycle [5], which make an HMM based model unsuitable. Most of these limitations are common across all energy disaggregation approaches.

We next evaluated the performance of our approach in comparison to LBM in the case higher resolution data was available to LBM. The LBM authors report high accuracy with 2-minute data. In Figure 4, we see that with an increase in resolution from 15 to 2 minutes, the accuracy of LBM improves. However, our approach still performs comparable or better across all appliances.

4.6 Analysis

4.6.1 *Sensitivity analysis on feature space*

Having established that Gemello gives better or comparable accuracy than state-of-the-art algorithms, we now analyse the performance of Gemello if additional data such as 15-minute AMI is also available. Figure 5 shows the performance of Gemello across different feature sets. 'All' set of features contains features from AMI data in addition to static household and monthly features. For fridge, the AMI features increase accuracy by 5%. On analysing the optimal set of features selected during nested cross validation when fed AMI features, we found that in 20/21 homes, proportion of step changes that are less than 500 Watts was among the optimal features. As we had hypothesised earlier, fridge is a duty cycled appliance often consuming less than 500 Watts. Thus, homes similar in proportion of step changes less than 500 Watts are similar in fridge consumption. We also found the fraction of energy consumed in afternoon hours to be an important feature as it occurs in 19/21 homes. Homes typically show low electrical activity during afternoon hours, which is often dominated by the always-on appliance fridge. Thus, fraction of energy in afternoon hours is useful for finding homes with similar fridge energy.

We now discuss the optimum set of features learnt by our approach. Table 3 shows the top features learnt over monthly and static features. For the HVAC, washing machine, lights and dryer the features match our intuition, as discussed earlier in Section 3. For dish washer, the top features include the aggregate monthly energy consumption in winter months (December and January). The mean correlation between dish washer energy usage across the months and aggregate energy in December and January is 0.49 and 0.43 respectively. We currently do not have an explanation for this correlation. Further, we find that the top features for fridge shown in the table do not have a high correlation with fridge energy consumption. However, these features do a good job in clustering homes, which ensures that homes similar in the feature are similar in fridge energy.

To find the best possible performance of Gemello on our dataset, we define the 'Oracle'. The 'Oracle' finds the 'optimal' subset of 'neighbourhood' homes from all possible sub sets of homes for each home for each load. The appliance energy data from these 'optimal' homes is then averaged to predict the appliance energy consumption for the test home. Here, 'optimality' is defined as giving the best energy breakdown accuracy. The substantially high accuracy of the 'Oracle' only highlights the potential of Gemello . That is to say, if we were to find *better* features for defining similarity between homes, our accuracies would tend towards the accuracy of the 'Oracle'.

4.6.2 *Sensitivity analysis on number of submetered homes*

We now discuss the impact of number of submetered homes ($\#SH$) on the accuracy of Gemello. A low number of homes is highly desirable as it will make the approach more cost-effective and thus scalable. To find the accuracy if only $\#SH$ homes were submetered, we do a cross-validation to tune our model (inner loop of the nested cross-validation described earlier). Then, on the remaining $Total - \#SH$ homes, we predict using the tuned model. We repeat this procedure 100 times ensuring that we pick different subsets of $\#SH$ homes randomly. Figure 6 shows the accuracy of HVAC and fridge with number of submetered homes ($\#SH$).

Appliance	Top features (prop. of homes present)
HVAC	Aggregate energy in summer months (.94), 75^{th} percentile of 12 month aggregate (.94)
Fridge	Area (1), Kurtosis of 12 month aggregate (.9), Range of 12 month aggregate (.86)
Washing machine	Area (1), Num. rooms (.94)
Lights	Area (1)
Dryer	Num. occupants (1), Area (1)
Dish washer	Monthly aggregate in December (.92), January (.77)

Table 3: Gemello automatically figures out which features are most effective at predicting energy usage for each appliance.

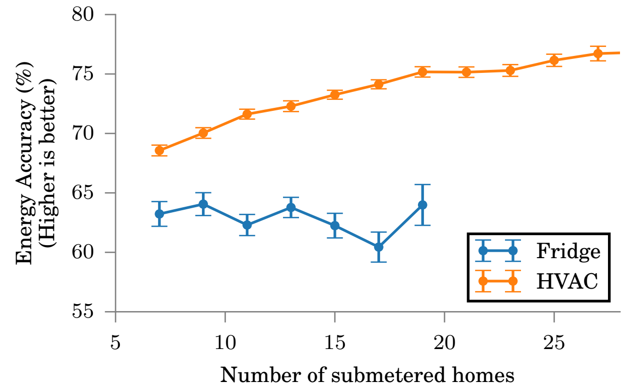

Figure 6: Gemello achieves high accuracy even with only a small fraction of submetered homes, making it a cost-effective and scalable solution. (There are fewer homes in the dataset with fridge data, thus fridge curve ends earlier.)

We chose the starting $\#SH$ as 7, as we have defined 6 as the maximum number of neighbours. We find that even with only 7 submetered homes, Gemello achieves $\approx$ 69% and $\approx$ 65% accuracy on HVAC and fridge respectively. The HVAC accuracy is still 8% better than the state-of-the-art. For the fridge, the accuracy is only 4% off from the accuracy when 21 homes were available. While we don't show the impact of number of submetered homes on accuracy for other appliances in interest of space, time and aesthetics, the conclusion holds that even with low number of submetered homes, our approach gives good performance.

5. LIMITATIONS AND FUTURE WORK

Our work has four major limitations, described below:

1. Since our approach only uses single reading per month, it can only estimate the monthly energy consumed by individual appliances, and cannot estimate the finer grained power signal. Estimating the fine grained signal has been shown to enable various applications such as activity recognition [1], anomaly detection [9, 3] and predicting household occupancy [6]. However, in homes that do have AMI metering infrastructure, the results of Gemello could potentially improve. This is a direction of future work. We believe that incorporating some of our features in the LBM approach could yield good results.

2. This approach is less likely to work for very uncommon appliances. It is likely limited to appliances which are commonly present across a large number of homes. This concern is common to current NILM algorithms as well, which are typically suited for more common appliances. The generality of this approach will grow linearly with the number of homes on which submetering infrastructure is deployed.

3. If people respond to energy breakdown and change their energy consumption, our feature vector consisting of historical energy usage is likely to be less effective. We believe that such changes can be captured by running changepoint detection algorithms over longer duration historical energy data.

4. 'Outlier' homes (such as homes consuming very high energy), which are likely candidates for energy feedback may have wrong predictions owing to the fact that 'neighbouring' homes are not representative of their energy consumption. We plan to address this concern in future work by first segregating these 'outlier' homes based on a model of home's energy consumption as a factor of various static and dynamic characteristics such as: area, number of occupants, etc.

6. CONCLUSIONS

A detailed energy breakdown is the first step towards saving energy in a home. We started with the goal of creating an energy breakdown solution that - works with whatever data is easily accessible, is able to scale across large number of homes and requires minimal capital expenditure involved. In order to achieve these objectives, we completely flipped the way we look at the problem. Rather than the existing bottom-up approach of using modelling to identify electrical signatures, we used the top-down approach of using modelling to identify home level characteristics that correlate well with appliance level energy consumption. We show that such home level characteristics can be easily calculated with static household information and monthly electricity data both of which are readily available. Results show that our proposed approach that satisfies the objectives we started off with, gives comparable or better energy breakdown accuracy than the state-of-the-art NILM techniques which rely on high resolution power consumption data from smart meters. Further, we found that our approach is robust with a small number of instrumented homes. We believe that Gemello has potential for immediate impact because all the information it requires is already available.

7. ACKNOWLEDGEMENTS

We would like to sincerely thank Fred Jiang, Nic Lane, Taiwoo Park, Polly Huang and Wen Hu who provided feedback on an early version of this work at Sensys 2015 PhD forum. Arunchandar Vasan, Venkatesh Sarangan and Mario Berges provided useful suggestions. Jack Kelly, Oliver Parson, Mingjun Zhong, Manoj Gulati, Milan Jain provided detailed pre-submission reviews for our paper, in addition to deep technical discussions. Mingjun Zhong very kindly helped us use his LBM algorithm through Skype discussions and 100+ emails. Grant Fisher provided information about the dataset.

This work was funded in part by NSF awards 1305362, 1038271, and 0845761. The first author is supported by TCS Research and Development PhD fellowship. We also thank Department of Electronics and Information Technology (Government of India) for funding the project (Grant Number DeitY/R&D/ITEA/4(2)/2012 and Grant Number ITRA/15(57)/Mobile/HumanSense/01). We would also like to thank Google for providing support for the conference.

8. REFERENCES

[1] J. Alcalá, O. Parson, and A. Rogers. Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 225–234. ACM, 2015.

[2] K. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe, and M. Bergés. BLUED: A fully labeled public dataset for Event-Based Non-Intrusive load monitoring research. In *Proceedings of 2nd KDD Workshop on Data Mining Applications in Sustainability*, pages 12–16, Beijing, China, 2012.

[3] P. Arjunan, H. D. Khadilkar, T. Ganu, Z. M. Charbiwala, A. Singh, and P. Singh. Multi-user energy consumption monitoring and anomaly detection with partial context information. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 35–44. ACM, 2015.

[4] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213–234, 2013.

[5] S. Barker, S. Kalra, D. Irwin, and P. Shenoy. Empirical characterization and modeling of electrical loads in smart homes. In *IEEE International Green Computing Conference*, pages 1–10, Arlington, VA, USA, 2013.

[6] N. Batra, R. Baijal, A. Singh, and K. Whitehouse. How good is good enough? re-evaluating the bar for energy disaggregation. *arXiv preprint arXiv:1510.08713*, 2015.

[7] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava. It's Different: Insights into home energy consumption in India. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2013.

[8] 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 *Fifth International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014.

[9] N. Batra, A. Singh, and K. Whitehouse. If you measure it, can you improve it? exploring the value of energy disaggregation. In *Proceedings of the second ACM International Conference on Embedded Systems For Energy-Efficient Built Environments*. ACM, 2015.

[10] 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 First ACM International Conference on Embedded Systems For Energy-Efficient Buildings*. ACM, 2014.

[11] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of industrial ecology*, 14(5):844–858, 2010.

[12] K.-Y. Chen, S. Gupta, E. C. Larson, and S. Patel. Dose: Detecting user-driven operating states of electronic devices from a single sensing point. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 46–54. IEEE, 2015.

[13] M. Clark, B. Campbell, and P. Dutta. Deltaflow: submetering by synthesizing uncalibrated pulse sensor streams. In *Proceedings of the 5th international conference on Future energy systems*, pages 301–311. ACM, 2014.

[14] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73, 1990.

[15] S. Darby. The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*, 2006.

[16] S. DeBruin, B. Ghena, Y.-S. Kuo, and P. Dutta. Powerblade: A low-profile, true-power, plug-through energy meter. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 17–29. ACM, 2015.

[17] EIA. Household energy use in texas, 2014.

[18] M. Evans, B. Shui, and S. Somasundaram. *Country report on building energy codes in india*. Pacific Northwest National Laboratory, 2009.

[19] M. Gulati, S. S. Ram, A. Majumdar, and A. Singh. Single point conducted emi sensor with intelligent inference for detecting it appliances. In *IIITD Technical Report*, 2015.

[20] S. Gupta, M. S. Reynolds, and S. N. Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139–148. ACM, 2010.

[21] G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.

[22] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.

[23] T. W. Hnat, V. Srinivasan, J. Lu, T. I. Sookoor, R. Dawson, J. Stankovic, and K. Whitehouse. The hitchhiker's guide to successful residential sensing deployments. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*,

pages 232–245. ACM, 2011.

[24] M. Iyer, W. Kempton, and C. Payne. Comparison groups on bills: Automated, personalized energy information. *Energy and Buildings*, 38(8):988–996, 2006.

[25] M. Jain, A. Singh, and V. Chandan. Non-intrusive estimation and prediction of residential ac energy consumption. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–9. IEEE, 2016.

[26] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. Culler. Design and implementation of a high-fidelity ac metering network. In *Information Processing in Sensor Networks, 2009. IPSN 2009. International Conference on*, pages 253–264. IEEE, 2009.

[27] J. Kelly, N. Batra, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava. Nilmtk v0. 2: a non-intrusive load monitoring toolkit for large scale data sets: demo abstract. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, pages 182–183. ACM, 2014.

[28] J. Kelly and W. Knottenbelt. UK-DALE: A dataset recording UK Domestic Appliance-Level Electricity demand and whole-house demand. *ArXiv e-prints*, 2014.

[29] J. Kelly and W. Knottenbelt. Neural nilm: Deep neural networks applied to energy disaggregation. *arXiv preprint arXiv:1507.06594*, 2015.

[30] J. Kelly and W. Knottenbelt. Does disaggregated electricity feedback reduce domestic electricity consumption? a systematic review of the literature. In *3rd International NILM Workshop*, 2016.

[31] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han. Unsupervised disaggregation of low frequency power measurements. SIAM.

[32] Y. Kim, T. Schmid, Z. M. Charbiwala, and M. B. Srivastava. Viridiscope: design and implementation of a fine grained power monitoring system for homes. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 245–254. ACM, 2009.

[33] J. Z. Kolter and T. Jaakkola. Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 1472–1482, La Palma, Canary Islands, 2012.

[34] J. Z. Kolter and M. J. Johnson. REDD: A public data set for energy disaggregation research. In *Proceedings of 1st KDD Workshop on Data Mining Applications in Sustainability*, San Diego, CA, USA, 2011.

[35] S. Makonin, F. Popowich, I. V. Bajic, B. Gill, and L. Bartram. Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring.

[36] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic. AMPds: A Public Dataset for Load Disaggregation and Eco-Feedback Research. In *IEEE Electrical Power and Energy Conference*, Halifax, NS, Canada, 2013.

[37] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle. Circuit-level load monitoring for household energy management. *Pervasive Computing, IEEE*, 10(1):40–48, 2011.

[38] O. Parson, G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt, and A. Rogers. Dataport and nilmtk: a building data set designed for non-intrusive load monitoring. 2015.

[39] O. Parson, S. Ghosh, M. Weal, and A. Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pages 356–362, Toronto, ON, Canada, 2012.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[41] L. Pérez-Lombard, J. Ortiz, and C. Pout. A review on buildings energy consumption information. *Energy and buildings*, 40(3):394–398, 2008.

[42] B. Shui, M. Evans, H. Lin, W. Jiang, B. Liu, B. Song, and S. Somasundaram. *Country Report on Building Energy Codes in China*. Citeseer, 2009.

[43] H. Wilhite and R. Ling. Measured energy savings from a more informative energy bill. *Energy and buildings*, 22(2):145–155, 1995.

[44] M. Zhong, N. Goddard, and C. Sutton. Signal aggregate constraints in additive factorial hmms, with application to energy disaggregation. In *Advances in Neural Information Processing Systems*, pages 3590–3598, 2014.

[45] M. Zhong, N. Goddard, and C. Sutton. Latent bayesian melding for integrating individual and population models. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 3617–3625. Curran Associates, Inc., 2015.

[46] J.-P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, and C. Evans. Household Electricity Survey. A study of domestic electrical product usage. Technical Report R66141, DEFRA, May 2012.