



Lawrence Berkeley
National Laboratory

Extracting Baseline Electricity Usage with Gradient Tree Boosting

Taehoon Kim
Dongeun Lee
Jaesik Choi
Anna Spurlock
Alex Sim
Annika Todd
Kesheng Wu

Lawrence Berkeley National Laboratory
One Cyclotron Road, Berkeley, CA 94720

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Extracting Baseline Electricity Usage with Gradient Tree Boosting

Taehoon Kim¹, Dongeun Lee¹, Jaesik Choi¹, Anna Spurlock², Alex Sim², Annika Todd², and
Kesheng Wu²

²Lawrence Berkeley National Laboratory, Berkeley CA, USA

¹Ulsan National Institute of Science and Technology, Ulsan, Korea

November 15, 2015

Abstract

To understand how specific interventions affect a process observed over time, we need to control for the other factors that influence outcomes. Such a model that captures all factors other than the one of interest is generally known as a baseline. In our study of how different pricing schemes affect residential electricity consumption, the baseline would need to capture the impact of outdoor temperature along with many other factors. In this work, we examine a number of different data mining techniques and demonstrate Gradient Tree Boosting (GTB) to be an effective method to build the baseline. We train GTB on data prior to the introduction of new pricing schemes, and apply the known temperature following the introduction of new pricing schemes to predict electricity usage with the expected temperature correction. Our experiments and analyses show that the baseline models generated by GTB capture the core characteristics over the two years with the new pricing schemes. In contrast to the majority of regression based techniques which fail to capture the lag between the peak of daily temperature and the peak of electricity usage, the GTB generated baselines are able to correctly capture the delay between the temperature peak and the electricity peak. Furthermore, subtracting this temperature-adjusted baseline from the observed electricity usage, we find that the resulting values are more amenable to interpretation, which demonstrates that the temperature-adjusted baseline is indeed effective.

1 Introduction

With measurements recorded for most customers in a service territory at hourly or more frequent intervals, advanced metering infrastructure (AMI) captures electricity consumption in unprecedented spatial and temporal detail. This vast and growing stream of data, together with cutting-edge data science techniques and behavioral theories, enables 'behavior analytics:' novel insights into patterns of electricity consumption and their underlying drivers [9, 33].

As electricity cannot be easily stored, electricity generation must match consumption. When the peak demand exceeds the generation capacity, a blackout would occur, typically during the time when consumers need electricity the most [19, 37]. Since increasing generation capacity is expensive and takes years to implement, regulators and the generators have devised a number of pricing schemes intended to discourage unnecessary consumption during peak demand periods.

To measure the effectiveness of a pricing policy on the peak demand, one can analyze electricity usage data generated from AMI. Our work focuses on extracting *baseline models* of household electricity usage for a behavior analytics study [4, 9, 33]. The baseline models would ideally capture the pattern of household electricity usage accurate enough to predict the future electricity usage of households for years into the future.

Although this work shares some similarities with other works on forecasting electricity demands and prices [29, 2, 31], there are a number of distinctive characteristics that necessitate us considering a different class of data mining methods. The fundamental difference between a baseline model and a forecast model is that the baseline model needs to capture the core behavior that persist for a long time, while the forecast model typically aims at making a forecast for the next few cycles of a time series in question. Typically, techniques that make forecasts for years into the future are based on highly aggregated time series with month or year as time steps [1, 2], whereas those that work on time series with shorter time steps typically focus on making forecasts for the next day or the next few hours [10, 23, 24, 32].

In the specific case that has motivated our work, the overall objective is to study the impacts of proposed pricing policies. The process of designing these pricing schemes, recruiting participants for a pilot study, implementing the pricing schemes, and monitoring the impacts have taken a few years. The baseline model is typically based on observed consumption prior to the implementation of the new pricing schemes, and applied to predict what consumer behavior would be without the pricing changes. This is challenging because the baseline model needs to not only capture intraday household electricity usage but also be applicable for years. Furthermore, in preliminary tests, we have noticed that the impact of the pricing schemes is weaker than the impact of other factors such as temperature, therefore, the baseline model must be able to incorporate outdoor temperature when making predictions.

This work examines a number of methods for developing the baseline models that could satisfy the above requirements. We use a large set of AMI data to exercise these methods and evaluate their relative strengths. The bulk of data in this work is hourly electricity usage from randomly chosen samples of households from a region of the US where the electricity usage is highest in the afternoon and evening during the months of May through August. The methods we choose to extract the baseline models all require a large amount of sample input, therefore the models developed represent average behavior, not behavior specific to any individual household.

In the remaining of this paper, we briefly present the background and related work in Section 2 and describe the residential electricity usage data used in this study in Section 3. We describe the methods used to extract the new type baseline in Section 4 and discuss the output from these methods in Section 5. A short summary is provided in Section 6.

2 Background

Energy management has become an important problem all around the world. The recent deployment of residential AMI makes hourly electricity consumption data available for research, which offers a unique opportunity to understand the electricity usage patterns of households. In particular, understanding how and when households use electricity is essential to regulators for increasing the efficiency of power distribution networks and enabling appropriate electricity pricing. One concrete objective from several current pricing studies is to design new rules and structures in order to reduce the peak demand and therefore level out total electricity usage [11, 33].

The recent influx of massive amounts of electricity data from AMIs lead to various research on energy behavior such as electricity consumption segmentation [7, 13, 36, 6, 35, 28, 20], forecasting and load pro-

filing [12, 18, 14], and targeting customers for an air-conditioning demand response program to maximize the likelihood of savings [21].

An important tool for this problem is classifying and representing different households with different load profiles [3, 14, 20]. Accurately identifying the load profiles will allow the researchers to associate observed electricity usage with consumer energy behavior. Load profiling could identify policy relevant energy lifestyle segmentation strategies, which can lead to better energy policy, improve program effectiveness, increase the accuracy of load forecasting, and create better program evaluation methods [20].

Accurate prediction or load forecasting of electricity usage is very important for the industry [22, 25]. For example, long-term usage forecasting for more than one year ahead is important for capacity planning and infrastructure investments. Short-term forecasting is used in the day-ahead electricity market, determining available demand response, and increasing demand side flexibility. Many statistical methods and machine learning methods are used in this process [1, 12, 18, 22, 25, 30]. For example, some authors prefer supervised machine learning methods such as support vector machines [5, 17], some use statistical models such as dynamic regression [22], while others advocate for neural networks and artificial intelligence approaches [25]. Typically, these methods transform the time series of historical data into a time scale such that the predictions are made for the next time step or the next few time steps.

Household electricity usage depends on many factors, such as outdoor temperature, appliances in the house, number of occupants, the energy behavior of the occupants, the time of day, day of the week, seasons, and so on [4, 34]. Some of the prediction models focus on aggregated demand and therefore could parameterize many factors affecting the usage of an individual household [30]. From the study of earlier models, we learned that a household’s electricity usage is strongly periodic, in that the daily electricity usage repeats every day and every week. Given any two consecutive days, their usage patterns are very similar to each other. Given any two consecutive weeks, their electricity uses are also similar to each other. Throughout a year, the overall electricity usage follows the pattern of temperature change. To predict correctly the electricity usage, we need to capture the same factors in our own models.

3 Dataset

The households in our dataset are divided into 6 different groups based on how they participate in the study and which pricing scheme is used. There is a control group following the practice of randomized controlled trials. In later discussion, this group is labelled **control**. As expected, the control group stay with the original pricing scheme throughout the testing period.

Some households are labeled as active participants because they explicitly opt in to new pricing schemes offered. There are two different pricing schemes offered. The group **active1** uses pricing scheme 1 and the group **active2** uses pricing scheme 2.

The other three groups correspond to households that *passively* participate in one of the two policies or both of them: **passive1** denotes a group of households with passive participation in the pricing scheme 1, **passive2** with passive participation in the pricing scheme 2, and lastly **passive3** with passive participation in both of the schemes.

In our study of baseline extraction methods, we use a subset of households from each of the 6 groups. Furthermore, we select households with measurement data for all three years during the study. The number of unique households in our dataset was 6,295.

3.1 Electricity Usage Data

Our electricity usage data have hourly electricity consumption records of individual households for three years. The unit of electricity is in kilowatt-hour (kWh). The total number of hourly data points is 160,125,432, from which we focus on data generated during the summer that is accountable for most of electricity usage (from June 1 to August 31), yielding 41,698,080 data records. These represent data records for three years, labelled as $(T - 1, T, T + 1)$, where year $T - 1$ corresponds to the year when the electricity has a fixed price throughout the day, and the new prices are used in year T and $T + 1$.

3.2 Features for regression models

To establish our baseline, we need to first determine the features that this model depends on. From information in the literature and our exploration of the dataset, we choose 8 features: 3 time variables (month, hour, and day of week), 2 historical electricity usage data (electricity usage of the same hours on a day before (yesterday) and a week before), and 3 hourly averaged weather conditions (temperature, atmospheric pressure, and dew point). The role of the historical usage data is to distinguish each household from others. Here, the weather data vary only over time, not across households, since all households belong to the same geographical region. Although some weather data such as the atmospheric pressure and the dew point do not seem to play major roles at first glance, we also want to take them into account to see whether there is a latent correlation between these data and electricity usage.

3.3 Overview of the data

Fig. 1 shows the average daily electricity usages of 6 different groups over three summer seasons. The data from each of the three years are plotted as a separate line. We note that even though different pricing schemes are used, the impact of the pricing schemes is not obvious. This can be partially explained by Fig. 2, where average hour temperatures and electricity usages are plotted against hour.

In Fig. 2, the temperatures of T and $T + 1$ are higher than the temperature of $T - 1$, which means households have experienced hotter summers in T and $T + 1$. As a result, the electricity usage increases in T and $T + 1$. Even though the new pricing schemes are designed to reduce electricity usage, but the increases in temperature complicates the analysis. Furthermore, the impact of temperature on electricity usage does not appear to be instantaneous; but its impact on electricity usage appears a few hours later. The increased electricity usage during the summer afternoon is mostly from airconditioning, which is more directly related to the indoor temperature, while the temperature reported in our dataset is outdoor temperature. It takes time for the increased outdoor temperature to impact the indoor temperature. Additionally, residents of a house typically return from work in late afternoon, which increase the number of occupants in a household.

The difficulties of identifying how 6 different groups behave differently from Figs. 1 and 2 necessitate a new prediction model for the baseline electricity usage. To this end, we compare various methods in Section 4.

4 Methodology

As we have explained before, the control group does not appear to accurately reflect the 'business-as-usual' in this study of the residential electricity usage, therefore, it is useful to consider alternative methods to extract a baseline. In this section we give a brief introduction of three different statistical models for this baseline: linear regression, gradient linear boosting, and gradient tree boosting.

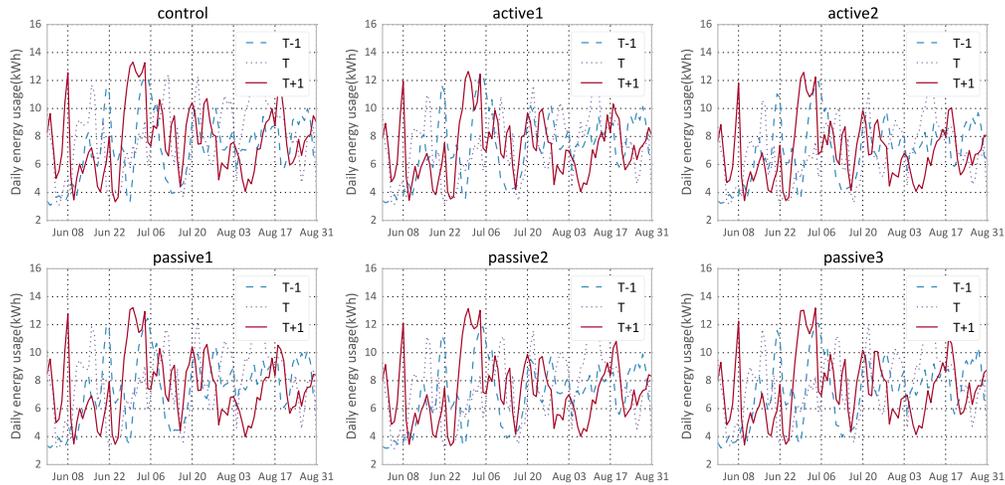


Figure 1: Daily electricity usages of 6 groups for $(T - 1, T, T + 1)$. Note that the effectiveness of differing pricing policies is not immediately visible.

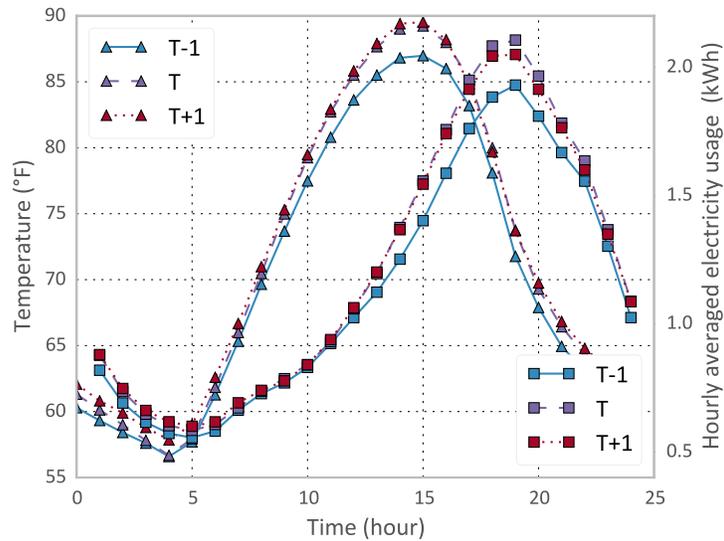


Figure 2: Hourly temperatures (triangle markers) and electricity usages (square markers) for $(T - 1, T, T + 1)$. Note the time lags between the peaks of temperatures and the peaks of electricity usages, which should be taken into consideration when we express a baseline usage model with outdoor temperatures. The temperatures of T and $T + 1$ are higher than that of $T - 1$, which results in the higher electricity usages in T and $T + 1$.

4.1 Linear Regression

One of popular and simple regression models is the linear regression (LR) where a model is represented in the form of linear equations. Multiple LRs can be used to forecast electricity consumption of households [2].

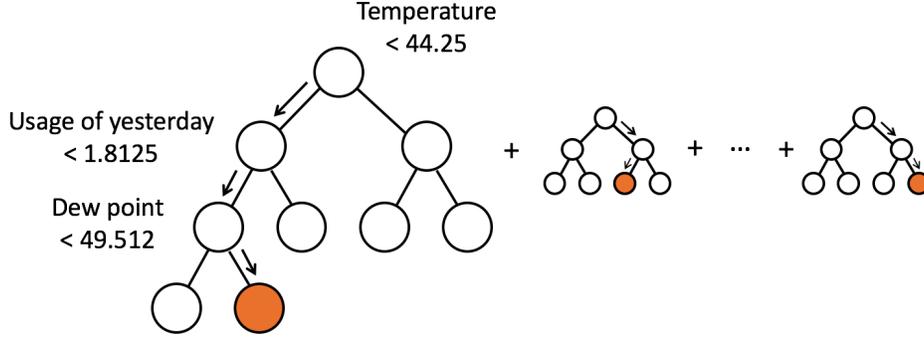


Figure 3: An example of Gradient Tree Boosting (GTB) model. The directed arrow represents a possible path of a sample during the test. Each decision tree decides which path a sample should traverse. Values of leaf nodes are summed to get the prediction.

Given a data set $\{y_i, x_{i,1}, \dots, x_{i,K}\}_{i=1}^n$ of n statistical units, an LR can be represented as follows:

$$\hat{y}_i = \epsilon + \sum_{k=1}^K \beta_k x_{i,k} \quad (1)$$

where \hat{y}_i is an estimated value of y_i , β_k is a k th regression coefficient of $x_{i,k}$, and ϵ is a bias.

4.2 Gradient Linear Boosting and Gradient Tree Boosting

Boosting is a prediction algorithm derived from machine learning literature based on the idea of combining a set of weak learners to create a single strong learner. The boosting method has attracted much attention due to its performance on various applications in both machine learning and statistics literature [26, 15, 27].

Gradient Boosting (GB) is one of the boosting methods which constructs an additive regression model by sequentially training weak learners in the gradient descent viewpoint [16]. GB can be further distinguished by choosing different weak learners. Here we choose two different weak learners: linear function and decision tree. Each model is called Gradient Linear Boosting (GLB) and Gradient Tree Boosting (GTB) respectively.* Fig. 3 shows an example of binary decision trees where each arrow shows a possible path of a sample during testing.

In general, GB can be represented as follows:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F}, \quad (2)$$

where K is the number of weak learners, f_k is a function (linear function or decision tree) in the functional space \mathcal{F} which is the set of all possible regression functions, x_i is an input value from a training set, and \hat{y}_i is the estimation of an output value y_i from the training set.

The objective of GB is to minimize the following objective function $obj(\cdot)$ of Θ which denotes the parameters of GB:

$$obj(\Theta) = L(\Theta) + \sum_{k=1}^K \Omega(f_k), \quad (3)$$

*XGBoost library (<https://github.com/dmlc/xgboost>) is used in this paper.

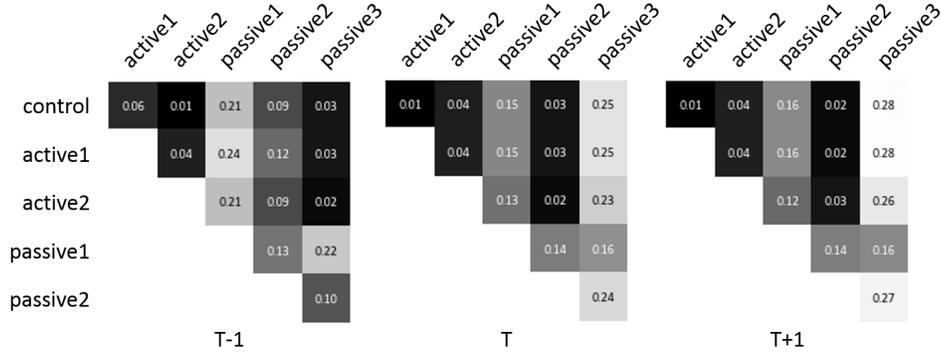


Figure 4: K-S tests for three years. Darker colors denote values close to 0 (small difference), while lighter colors denote values close to 1 (large difference). Note that **passive3** shows large differences in T and $T + 1$.

where $L(\cdot)$ is a training loss function, $\Omega(\cdot)$ is a regularization term. Specifically, we use the root-mean-square error (RMSE) as the training loss function $L(\cdot)$ which is written as:

$$L(\Theta) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad (4)$$

where n is the number of elements in the training set. We employ hourly training datasets (x_i, y_i) for experiments.

5 Experimental Results

5.1 Kolmogorov-Smirnov Test

Before we extract the baseline electricity usage models, we can first examine how each group is similar to/distinguishable from other groups. Even though we separate households according to the criteria explained in Section 3, we want to confirm that different groups can be distinguished from each other in a statistical sense.

To achieve this goal, we use the Kolmogorov-Smirnov test (K-S test), which tests whether two different samples are drawn from the same distribution [8]. In particular, the K-S test quantifies a distance between the cumulative distribution functions (CDFs) of two sample distributions. When comparing two samples with their (empirical) CDFs $F(x)$ and $G(x)$, the K-S test is defined as:

$$D = \sup_x |F(x) - G(x)|, \quad (5)$$

where \sup is the supremum of the set of distances. Therefore, if two CDFs are the same, a K-S test value D becomes 0 because their distance is zero. ($0 \leq D \leq 1$) Since the K-S test only checks whether samples from two different groups are drawn from the same distribution, the sequential characteristic of samples is ignored.

We use combinations of two groups from six different groups as two sets of samples. Fig. 4 shows three K-S tests for each year, where the color of a value is darker if the value is close to 0 (small difference) and lighter if the value is close to 1 (large difference). It should be noted that the same flat rate was adopted across different groups in $T - 1$, whereas different pricing policies were adopted in T and $T + 1$.

Table 1: Yearly Averaged All Hour Usage for Six Groups, Usage Increments in T and $T + 1$, and Their Ranks in Decreasing Order

	D_{T-1}	D_T	D_{T+1}	$D_T - D_{T-1}$	$D_{T+1} - D_{T-1}$
control	1.128	1.205	1.197	0.078 (1)	0.070 (1)
active1	1.136	1.163	1.161	0.027 (6)	0.025 (5)
active2	1.125	1.160	1.173	0.035 (5)	0.048 (4)
passive1	1.157	1.206	1.181	0.049 (3)	0.023 (6)
passive2	1.100	1.152	1.154	0.051 (2)	0.054 (2)
passive3	1.174	1.216	1.228	0.042 (4)	0.053 (3)

Table 2: Yearly Averaged Peak Hour Usage for Six Groups, Usage Increments in T and $T + 1$, and Their Ranks in Decreasing Order

	P_{T-1}	P_T	P_{T+1}	$P_T - P_{T-1}$	$P_{T+1} - P_{T-1}$
control	1.790	1.973	1.937	0.183 (1)	0.147 (1)
active1	1.805	1.796	1.806	-0.009 (5)	0.001 (5)
active2	1.752	1.696	1.739	-0.056 (6)	-0.013 (6)
passive1	1.853	1.952	1.877	0.098 (2)	0.024 (4)
passive2	1.742	1.822	1.818	0.080 (3)	0.076 (2)
passive3	1.809	1.870	1.854	0.061 (4)	0.046 (3)

In Fig. 4, **passive3** shows large differences when compared to other groups in T and $T + 1$. In addition, **passive1** also shows large differences in $T - 1$. From these results, we could say that the actively participating groups are intrinsically different from the passively participating groups. However, the K-S test has limitations that it is difficult to gain further insight, mostly because the sequential characteristic of data is totally ignored.

5.2 Yearly Averaged Usage Analysis

Tables 1 and 2 contain the statistics of electricity consumption for all hour and peak hour electricity usage data. D_t is daily averaged electricity usage averaged by each group in year t and P_t is daily peak hour averaged electricity usage averaged by each group in year t . As we have seen in Fig. 2, most of the groups use more electricity in T and $T + 1$ than in $T - 1$, due to their higher temperatures.

In Table 1, $D_T - D_{T-1}$ and $D_{T+1} - D_{T-1}$ show how daily averaged electricity usage increases from $T - 1$ to T and $T + 1$. The number in the parenthesis is a rank of the value among usage increments of six groups in the same year. Here, all six groups show usage increments in T and $T + 1$. Similar trends can be identified in Table 2 as well. However, the active groups more aggressively reduce electricity consumption and even use less electricity than $T - 1$. Note that the active groups (**active1** and **active2**) indeed reduce Peak Hour electricity usage in T and $T + 1$, as compared to the control and the passive groups (**passive1**, **passive2**, and **passive3**).

Table 3: RMSE for Three Different Models: Linear Regression (LR), Gradient Linear Boosting (GLB), And Gradient Tree Boosting (GTB).

	LR	GLB	GTB
control	1.841	0.952	0.845
active1	1.821	0.983	0.863
active2	1.731	0.957	0.839
passive1	1.938	0.972	0.848
passive2	1.862	0.951	0.838
passive3	2.350	0.982	0.853

Even though these results provide us with understanding of the effectiveness of differing pricing policies, it is difficult to immediately distinguish their impacts, which is due to the increased temperatures in T and $T + 1$. We want to find a way to *normalize* these results in order to identify the effectiveness of differing pricing policies. With the baseline model, we can get rid of effects from features explained in Section 3.2, especially the temperature.

5.3 Model Comparison

We explore three different models: LR, GLB, and GTB, described in Section 4, and plan to choose a single model that best represents the core behavior. Specifically, we trained the three models with the usage data in $T - 1$ by random sampling 70% of data as a training set and using the remaining 30% of data as a test set. In the case of GLB and GTB, we trained 1,000 decision trees for a single GTB. If the sum of child nodes' weights was less than 2, we kept partitioning a tree before the max depth of tree surpassed 5. For each step, we randomly collected half of the data set and shrinker the feature weights to 0.3 so as to avoid over fitting. These parameters were provided by XGBoost package and we tuned hyper parameters by using 5-fold cross-validation with a grid-search method in the parameter spaces.

Table 3 shows the result of RMSE for the three models. We see that the errors of LR, GLB are larger than GTB. This is not unexpected since the relationship between electricity usage and temperature is not only non-linear but also delayed. In this work, we choose GTB to extract the baseline.

5.4 Training Gradient Tree Boosting

Our goal is to predict residential electricity consumption model that captures the effect of outdoor temperature, including its delayed effect. To achieve this goal, we trained a GTB model with the usage data of $T - 1$ for all households regardless of different groups they belong to (policy-agnostic way). Again we randomly sampled 70% of data as a training set and used the remaining 30% of data as a test set. Fig. 5 shows RMSE and residual of six different groups for three years, which was calculated with the test set using the trained GTB. RMSE was calculated by (4) and the residual was the averaged sum of error.

Similarly to Tables 1 and 2, we present predicted usage for all hours and peak hours in Tables 4 and 5 using the baseline model. We also calculate the difference between actual usage and predicted usage during all hours and peak hours in Tables 4 and 5 in T and $T + 1$, in order to see how differing pricing schemes affect each group. The ranks of differences are also provided in parentheses.

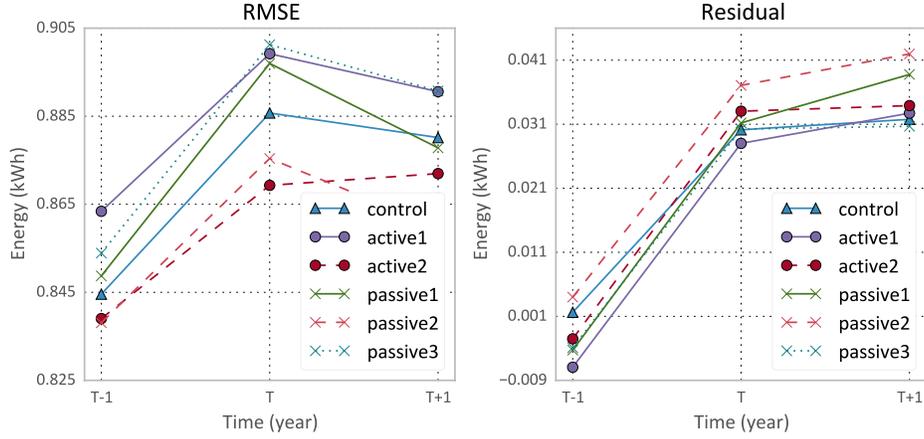


Figure 5: RMSE and residual of six groups for three years by GTB.

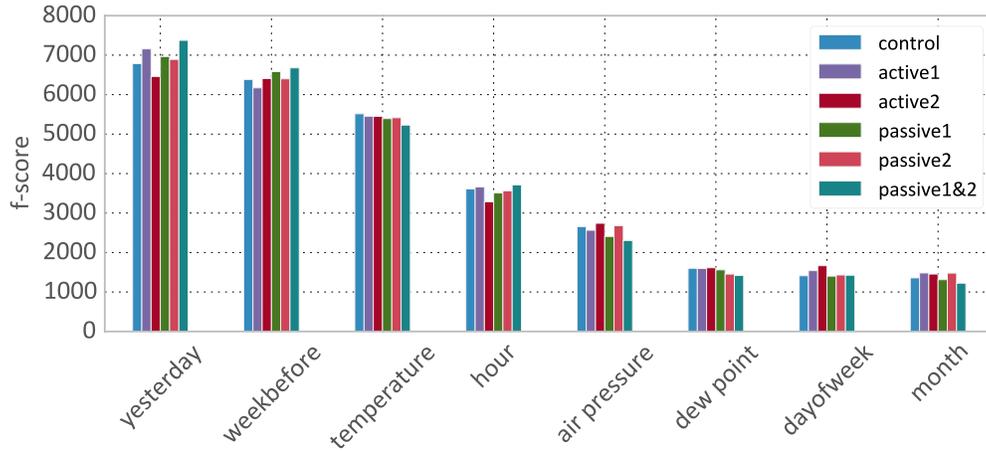


Figure 6: F-score representing the importance of a feature in the decision trees of GBT, which is calculated by counting the appearance of a feature.

In Tables 4 and 5, most groups reduce their electricity usage in T and $T + 1$, as compared to the baseline model prediction. Among six groups, the active groups (**active1** and **active2**) reduce the most, which is expected behavior. The results in Tables 4 and 5 are compelling, since the effect of temperature is now controlled. Therefore, we can identify the temperature-adjusted electricity usage of each group.

Fig. 6 shows f-score of each feature in GTB, where the f-score is the number of appearances of a feature in all of weak decision trees in GTB. If the f-score of one feature is higher, the feature is more important than other features. The two most powerful features are historical electricity usage data (yesterday and week before usage) and the third most influential feature is temperature. In Fig. 6, we can see how GTB finds which features are important. It is also interesting to note that ‘day of week’ is not that effective as other features, because we originally assumed that GTB might detect the difference between weekend and weekday from the dataset.

Table 4: Yearly Averaged All Hour Prediction for Six Groups, Difference Between Actual and Predicted Usage in T and $T + 1$, and Their Ranks in Decreasing Order

group	E_{T-1}	E_T	E_{T+1}	$D_T - E_T$	$D_{T+1} - E_{T+1}$
control	1.129	1.235	1.229	-0.030 (1)	-0.032 (3)
active1	1.165	1.258	1.248	-0.095 (6)	-0.087 (6)
active2	1.139	1.228	1.219	-0.069 (5)	-0.046 (5)
passive1	1.153	1.237	1.219	-0.031 (3)	-0.039 (4)
passive2	1.118	1.212	1.218	-0.061 (4)	0.064 (1)
passive3	1.170	1.247	1.258	-0.030 (1)	-0.031 (2)

Table 5: Yearly Averaged Peak Hour Prediction for Six Groups, Difference Between Actual and Predicted Usage in T and $T + 1$, and Their Ranks in Decreasing Order

group	Q_{T-1}	Q_T	Q_{T+1}	$P_T - Q_T$	$P_{T+1} - Q_{T+1}$
control	1.742	1.947	1.986	0.026 (1)	-0.049 (1)
active1	1.756	1.821	1.888	-0.025 (5)	-0.082 (5)
active2	1.722	1.752	1.847	-0.056 (6)	-0.108 (6)
passive1	1.783	1.937	1.956	0.015 (2)	-0.079 (4)
passive2	1.706	1.837	1.908	-0.014 (4)	-0.070 (2)
passive3	1.755	1.879	1.929	-0.010 (3)	-0.075 (3)

5.5 Hourly Averaged Prediction

Fig. 7 shows the hourly usage prediction and hourly averaged temperature of different groups for three years. In year T and $T + 1$, we can see the flat tops of electricity usage curves during peak hours which indicates that the users have curbed their electricity usage as the new pricing schemes intended. Even though this reduction in peak-hour electricity usage is expected for all new pricing schemes, we only observed this flat top in group **active2**. We can also see that the GTB model effectively has learned the lagged effect of temperature explained in Fig. 2.

Fig. 8 shows how the predicted electricity usage of three models described in Section 4 and the actual usage are different by year and group. Here we can clearly see a big gap between actual usage and predicted usage. The prediction using GTB shows the most accurate hourly prediction, while LR shows the least accurate hourly prediction. Note that the electricity usage curve of a randomly selected household is not as smooth as the actual usage curve averaged over each group shown in Fig. 7.

6 Summary and Future Work

After observing the shortcomings of randomized control group in the residential electricity usage data, we propose to explore a new type of baseline extracted through machine learning methods. In this process, we explored a linear regression method and two variants of Gradient Boosting techniques. Instead of providing

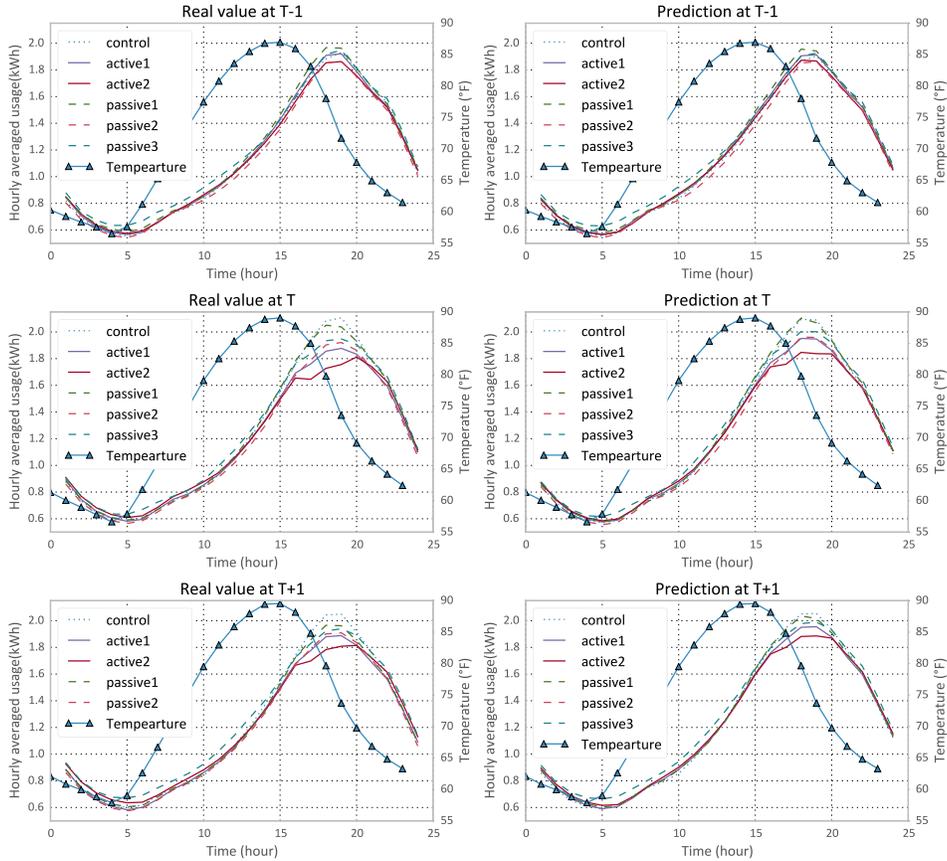


Figure 7: Hourly averaged actual usage is shown on the left. And hourly averaged predicted usage is shown on the right. Triangles markers show the averaged temperature. As presented in Tables 4 and 5, the predicted usage shows higher values than the actual usage, demonstrating that differing pricing policies affect household usage patterns.

accurate short-term forecasts, our baseline model aims to capture intraday characteristics that persists for years. Our tests show that one of the boosting technique, GTB, could incorporate important features such as outdoor temperature and capture the core user behavior. For example, the baseline model from GTB accurately reproduces the lag between the daily peak temperature and peak electricity usage.

The ultimate objective of our work is to evaluate the effectiveness of the different pricing schemes. The new baseline is an important component. This preliminary work demonstrate that new approach is promising, but additional work is needed to evaluate the effectiveness of this approach. For example, we should to re-evaluate the features used in the regression models and systematically measure their impact.

Acknowledgment

The authors gratefully acknowledge the helpful discussions with Sam Borgeson, Daniel Fredman, Liesel Hans, Ling Jin, and Sid Patel.

This work is supported in part by the Director, Office of Laboratory Policy and Infrastructure Man-

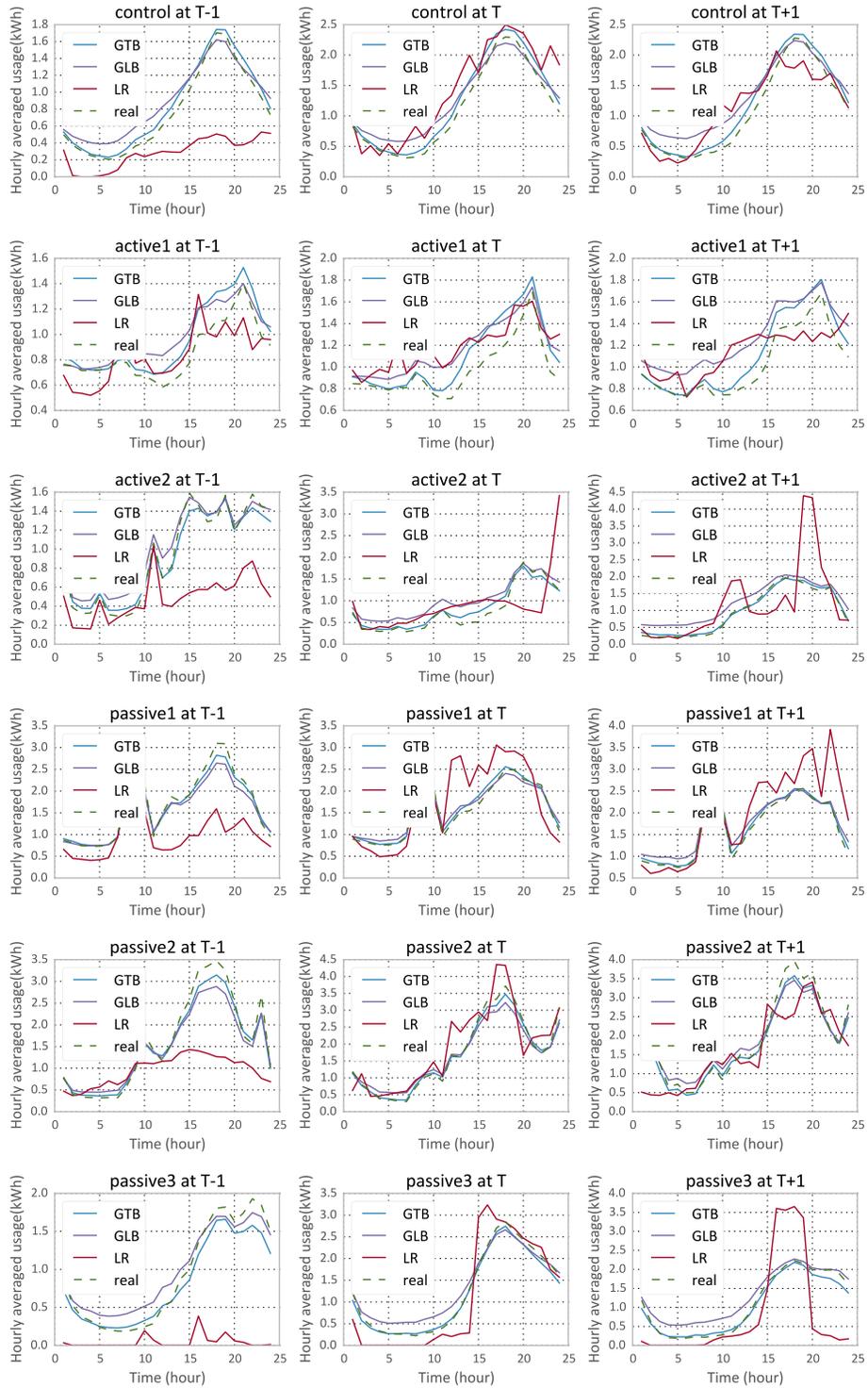


Figure 8: Predicted electricity usage (solid line) of the Linear Regression (LR), Gradient Linear Boosting (GLB), and Gradient Tree Boosting (GTB); and hourly averaged actual usage (dashed line) of a randomly selected household from each group.

agement of the U.S. Department of Energy under contract No. DE-AC02-05CH11231. This work is also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science, ICT & Future Planning (MSIP) (NRF-2014R1A1A1002662) and the NRF grant funded by the MSIP (NRF-2014M2A8A2074096).

References

- [1] Hesham K. Alfares and Mohammad Nazeeruddin. Electric load forecasting: Literature survey and classification of methods. *International Journal of Systems Science*, 33(1):23–34, January 2002.
- [2] Vincenzo Bianco, Oronzio Manca, and Sergio Nardini. Electricity consumption forecasting in Italy using linear regression models. *Energy*, 34(9):1413–1421, 2009.
- [3] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi. A bottom-up approach to residential load modeling. *IEEE Transactions on Power Systems*, 9(2):957–964, May 1994.
- [4] Peter Cappers, Annika Todd, Michael Perry, Bernard Neenan, and Richard Boisvert. Quantifying the impacts of time-based rates, enabling technology, and other treatments in consumer behavior studies: Protocols and guidelines. Technical Report LBNL-6301E, Lawrence Berkeley National Laboratory, 2013.
- [5] Bo-Juen Chen, Ming-Wei Chang, and Chih-Jen Lin. Load forecasting using support vector machines: A study on EUNITE competition 2001. *IEEE Transactions on Power Systems*, 19(4):1821–1830, 2004.
- [6] Gianfranco Chicco, Roberto Napoli, and Federico Piglione. Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, 21(2):933–940, 2006.
- [7] Gianfranco Chicco, Roberto Napoli, Federico Piglione, Petru Postolache, Mircea Scutariu, and Cornel Toader. Load pattern-based classification of electricity customers. *IEEE Transactions on Power Systems*, 19(2):1232–1239, 2004.
- [8] William Jay Conover and WJ Conover. *Practical nonparametric statistics*. Wiley New York, 1980.
- [9] Dora L. Costa and Matthew E. Kahn. Energy conservation ”nudges” and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702, June 2013.
- [10] Remy Cottet and Michael Smith. Bayesian modeling and forecasting of intraday electricity load. *Journal of the American Statistical Association*, 98(464):839–849, 2003.
- [11] James A. Espey and Molly Espey. Turning on the lights: A meta-analysis of residential electricity demand elasticities. *Journal of Agricultural and Applied Economics*, 36:65–81, 4 2004.
- [12] Marcelo Espinoza, Caroline Joye, Ronnie Belmans, and Bart De Moor. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. *IEEE Transactions on Power Systems*, 20(3):1622–1630, 2005.
- [13] Vera Figueiredo, Fátima Rodrigues, Zita Vale, and Joaquim Borges Gouveia. An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems*, 20(2):596–602, 2005.

- [14] Dipl-Wi-Ing Christoph Flath, Dipl-Wi-Ing David Nicolay, Tobias Conte, PD Dr Clemens van Dinther, and Lilia Filipova-Neumann. Cluster analysis of smart metering data. *Business & Information Systems Engineering*, 4(1):31–39, 2012.
- [15] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In *Proceedings of International Conference on Machine Learning*, volume 96, pages 148–156, 1996.
- [16] Jerome H Friedman. Greedy function approximation: A gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [17] Samuel Humeau, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer. Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households. In *Proceedings of Sustainable Internet and ICT for Sustainability*, pages 1–6. IEEE, 2013.
- [18] GW Irwin, W Monteith, and WC Beattie. Statistical electricity demand modelling from consumer billing data. In *IEE Proceedings C (Generation, Transmission and Distribution)*, volume 133, pages 328–335. IET, 1986.
- [19] Paul L Joskow. California’s electricity crisis. *Oxford Review of Economic Policy*, 17(3):365–388, 2001.
- [20] Jungsuk Kwac, June Flora, and Ram Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1):420–430, 2014.
- [21] Jungsuk Kwac and Ram Rajagopal. Demand response targeting using big data analytics. In *Proceedings of IEEE International Conference on Big Data*, pages 683–690. IEEE, 2013.
- [22] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17(2):342–348, May 2002.
- [23] F. Oldewurtel, A. Ulbig, A. Parisio, G. Andersson, and M. Morari. Reducing peak electricity demand in building climate control using real-time pricing and model predictive control. In *Proceeding of IEEE Conference on Decision and Control*, pages 1927–1932, December 2010.
- [24] Anastasios Panagiotelis and Michael Smith. Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. *International Journal of Forecasting*, 24(4):710–727, 2008.
- [25] Sarvapali D. Ramchurn, Perukrishnen Vytelingum, Alex Rogers, and Nicholas R. Jennings. Putting the ‘smarts’ into the smart grid: A grand challenge for artificial intelligence. *Communications of the ACM*, 55(4):86–97, April 2012.
- [26] Robert E Schapire. The strength of weak learnability. *Machine learning*, 5(2):197–227, 1990.
- [27] Robert E Schapire and Yoav Freund. *Boosting: Foundations and algorithms*. MIT press, 2012.
- [28] Brian Artur Smith, Jeffrey Wong, and Ram Rajagopal. A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting. In *Proceedings of ACEEE Summer Study on Energy Efficiency in Buildings*, 2012.
- [29] L. Suganthi and Anand A. Samuel. Energy models for demand forecasting – a review. *Renewable and Sustainable Energy Reviews*, 16(2):1223–1240, 2012.

- [30] Lukas G. Swan and V. Ismet Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8):1819–1835, 2009.
- [31] J. W. Taylor and P. E. McSharry. Short-term load forecasting methods: An evaluation based on European data. *IEEE Transactions on Power Systems*, 22(4):2213–2219, November 2007.
- [32] James W. Taylor. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1):139–152, 2010.
- [33] Annika Todd, Michael Perry, Brian Smith, Michael J. Sullivan, Peter Cappers, and Charles A. Goldman. Insights from smart meters: The potential for peak hour savings from behavior-based programs. Technical Report LBNL-6598E, Lawrence Berkeley National Laboratory, 2014.
- [34] Annika Todd, Elizabeth Stuart, Steven R. Schiller, and Charles A. Goldman. Evaluation, measurement, and verification (EM&V) of residential behavior-based energy efficiency programs: Issues and recommendations. Technical Report DOE/EE-0734, US Department of Energy, 2012.
- [35] George J Tsekouras, Nikos D Hatziargyriou, and Evangelos N Dialynas. Two-stage pattern recognition of load curves for classification of electricity customers. *IEEE Transactions on Power Systems*, 22(3):1120–1128, 2007.
- [36] Sergio Valero Verdú, Mario Ortiz Garcia, Carolina Senabre, Antonio Gabaldón Marin, and Francisco J García Franco. Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. *IEEE Transactions on Power Systems*, 21(4):1672–1682, 2006.
- [37] Frank A Wolak. Diagnosing the California electricity crisis. *The Electricity Journal*, 16(7):11–37, 2003.