

# Electricity Demand Simulations on the Distribution Edge: Developing a Granular Representation of End-User Preferences using Smart Meter Data

Ashwini Bharatkumar<sup>1</sup>, Ricardo Esparza<sup>2</sup>, Kristina Mohlin<sup>\*2</sup>, Elisheba Spiller<sup>2</sup>,  
Karen Tapia-Ahumada<sup>1</sup> and Burcin Unel<sup>†3</sup>

<sup>1</sup>*Massachusetts Institute of Technology Energy Initiative*

<sup>2</sup>*Environmental Defense Fund*

<sup>3</sup>*Institute for Policy Integrity at New York University School of Law*

June 7, 2019

## Abstract

We combine insights from engineering and economics literature by incorporating microeconomic theory on consumer preferences into an existing electricity simulation model to provide an improved representation of residential customers' electricity consumption preferences. The resulting model can be used to evaluate and compare residential end-users' responses to electric rate designs and their decisions to invest in and operate distributed energy resources (DERs). In order to represent how end-users are likely to respond to different rate designs, we model residential end-users' preferences for consuming electricity by incorporating a constraint that represents the consumer welfare derived from thermal and non-thermal electricity services throughout the day. We then calibrate this model using advanced meter infrastructure (AMI) data from a large U.S. electric utility. In future research, this model can provide new insights by combining engineering simulation techniques with economic theory and econometric methods using real-world smart grid data.

---

\*Corresponding author: kmohlin@edf.org

†We want to thank Alex Bi and Nelson Lee for excellent research assistance.

## 1 Introduction

The electric distribution grid is transitioning toward a model in which customers can themselves provide a variety of services to the grid, including electricity generation, by investing in distributed energy resources (DERs) such as distributed solar generation, programmable appliances, and energy storage. However, customers' incentives to make these investments depend on how they are being charged for electric service. Specifically, the way the electric distribution company allocates the cost of service into the different elements of the rate (tariff) design, such as volumetric or demand charges and time-variant or flat charges, determines the returns on investment for different types of DERs. The rate design will also be a main factor in determining where and to what extent investment in DERs are made, and whether DERs will contribute to improving system reliability and reducing electric system costs.

Despite the topic's importance for the electric distribution system of the future, the body of literature on the impact of electric rate design on the proliferation of DERs is still limited, see e.g., [Darghouth et al. \[2016\]](#), [Hledik and Greenstein \[2016\]](#), [Schittekatte et al. \[2018\]](#), and [Simshauser \[2016\]](#). While these studies look at important topics such as the potential for cost shifting, they all hold electricity consumption patterns constant, and, hence, do not take into account how customers' use of electricity may shift in response to new electric rate designs. As a result, their approaches are more limited in their ability to capture the impact of rate design on the return on investment for different DERs.<sup>1</sup>

In this paper, we improve upon this common feature of typical engineering models analyzing the impacts of electric rate design by incorporating microeconomic theory into an existing electric simulation model. Engineering simulation models are generally cost minimization problems with ad-hoc monetized penalties for deviations from a reference electricity use profile, and, thus, do not provide a very good representation of consumer

---

<sup>1</sup> A notable exception is [Hledik and Greenstein \[2016\]](#) who estimate the effect of demand charges under smart, or predictive, charging patterns of distributed storage owners and simple charging, and show that the benefit of cost-reflective pricing is larger under smart charging. However, this different charging pattern is not based on observed user preferences or behavior when faced with a different tariff.

preferences and are limited in their ability to predict changes in consumer behavior. In contrast, our model specification provides an improved representation of residential customers' electricity consumption preferences, and can therefore be used to evaluate and compare residential customers' responses to different electric rate designs and their decisions to invest in and operate DERs.

Specifically, we include a consumer welfare constraint (henceforth utility constraint) in the optimization (electric bill minimization) model to represent consumer preferences for non-thermal and thermal electric loads. This representation allows us to separate consumer preferences related to heating and cooling needs, which are weather dependent, and other needs, which depend on individual preferences for appliance usage. We then calibrate the model using AMI data from a large US electric distribution company. Using 30-min data for over 50,000 customers, we calibrate unique preference parameters for each customer. As a result, we can provide synthetic model-generated loads that closely replicates observed loads.

This paper is structured as follows. We first describe the simulation model and how the utility constraint is specified. We next describe how we calibrate the new specification to AMI data. Finally, we present results from a randomly selected subset of users to illustrate how well the model output with the new specification replicates observed electric load patterns.

## **2 Demand Response and Distributed Resources Economic Model (DR-DRE)**

The Demand Response and Distributed Resources Economic Model (DR-DRE), which this research improves upon, was originally developed by engineers at the Massachusetts Institute of Technology (MIT) (see [Huntington \[2016\]](#) for a more complete description of the original specification). DR-DRE is designed to represent how electricity tariffs affect households' electricity consumption and their decisions to invest in and operate DERs. The model can represent a large number of different households/users such that aggregating the DR-DRE model output across a large set of users gives results for the impact of an electricity tariff on

aggregate (net) load.

DR-DRE’s available investment options in DERs are rooftop photovoltaics (PV) systems, batteries and heat pumps. The model’s representative customers choose their electricity consumption and investment in DERs to minimize their net electricity expenditure (i.e., any expenditure minus any DER-related revenues), subject to customer-specific constraints related to preferences for indoor temperature and load shifting as well as constraints related to heating, ventilation, and air conditioning (HVAC) technology and the building’s thermal characteristics. The customers’ decisions to conserve or shift electricity consumption, and whether to adopt any DER therefore depend on the relative prices and the associated welfare losses (such as deviations from the ideal indoor temperature) from changing their electricity consumption across the hours of the day.

DR-DRE can represent a variety of rate designs which can include time-varying energy (\$/kWh), capacity (\$/kW), and fixed charge components, or any combination thereof. The model can also simulate the provision of different services from DERs, such as energy, operating reserves, firm capacity, and network services.

The DR-DRE model is proprietary and a mixed integer linear program written in Julia/JuMP, using a nonlinear solver.

### **3 Updating the Consumer Preference Specification in DR-DRE**

In the original DR-DRE version, households’ cost-minimizing responses to electricity rate and structure were determined by two disutility parameters (i.e., penalties) for thermal and non-thermal load responses, respectively. For thermal load, this disutility parameter was a monetized disutility for indoor temperature deviations outside of a set range around the ideal indoor temperature (bliss point). This results in an end-user optimization around cost and temperature. For example, during summer hours when the cost of electricity for cooling is higher than the monetized disutility from having a higher indoor temperature, the model provides a cost-minimizing solution with higher indoor temperature and lower thermal load but with an additional disutility cost tied to the resulting hotter indoor temperature.

Similarly, for non-thermal load, in the original DR-DRE specification, there was a parameter which represented the monetized disutility of curtailment of non-thermal electricity use. However, the values of these parameters were not based on empirical studies or observed user preferences, and this specification furthermore did not allow for a difference in the utility/preference for electricity load during different parts of the day.

We therefore replaced the constraints related to these parameters with a household utility constraint (1) as specified below:

$$Utility \geq \bar{u}, \tag{1}$$

where

$$Utility = \prod_{t=1}^T (NonThermalLoad(t) - MinLoad(t))^{a_t} - \sum_{t=1}^T b(TempInt(t) - BlissPoint)^2 \tag{2}$$

and where  $0 < a_t < 1$  and  $t = [1, 24]$  represents the hour of the day. *MinLoad* is the absolute minimum non-thermal electricity use in each hour, *TempInt*( $t$ ) is the indoor temperature in hour  $t$  as generated by the HVAC system and *BlissPoint* is the most comfortable indoor temperature which would be chosen if the electricity costs of running the HVAC system were not a consideration.  $\bar{u}$  is the minimum level of utility (household welfare) that the cost-minimizing solution needs to achieve.

The first part of the utility function represents each user's relative preference for electricity services other than heating or cooling (non-thermal loads) during each hour. This part of the utility function is an adapted version of the Klein-Rubin (Stone-Geary) utility function.<sup>2</sup> This formulation features a necessary or minimum consumption of each good (here represented by the parameter *MinLoad*). As a result, this specification better reflects the choice problem facing an electricity customer considering substitution of electricity consumption in one period with consumption in another period based on the relative prices across time periods. The relative size of parameters  $a_t$ , in turn, indicates the relative utility of non-thermal

---

<sup>2</sup>The Klein-Rubin (Stone-Geary) utility function has been widely used in the study of private consumption patterns, see e.g., Gaudin et al. [2001], and derives from the seminal studies by Klein and Rubin [1947], Geary [1950] and Stone and Rowe [1954]

electricity use in different hours of the day.

The second part of the utility function models each user’s relative preference for meeting their heating and cooling demand (thermal loads) during each hour by representing the disutility from any deviation between the indoor temperature provided by the user’s HVAC system and the end-user’s ideal indoor temperature (bliss point) as a convex function.<sup>3</sup> The squared term implies an increasing marginal disutility of further temperature deviations away from the bliss point.

With this specification, substitution and shifting of loads amongst hours and across uses (thermal versus non-thermal) are better captured, while avoiding the type of ad-hoc monetary disutility penalty featured in the original specification.

#### 4 Calibration against AMI Data

DR-DRE previously relied on simulated, representative residential load profiles based on average thermal and non-thermal load patterns. Through access to a very large and granular dataset of customer-level AMI load data from Commonwealth Edison (ComEd) in Chicago, we are able to represent a much more varied set of preferences and calibrate them to observed load data rather than simulated load profiles. This representation of a large set of different end-user preferences based on real load data enables us to better capture the variation in household responses that can be expected under different rate designs. However, our data show only the total household-level consumption, and do not differentiate between different end uses of electricity, such as heating and other needs. As a result, we need to estimate the fraction of each end-user’s load that is used for thermal and non-thermal purposes. Thus, as further described below, we first use regression analysis to estimate the thermal and non-thermal portions of the total household electric load. Next, we use the estimated thermal and non-thermal electric loads to calibrate the parameters of the utility function and the building thermal properties in DR-DRE.

---

<sup>3</sup> We assume a bliss point of 21 degrees Celsius.

## 4.1 Econometric estimation of thermal and non-thermal loads

We first estimate hourly thermal load for each household by regressing the total household load on outdoor temperature to capture each household’s electricity use for indoor space heating and cooling. In the next step, we calculate hourly non-thermal loads as the difference between the observed total loads and the estimated thermal loads.

### 4.1.1 Data

The dependent variable is customer-level AMI load data for 55,635 households with 30-min resolution for the year 2016 from ComEd in Chicago. We aggregate the data up to hourly loads resulting in 8,784 data points per household (2016 was a leap year). We eliminate commercial users and users with a large number of missing data, which reduces our initial sample of 55,635 to 54,412.

Table 1 below shows the customer class and tariffs of the 54,412 households sample. In 2016, ComEd introduced a Residential Real Time Pricing (now more accurately re-name Hourly Pricing) program, however, since participation has been low (less than 0.6% of residential customers with supply service with ComEd), we assume all households in our sample face a flat (time-invariant) volumetric charge per kWh.

Table 1: Characteristics of the 54,412 users sample

Customer class	Frequency	Fixed Charge	Volumetric Charge (per kWh)
Residential Multi	9,040	\$11.98	\$0.107
Residential Multi (Space Heat)	1,076	\$12.46	\$0.095
Residential Single	44,185	\$14.89	\$0.106
Residential Single (Space Heat)	111	\$16.32	\$0.94

Table 2: Distribution of hourly loads (kWh) for the 54,412 users sample

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
0.00	0.25	0.48	0.76	0.92	28.81

Table 3: Distribution of yearly loads (kWh) for the 54,412 users sample

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
10.17	3657.65	5909.93	6638.45	8843.46	61323.89

Tables 2 and 3 show the distribution of the hourly loads and total annual loads, respectively.

For the regression analysis, the main independent variables are outdoor temperature and relative humidity in 2016 at Chicago Midway Airport Climatological Data Station, retrieved from a public dataset by the National Centers for Environmental Information (NCEI) with 60-min resolution.

#### 4.1.2 Econometric specification

Our econometric specification estimates the responsiveness of customers' electricity usage to outdoor temperature. We assume that this response is due to a change in thermal loads; so, for example, during the summer, we attribute a positive correlation between total electricity consumption and hourly outdoor temperature to increased A/C usage. This assumption disregards the fact that it is possible for there to be non-space cooling or space heating reasons for changes in electricity consumption in response to outdoor temperature, such as changes in the use of electric appliances due to weather (either through behavioral change, such as households staying in to avoid high outdoor temperatures, or technological effects, such as a refrigerator having to cycle more due to hotter temperatures). However, because we do not observe appliance-level electricity consumption, this assumption is required. Note also that since the customers in our sample, as noted in the previous section, were facing time-invariant volumetric electricity charges in 2016 there is no need to control for electricity price in the regressions.

To estimate the responsiveness of electricity consumption to changes in outdoor temperature, we run the following two regressions for each individual household separately: one for the hours in which the temperature was above 65 degrees Fahrenheit (F), and another

for the hours in which the temperature was below 65F. This cut-off point of 65F is based on the common assumption that energy consumption is influenced by the need to keep a comfortable indoor temperature - estimated to be 65F - in residential buildings.<sup>4</sup> We are thus attempting to estimate the responsiveness of electricity demand to outdoor temperature due to residential space cooling and heating, respectively. We define this incremental response as thermal load.<sup>5</sup>

Dividing the regressions based on temperature (heating degree hours vs. cooling degree hours) provides two important benefits. First, it lets us avoid having to place unnecessary structure on the regression specification to allow for a U-shaped functional form around 65F. Second, it provides more flexibility in temperature responsiveness, due to differences in behavior, preferences, and consumption patterns during different times of year. This flexibility is important because many individuals in our sample do not have electric heat; behavioral responsiveness such as opening windows and/or doors during summer nights can lead to greater heat transfer losses; and so on. Thus, by allowing the coefficients to vary for heating and cooling preferences, we increase the accuracy of our regression results over the entirety of the temperature range.

Our estimation equations are as follows:

For hours with temperature >65F:

$$L_t = \beta_0 + \beta_{CDH} CDH_t + \beta_{CDH^2} [CDH_t]^2 + \beta_{humidity} Humidity_t + \beta_{weekend} I_{weekend} + \sum_{month \in 2...12} \beta_{month} I_{month} + \sum_{hour \in 2...24} \beta_{hour} I_{hour} + \epsilon_t;$$

and for hours with temperature <65F:

$$L_t = \beta_0 + \beta_{HDH} HDH_t + \beta_{HDH^2} [HDH_t]^2 + \beta_{humidity} Humidity_t + \beta_{weekend} I_{weekend} + \sum_{month \in 2...12} \beta_{month} I_{month} + \sum_{hour \in 2...24} \beta_{hour} I_{hour} + \epsilon_t;$$

<sup>4</sup> This is, for example, how Cooling and Heating Degree Days are defined.

<sup>5</sup> As alluded to earlier, this specification assumes that water heating usage is not a part of our thermal load.

where

- $L_t$  is total household load in hour  $t$ .
- $CDH_t$  is Cooling Degree Hours (CDH) in hour  $t$ . The cooling degree hours are calculated as the difference between the hourly temperature and 65F for each hour in which the temperature is higher than 65F. A square term is added for each cooling degree hour to account for a non linear response due to extreme temperatures.
- $HDH_t$  is Heating Degree Hours (HDH) in hour  $t$ . The heating degree hours are equal to the difference between 65F and the current hourly temperature whenever it is below 65F. A square term is added here as well.
- $Humidity_t$  is the hourly relative humidity, which measures moisture in the air as a percentage of the maximum water vapor possible at a given temperature and pressure. This variable affects individuals' sensation of outdoor temperature.
- $I_{weekend}$  is a dummy variable that indicates whether the given hour belongs to a weekend or weekday. Weekday is the omitted category.
- $I_{month}$  are dummy variables that indicate to which month of year the hour belongs to. January is the omitted category.
- $I_{hour}$  are dummy variables that indicate hour of day. Midnight to 1 am is the omitted category.

We use the regression results to predict the hourly thermal load for each household. We first replace coefficients that were not significant at the 5% level with zeros. That is, we only use statistically significant coefficients to estimate the thermal load. Our regression-predicted thermal loads are thus calculated with the following equations:

For hours with temperature  $>65$  F:

$$\widehat{ThermalLoad}_t = \hat{\beta}_{CDH,t} CDH_t + \hat{\beta}_{CDH^2} [CDH_t]^2,$$

and for hours with temperature <65 F:

$$\widehat{ThermalLoad}_t = \hat{\beta}_{HDH,t} HDH_t + \hat{\beta}_{HDH^2} [HDH_t]^2,$$

where  $\hat{\beta}$  are the user-specific coefficient estimates from the regressions.

### 4.1.3 Econometric Results

In the figures below, we show the regression results for a sub-sample of 50 randomly selected users from the whole sample of 54,412 households (henceforth users). We randomly sampled 50 users to first test out our calibration approach before expanding it for the full sample.

Figures 1 and 2 show the regression-estimated average seasonal thermal load profile across the hours of the day for our sub-sample, while figures 3 and 4 shows the hourly residuals across the year from the regression. These residuals would be equivalent to the non-thermal loads per hour, since we estimate non thermal loads by subtracting thermal loads from the total load.

Figure 1: Average Hourly Regression-Predicted Thermal Load Subsample 1

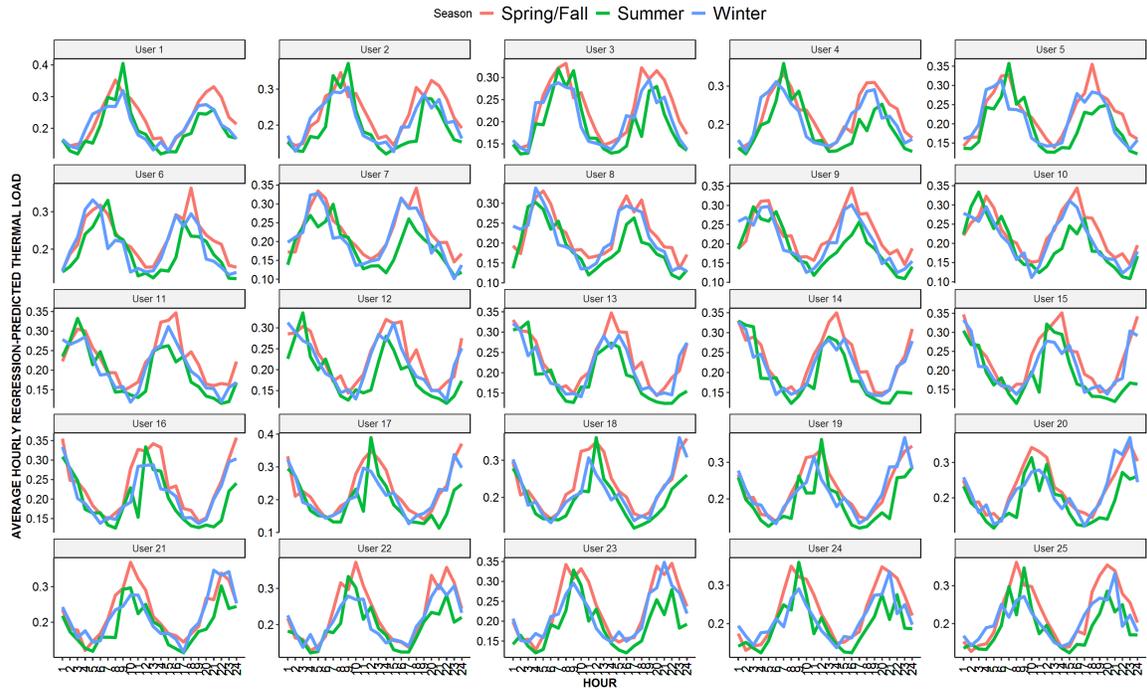


Figure 2: Average Hourly Regression-Predicted Thermal Load Subsample 2

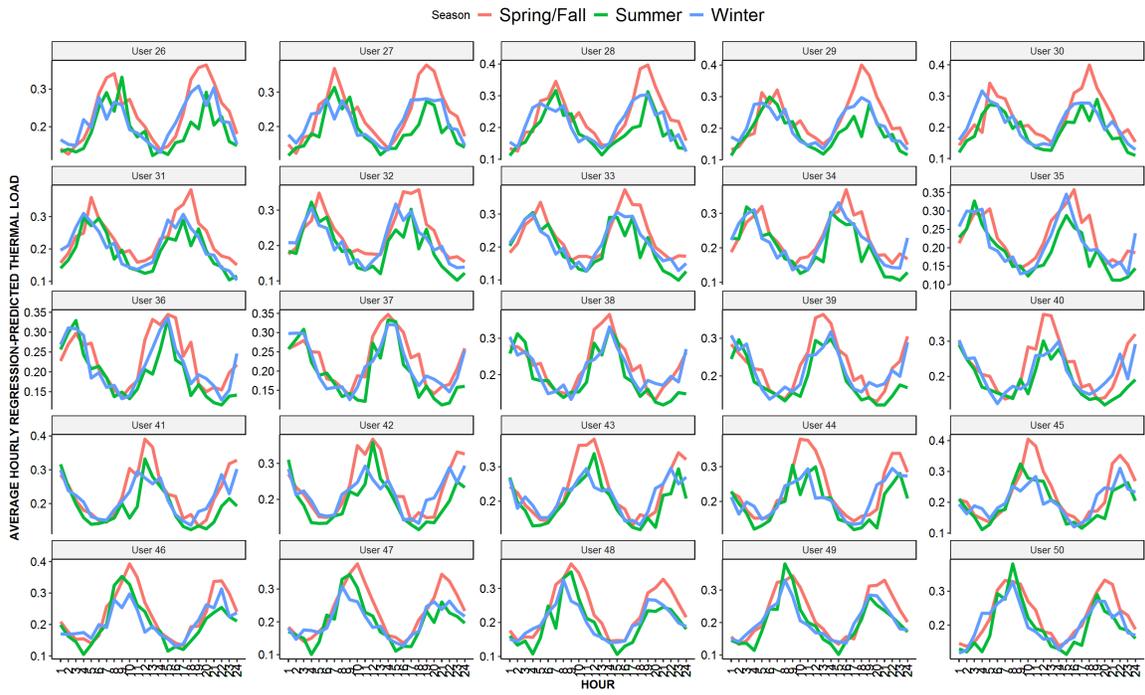


Figure 3: Hourly Residuals Subsample 1

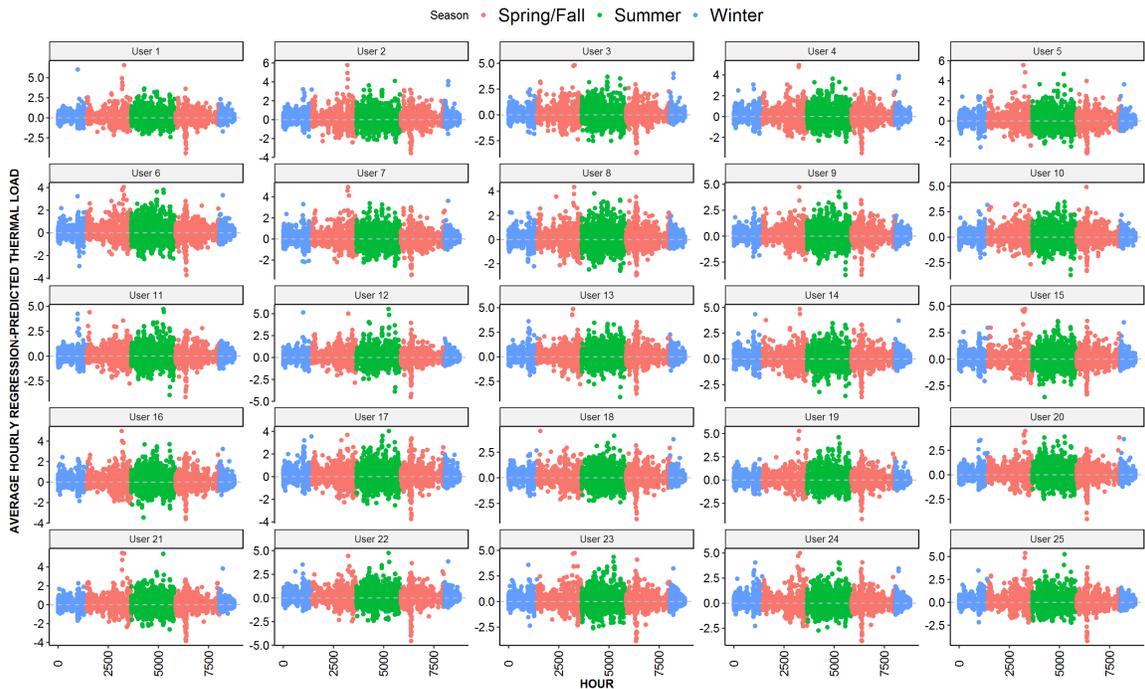
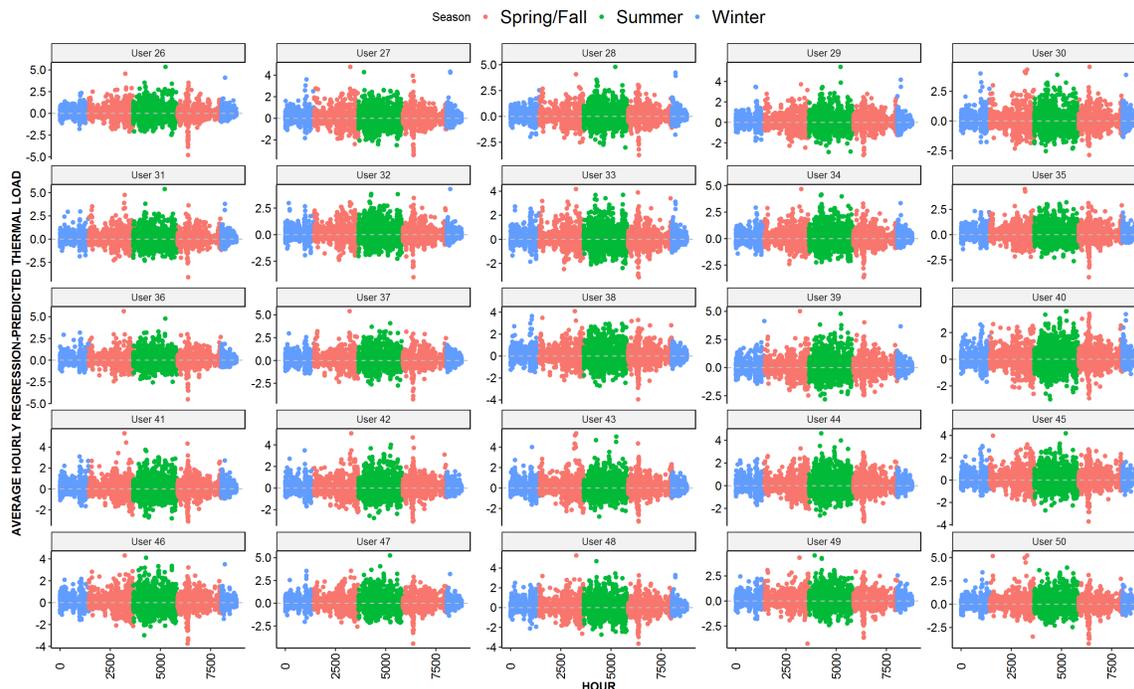


Figure 4: Average Residuals Subsample 2



## 4.2 Calibration using the estimated loads

### 4.2.1 Calibration of household preference parameters determining non-thermal load

The preference parameters we need to calibrate with the AMI data are the  $a_t$  values and the  $\bar{u}$  values from expression (2) for each household. Using the estimated non-thermal load for each hour (which is estimated as the difference between total observed loads and regression-predicted thermal loads), we can approximate the value of the  $a_t$  parameters in (2) for each household and hour by making use of the assumption that the estimated non-thermal load is the load that maximizes (non-thermal) utility subject to a budget constraint.<sup>6</sup>

<sup>6</sup> This approach for calculating the  $a_t$  values additionally requires an assumption that the non-thermal load is determined independently of the thermal load, which is strictly speaking not fulfilled as we have defined our utility function, but is the only assumption that makes it possible for us to approximate the  $a_t$  values without having to run the model and correct for the endogenous indoor temperature and related disutility.

Maximizing the first term of (2) subject to daily electricity expenditure  $m$ , we can solve for  $a_t$  and get:

$$a_t = \frac{\text{price}(t) [\text{NonThermal Load}(t) - \text{MinLoad}]}{m - \sum_{t=1}^T \text{price}(t) \text{MinLoad}} = \frac{\text{NonThermal Load}(t) - \text{MinLoad}}{\sum_{t=1}^T \text{NonThermal Load}(t) - \sum_{t=1}^T \text{MinLoad}} \quad (3)$$

Because the 2016 electricity prices were time-invariant, the  $a_t$  value for each hour is simply given by the share of that hour's non-thermal load in total daily non-thermal load.<sup>7</sup> To arrive at a reasonable number of preference parameters for each household that still represent the household's preference for electricity use in each hour of the day, we calculate 48 average values to represent each hour of the day on weekends and weekdays. For example, we calculate the mean for all the values of  $a_t$  for 1 am on weekdays and weekends in 2016, similarly for 2 am, and so on for each hour of the day.

Once we have calculated the 48  $a_t$  parameters, we use these parameters to calculate the value of the first term of the utility function (2). We thus estimate the value of the first (non-thermal) term of the utility function given the estimated non-thermal load for each day and the respective  $a_t$ 's. We then average these values within three different seasons (winter, summer, and spring/fall) to arrive at three different seasonal utility values for the non-thermal part of the utility function.

For the spring/fall months, where there is little to no thermal loads, this value represents the average level of daily household utility achieved with the loads we observe for 2016; essentially, the  $\bar{u}$  for those two seasons. However, for the summer, the second, thermal term of the utility function plays a larger role as space cooling loads are significant during those months. For the relatively few users with electric space heating, the winter will similarly require the  $\bar{u}$  to be calibrated to consider the heating electric load. For all users, we therefore choose to have two  $\bar{u}$  values for weekday and weekends during the spring/fall, two values for

---

<sup>7</sup> Essentially, in this step, we assume the value of the second, thermal part of the utility function is equal to 0. We assume a value of the *Minload* parameter equal to 0.001

weekday and weekend during the summer and two values for weekday and weekend during the winter, i.e., a total of 6 different  $\bar{u}$  values. For the summer, the calculations of the  $\bar{u}$  values is complicated by the fact that they are not directly determined by cooling load but instead by model-determined indoor temperatures and is further described in the next section. The calibration of the winter  $\bar{u}$  values remains for future calibration work.

In summary, for each household, we have 54 parameters to represent its preferences for electricity use: the 48  $a_t$  values, which represent the relative utility of non-thermal electricity consumption in each hour on weekdays and weekends respectively, and 6  $\bar{u}$  values (one for weekdays and one for weekends in each season) representing the seasonal average daily level of household utility achieved with the loads we observe for 2016.

Once calibrated to the AMI data and the load patterns observed under the business as usual scenario (BAU) scenario with flat volumetric charges in 2016, these  $\bar{u}$  values will define the minimum utility level which the household needs to achieve when minimizing expenditure subject to a new set of electricity prices in the new rate design scenarios we will run in future research.<sup>8</sup>

#### 4.2.2 Calibration of the summer $\bar{u}$ values

To calculate the weekday and weekend  $\bar{u}$  values for the summer, we need to calculate the value of the second term in the utility function (2), i.e., we need to subtract the disutility from any deviation from the assumed blisspoint of 18.3 degrees C. While the value of the first term of the utility function could be calculated based on our regression results as described in the previous section, the second term depends not directly on cooling load but on indoor temperature, which is endogenous to the simulations from DRE.

We therefore used an iterative approach where we first ran DRE using a value of zero for the second term of the utility function, after which the simulation gave indoor temperature results. With the new indoor temperature estimates, we recalculated the value of the utility function for each user based on the value of the first term that we wanted to replicate

---

<sup>8</sup> This implicitly represents an assumption that the elasticity of substitution between consumption of electricity services and other household consumption is zero.

as described in the previous section and the value of the second term as implied by the model-generated indoor temperatures until we reached a point where indoor temperatures remained stable.

Figure 5 shows outdoor hourly temperature data for the summer months compared to indoor hourly summer temperatures predicted by the model. Indoor temperatures drop with some time lag whenever outdoor temperatures go below the bliss point, and rise again to the levels of the bliss point when the outdoor temperature rises again above the bliss point. The figure also represents the final indoor temperatures used for the calculation of the value of the utility function when we reached a point sufficiently close to equilibrium.

Figure 5: Hourly Summer Outdoor Temperature vs Hourly Summer Model-Simulated Indoor Temperature per User

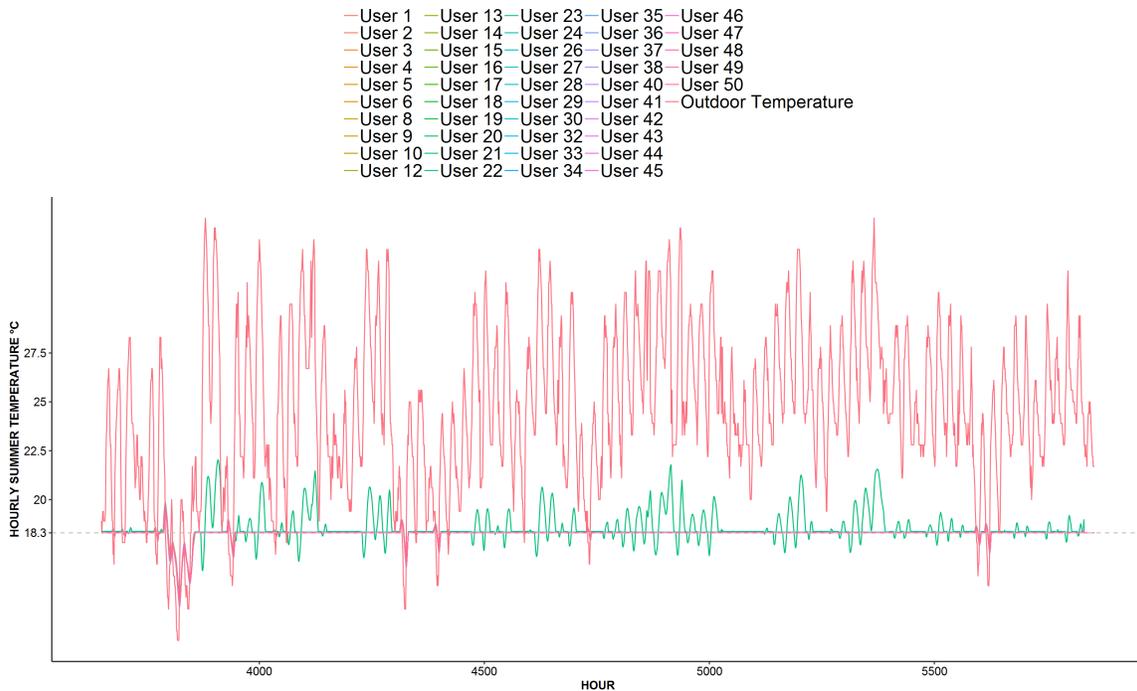
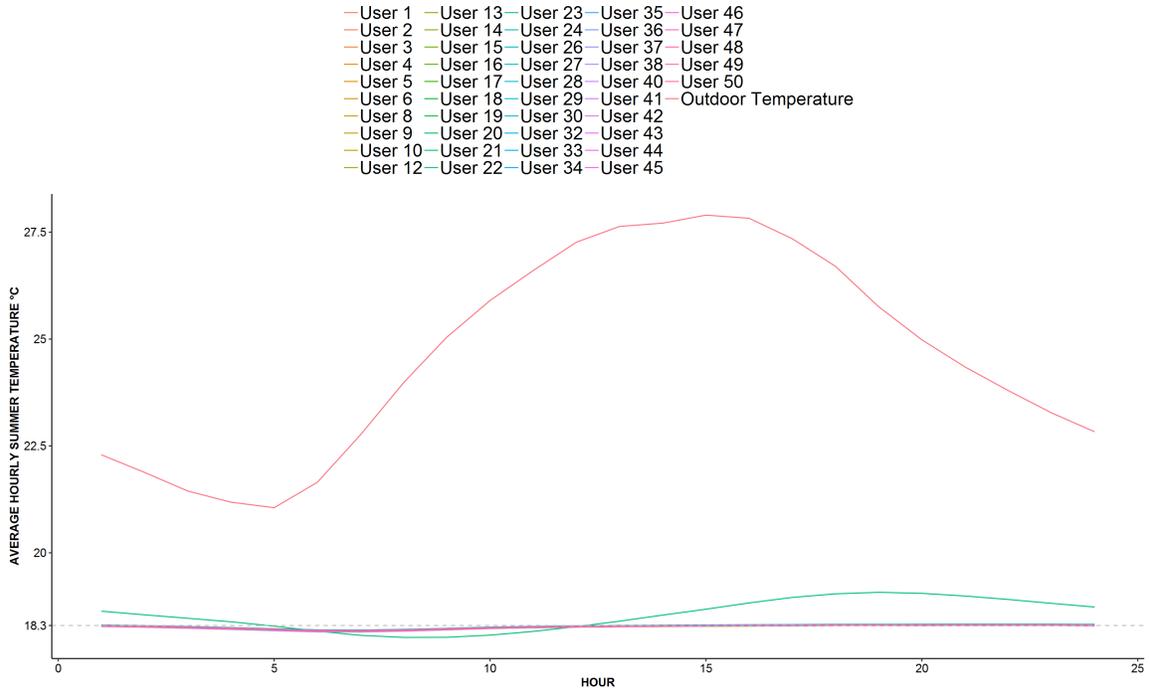


Figure 6 instead illustrates the summer average temperature profiles across the hours of the day. As can be seen, there is very little variation in indoor temperature across users, most being within decimal points to the assumed bliss point of 18.3 degrees, except for user

Figure 6: Average Hourly Summer Outdoor Temperature vs Average Hourly Summer Model-Simulated Indoor Temperature per User



21 which has a high AC load and needs to be further calibrated in future work.

### 4.2.3 Calibration of space cooling load

We next calibrate the model such that the total cooling loads simulated by DRE during the summer months of June, July and August matches the users' regression-predicted cooling loads during those months in 2016. As we show in the next section, this calibration technique will allow us to introduce variation in the model-simulated space cooling loads across all users.

The DRE thermal model depends on multiple input parameters<sup>9</sup> such as an outdoor heat index, thermostat setpoints, air conditioning equipment capacity and coefficient of performance (COP), building resistance value  $R$  (i.e., a measure of the resistance to heat transfer from inside to outside), and building thermal capacity value  $C$  (i.e., a measure of

<sup>9</sup> See [Huntington \[2016\]](#) for a description of these input values and associated references.

the building’s ability to store thermal energy in the building materials). The calibration approach relies on adjusting these two main building thermal parameters for every user such that the regression-predicted space cooling load during the summer months matches the magnitude of the model’s simulated space cooling load for that user.

This part of the calibration process are based on the followin assumptions/parameter values: an indoor temperature “bliss” point of 18.3°C (i.e. an ideal indoor temperature of 65°F consistent with our regression specification) and the value of the  $b$  parameter in the utility function (2) set to 0.1.<sup>10</sup> Furthermore, based on typical appliance specifications we used an air conditioning COP of 3, and a maximum cooling capacity of 4kWe.<sup>11</sup> We also assumed a thermal time constant ( $\tau$ ) of 10 hours based on the old age of the buildings in area we’re modeling. The  $\tau$  defines the ability of a building to retain heat and depends on the building thermal parameters such that  $\tau=R*C$ . With these assumptions, DRE is constrained to a single independent variable and, therefore, when provided a single  $R$  value the model outputs a predictable total cooling load for the summer season.

We run DRE several times for each of the 50 users using multiple  $R$  values. Of these 50 profiles, we pick the mean value in order to get a total summer cooling load for each of the  $R$  values used in the simulations (the standard deviation of the model-generated cooling loads for the sample of users was so small that the mean was representative for the level of cooling load for all the users). Table 4 shows the simulation results.

---

<sup>10</sup> This was the value at which DRE could be made to replicate the regression estimated thermal and non-thermal loads, and was found through iteration.

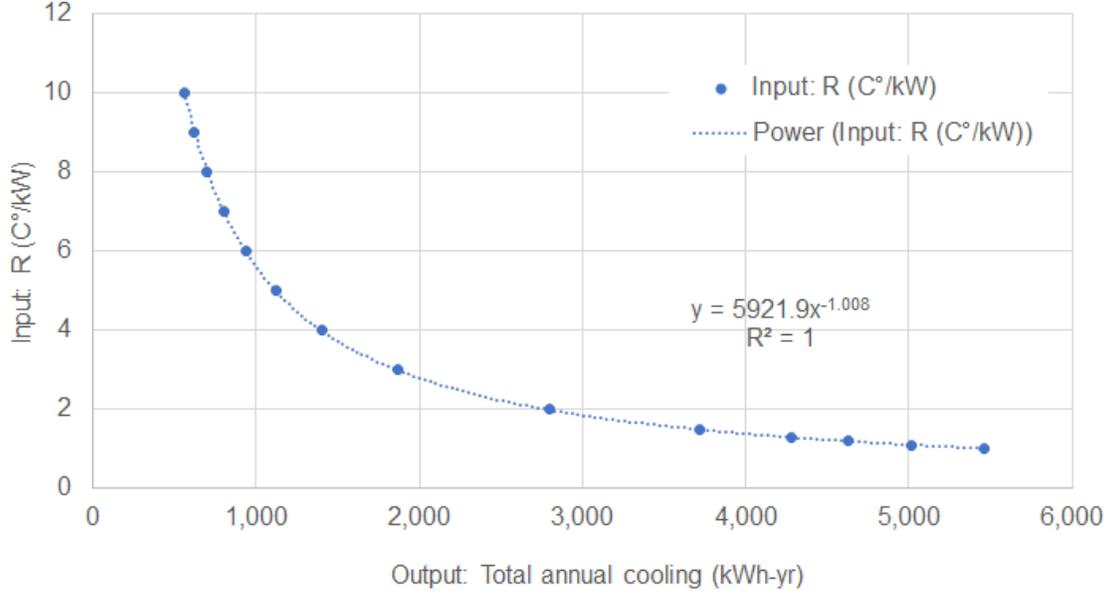
<sup>11</sup> These values are adopted to represent typical AC units found in the ComEd area of study. Based on the 2015 Residential Energy Consumption Survey (RECS), 92% of homes in the Northeast and Midwest regions have air conditioning systems, of which almost 70% correspond to central units. COP values are not provided in the survey, but EIA (2018) reports that installed base for residential central air conditioners in 2009 and 2015 for the North (Not Hot-Dry or Hot-Humid) region had a typical SEER of 11.4-12.5 (equivalent to 3.0 - 3.2 COP). Regarding the size of a AC system, it will depending on the footage of the building and the climate zone, which can range anything between 18kBTU/h and 60kBTU/h, i.e. 1500-5000kW, with central units on the larger range.

Table 4: Input and Output Values for a Constrained Thermal Cooling Model Simulated by DRE

Output: Total summer space cooling load (kWh-yr)	Input: $R$ ( $^{\circ}\text{C}/\text{kW}$ )	$C = \text{Tau}/R$ ( $\text{kWh}/^{\circ}\text{C}$ )
5,459	1.0	10.0
5,019	1.1	9.1
4,625	1.2	8.3
4,281	1.3	7.7
3,718	1.5	6.7
2,794	2.0	5.0
1,867	3.0	3.3
1,401	4.0	2.5
1,122	5.0	2.0
935	6.0	1.7
802	7.0	1.4
702	8.0	1.3
624	9.0	1.1
562	10.0	1.0

The  $R$  and total space cooling load values presented in Table 4 are then used to derive a mathematical expression that approximates the relationship between both variables, i.e.,  $R$  and summer cooling load. By plotting these variables against each other, a power trend-line is fitted to these values as Figure 7 shows.

Figure 7: R Values vs DRE Simulated Summer Cooling Load



An expression shown in Equation (4) is derived to describe the relationship between  $R$  and DRE simulated space cooling load.

$$R = 5921.9 * DREThermalLoad^{-1.008} \quad (4)$$

where  $DREThermalLoad$  corresponds to the total DRE simulated AC load for June, July and August.

Finally, using this expression and the regression-predicted summer space cooling loads for each user, we are able to estimate a unique  $R$  value and (because we assumed a  $\tau$  of 10 hours) also a unique  $C$  value for every user. Fortunately, when re-running DRE after having calibrated the  $R$  values for each user, the model-generated indoor temperatures in general remained sufficiently close to the previous model runs such that the summer  $\bar{u}$  values calculated in the previous step with the same  $R$  value for each user were still valid.

## 5 Calibration Results

### 5.1 Simulation of thermal loads

In this section we present a number of figures to illustrate DRE’s capability of simulating summer cooling loads that match the regression-predicted space cooling loads after we performed the calibration steps described in the previous sections.

Figures 8 and 9 plot the average DRE-generated hourly summer cooling loads across 45 end-users and compares to the results from the thermal load regressions. We removed 5 users out of the 50 end-users random sample from the visualization and analysis because their thermal loads were equal or close to zero based on our regression analysis, hence, in future calibration work we will need to adjust the parameters to reflect a user that does not have an AC.

The figures show that, in general, both DRE-generated and regression-estimated profiles follow a similar pattern and have for most users similar magnitudes. The degree to which both magnitude and pattern are replicated depends on each user’s characteristics and profiles. Some users have higher AC loads for which we still need to calibrate. Even so, the shapes are in general similar and represent the expected peak around 4 and 5 pm.

Figure 8: Average Summer Weekday Thermal Loads Estimated by Regression vs Simulated by DR-DRE Subsample 1

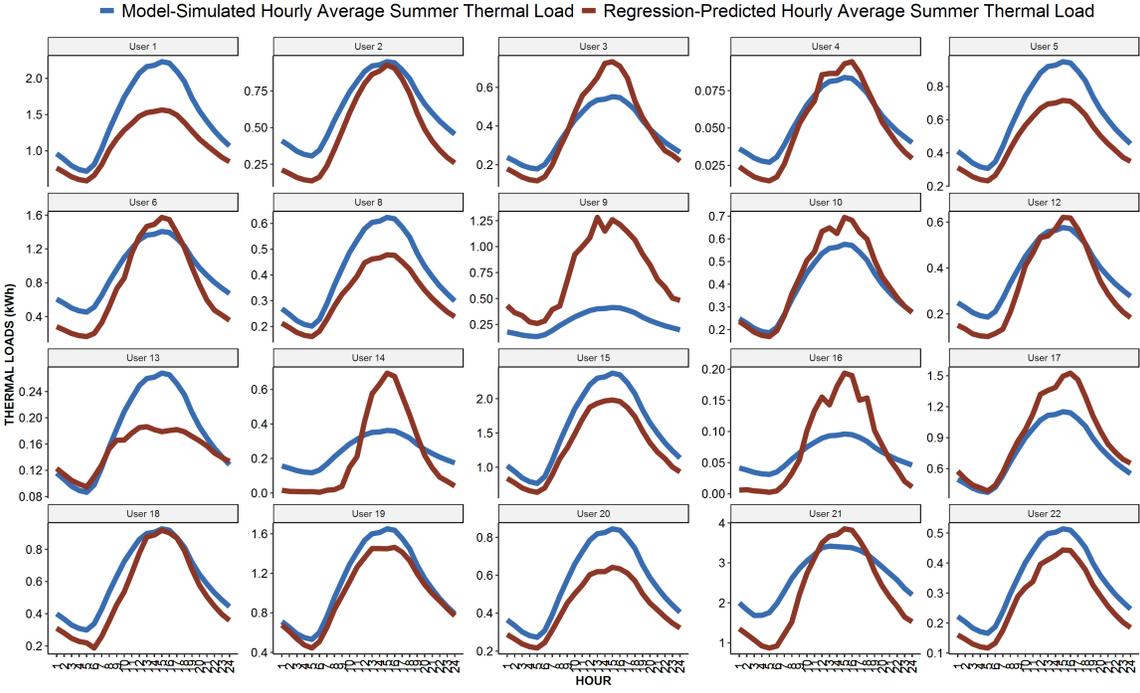
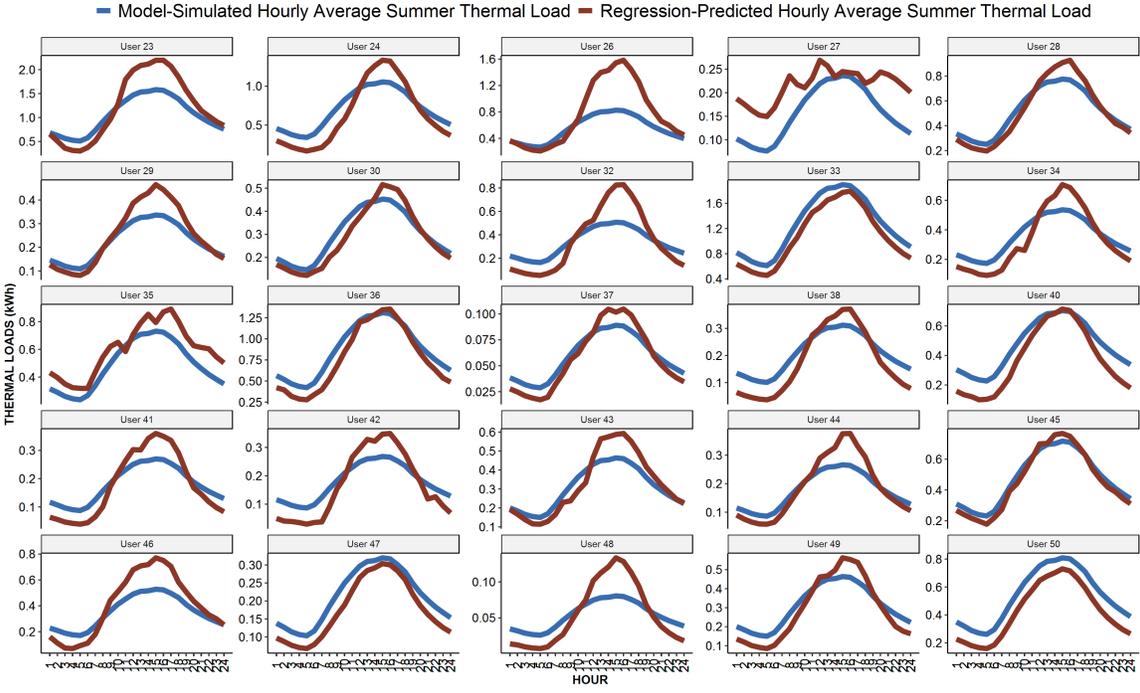


Figure 9: Average Summer Thermal Loads Estimated by Regression vs Simulated by DR-DRE Subsample 2



The cooling loads estimated from the regression share a strong linear relationship with the cooling loads generated by DRE, as shown in tables 5 and 6. The complete hourly series for the whole summer have a mean correlation of .94 across the users, while the hourly average loads (i.e., the data points represented in Figures 8 and 9) have a mean correlation of .98.

Table 5: Correlation Between DR-DRE Simulated Hourly Summer Averaged Thermal Loads and Hourly Summer Averaged Regression-Predicted Thermal Loads

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
0.8530	0.9709	0.9885	0.9803	0.9959	0.9998

Table 6: Correlation Between DR-DRE Simulated Hourly Summer Thermal Loads and Hourly Summer Regression-Predicted Thermal Loads

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
0.7579	0.9318	0.9572	0.9395	0.9788	0.9944

## 5.2 Simulation of non-thermal loads

As described previously, we estimated non-thermal loads as the difference between the observed total loads and the regression-predicted thermal loads. In this section, we present figures to illustrate DRE’s capability of replicating these non-thermal load patterns as estimated in our regressions.

Figures 10 and 11 show how different load shapes are across the 45 users and the ability of DR-DRE to accurately replicate the regression-estimated user specific profiles. Since the profiles were generated based on  $a_t$  values in the utility function (2) that were calculated as averages across the full year, we here illustrate the yearly average non-thermal load profile across the hours of the day for each user.

As with the thermal loads, we removed 5 users with thermal load equal to zero because the model would not be able to correctly replicate correctly the profile, which we will address

in future work.

Figure 10: Average Non-Thermal Loads Estimated by Regression Analysis vs Non-Thermal Loads Simulated by DR-DR Subsample 1

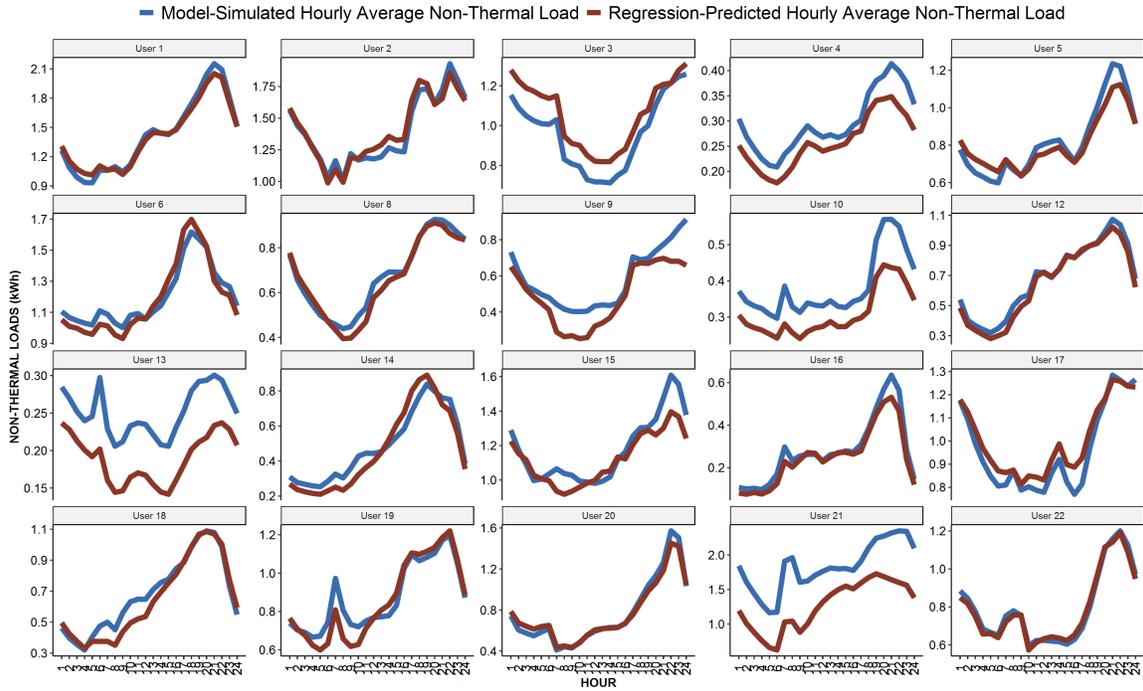


Figure 11: Average Non-Thermal Loads Estimated by Regression Analysis vs Non-Thermal Loads Simulated by DR-DR Subsample 2

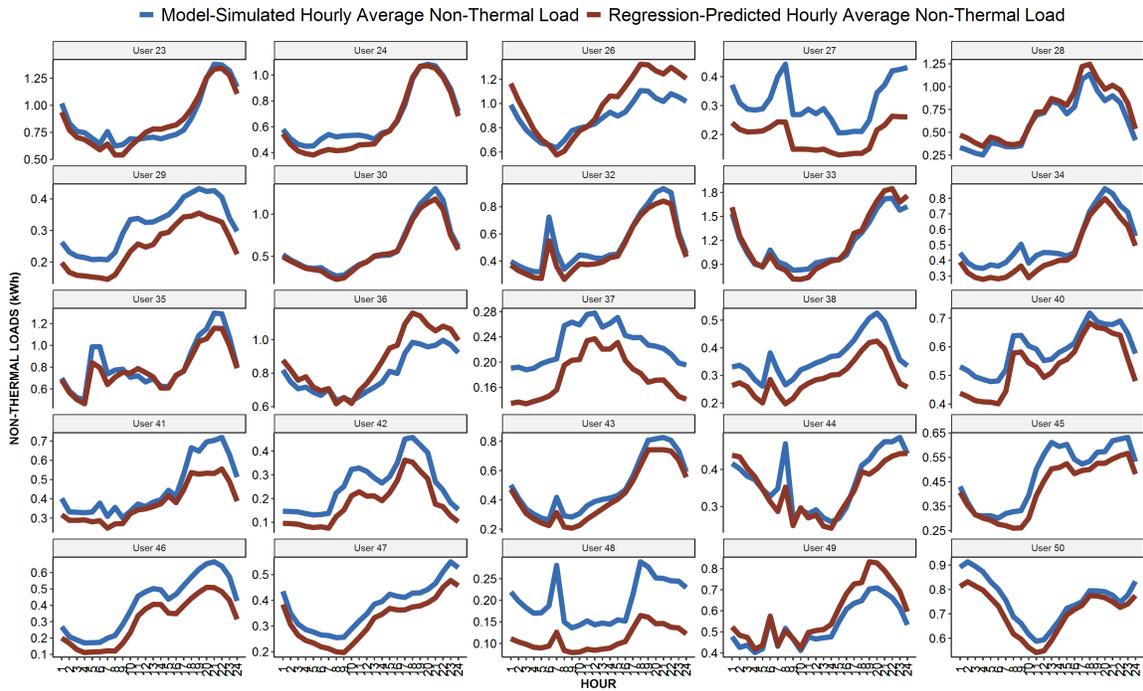


Table 7 below shows the distribution of the correlations across the 45 out of 50 randomly selected users. There are very high correlations between the averaged non-thermal loads predicted by the regression analysis and those simulated by DR-DRE, with a mean coefficient of 0.97 across users, thus demonstrating that the DR-DRE model does a good job of replicating the average non-thermal load shape as estimated by the regressions. This is also illustrated by the correlation plots in Figures 12 and 13.

Table 7: Correlation Between DR-DRE Simulated Hourly Yearly Averaged Non-Thermal Loads and Hourly Yearly Averaged Regression-Predicted Loads

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
0.8638	0.9683	0.9814	0.9713	0.9888	0.9968

Figure 12: Yearly Hourly Average Non-Thermal Loads Correlation Plots Subsample 1

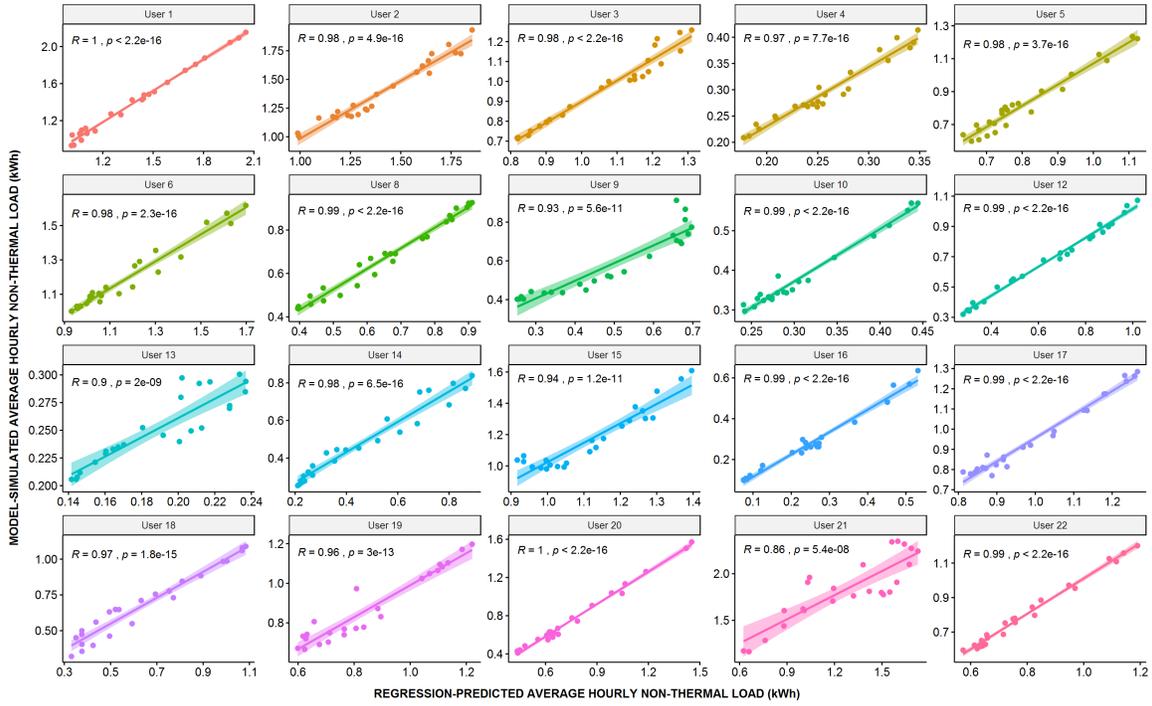
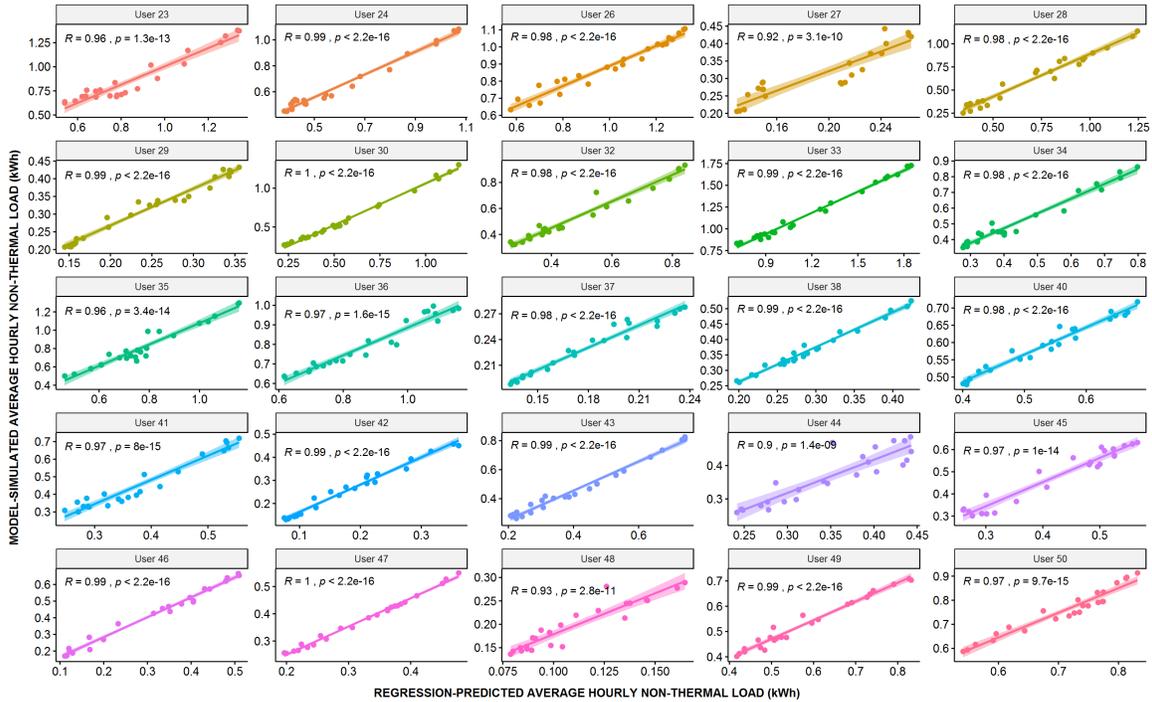


Figure 13: Yearly Hourly Average Non-Thermal Loads Correlation Plots Subsample 2



Unfortunately, we are not as well able to replicate the hourly non-thermal load series over the whole year as shown in Table 8. This is explained by the fact that there is large seasonal variation in the non-thermal patterns across the users, hence, the 48  $a_t$  parameters that determines the shape of the DRE-generated loads will not accurately replicate the non-thermal load in one specific hour of the year. To address this issue, we would need to significantly increase the number of  $a$  parameters which is unwieldy because 54 user-specific parameters is already a lot for a model that is intended to be run for more than 50,000 users. Thus there is a trade-off between how well one specific user’s hourly profile can be replicated by increasing the number of parameters and how many users the model can represent.

Table 8: Correlation Between DR-DRE Simulated Hourly Non-Thermal Loads and Hourly Regression-Predicted Loads

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
-0.47009	0.01964	0.06009	0.05831	0.12084	0.31943

## 6 Future Calibration Finetuning

So far we have calibrated the  $R$ ’s and  $C$  values for users with a space cooling load lower than 3000kWh. Thus, future work consists of estimating a new relationship between  $R$  values and cooling load also for users with cooling loads above that range. This exercise will further improve the calibration results for users with large space cooling loads.

Another future task is to adjust the user-specific parameters to better represent those users with regression-estimated cooling load close to or equal to zero - presumably due to a lack of AC.

In addition, we will parameterize the model to also replicate the space heating load in the winter for those (relatively few) users who are electric space heating customers.

## 7 Conclusion

This research provides an example of how to improve the representation of end-user preferences in electricity simulation models. Each user's preferences are captured in a utility constraint calibrated to advanced metering infrastructure data from a large U.S. electric distribution company. The results of this research demonstrate the capabilities of our modeling tool for creating large numbers of synthetic end-user profiles that can replicate observed load data, relying on a combination of econometric techniques and engineering simulation methods. In future research, we will further improve the calibrations, and then use the resulting individually calibrated preferences to assess how end-users may respond to different electric rate designs by changing their electricity load and investing in DERs.

## References

- Naïm R. Darghouth, Ryan H. Wiser, Galen Barbose, and Andrew D. Mills. Net metering and market feedback loops: Exploring the impact of retail rate design on distributed pv deployment. *Applied Energy*, 162(C):713–722, 2016.
- Sylvestre Gaudin, Ronald C Griffin, and Robin C Sickles. Demand specification for municipal water management: evaluation of the stone-geary form. *Land Economics*, 77(3):399–422, 2001.
- Roy C Geary. A note on “a constant-utility index of the cost of living”. *The Review of Economic Studies*, 18(1):65–66, 1950.
- Ryan Hledik and Gus Greenstein. The distributional impacts of residential demand charges. *The Electricity Journal*, 29(6):33–41, 2016.
- Samuel C. Huntington. Unlocking the value of distributed energy resources in electric distribution networks: Exploring the role of locational tariffs. Master’s thesis, Massachusetts Institute of Technology, 2016.
- Lawrence R Klein and Herman Rubin. A constant-utility index of the cost of living. *The Review of Economic Studies*, 15(2):84–87, 1947.
- Tim Schittekatte, Ilan Momber, and Leonardo Meeus. Future-proof tariff design: recovering sunk grid costs in a world where consumers are pushing back. *Energy Economics*, 70: 484–498, 2018.
- Paul Simshauser. Distribution network prices and solar pv: Resolving rate instability and wealth transfers through demand tariffs. *Energy Economics*, 54:108–122, 2016.
- Richard Stone and Deryck Almond Rowe. *The measurement of consumers’ expenditure and behaviour in the United Kingdom, 1920-1938*. CUP Archive, 1954.