Electricity Demand Simulations on the Distribution Edge:

Developing a Granular Representation of End-User Preferences using Smart Meter Data

Ashwini Bharatkumar[1], Ricardo Esparza[2], Kristina Mohlin[*2], Elisheba Spiller[2], Karen Tapia-Ahumada[1] and Burcin Unel[†3]

[1] *Massachusetts Institute of Technology Energy Initiative*

[2] *Environmental Defense Fund*

[3] *Institute for Policy Integrity at New York University School of Law*

October 8, 2020

Abstract

We combine insights from the 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 assess 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 companion papers, the calibrated model is used to 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. An earlier version of this paper was presented at the IAEE 2019 conference in Montreal.

Contents

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 is thereby 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 between different customer groups, 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.[1]

In our research, we improve upon this common feature of typical engineering models analyzing the impacts of electric rate design by incorporating microeconomic theory into the Demand Response and Distributed Resources Economic model (DR-DRE) developed at MIT. Engineering simulation models like DR-DRE are generally formulated as cost minimization

[1] 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.

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 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 better assess residential customers' potential 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 30 min interval AMI data from a large US electric distribution company to determine parameter values for electricity usage preferences and building thermal characteristics for each of our representative customers.

In this paper, we describe the methodology that is used to include a utility function into the existing DR-DRE model and then calibrate it to observed data. In related papers, we utilize this improved model to estimate the effect of rate design on outcomes such as environmental impacts (Unel et al. [2020]) and DER adoption (Spiller et al. [2020]).

This paper is structured as follows. We first describe the simulation model and how the utility constraint was specified. We next describe how we developed a calibration approach for this new specification using our AMI data. Finally, we present calibration results for our 45 representative residential customers.

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 the HVAC (heating, ventilation, and air conditioning) 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 of electricity across time periods 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 ($/kWh), capacity or demand charges ($/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 the electricity rate and tariff structure were - together with other relevant constraints as described in Huntington [2016]- determined by two disutility parameters (i.e., monetary penalties) for thermal and non-thermal load responses, respectively. For thermal loads, this disutility parameter was a

monetized disutility for indoor temperature deviations outside of a set range around the ideal indoor temperature (or so called bliss point). This results in an end-user optimization around cost and temperature. For example, if the cost of electricity for cooling during summer hours 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 loads, 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 household preferences. Furthermore, this specification did not allow for a difference in the utility/preference for electricity load (or correspondingly disutility of curtailment) during different parts of the day, which is crucial for identifying the effect of time-variant electricity prices on load shifting across hours.

To address the limitations brought by the use of these disutility parameters, we replace the constraints related to these parameters with a household utility constraint (1) as specified below:

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

where

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

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 in (2) represents each household's relative preference

for electricity services other than space heating or cooling (what we will henceforth call non-thermal loads) during each hour. This part of the utility function is an adapted version of the Klein-Rubin (Stone-Geary) utility function.[2] This formulation features a necessary or minimum consumption of each good (here represented by the parameter *MinLoad*). This specification better reflects the choice problem facing an electricity customer considering substitution of electricity consumption in one hour with consumption in another hour based on the relative prices across the hours of the day. The relative size of parameters a_t, in turn, indicates the relative utility of non-thermal electricity use in different hours of the day.

The second and negative part of the utility function represents each household's relative preference for meeting their space heating and cooling demand (what we will henceforth call thermal loads) during each hour by representing the disutility from any deviation between the indoor temperature provided by the household's HVAC system and its ideal indoor temperature (bliss point) as a convex function.[3] 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 model specification.

4 The AMI Load Data

Our smart deter data are customer-level AMI load data for 54,412 residential customers for the year 2016 from Commonwealth Edison (ComEd) in Chicago. We aggregate the 30-min interval data up to hourly loads resulting in 8,784 data points per household (2016 was a leap year).

Table 1 below shows the customer class and tariffs for these customers. In 2016, ComEd introduced a Residential Real Time Pricing (now more accurately re-named Hourly Pricing)

[2] 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]

[3] We assume a bliss point of 18.3 degrees Celsius.

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 dataset face the flat (time-invariant) volumetric charge per kWh in the default residential rate as shown in the table.

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

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

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

5 Calibration approach

DR-DRE previously relied on simulated residential load profiles based on average thermal and non-thermal load patterns. Through access to the AMI load data from ComEd described in the previous section, we are able to represent a more varied set of preferences as represented in these observed load data.

However, our AMI data show only each household's total electricity use in each hour, and do not differentiate between different end uses of electricity, such as space cooling versus

other appliance usage. 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.

To develop our calibration approach, we use a random sample of 50 households from the AMI load data set described in the previous section. In the results section, we validate our calibration approach by comparing model-simulated profiles produced with the calibrated model against the thermal and non-thermal load profiles predicted with our regressions.

5.1 Econometric estimation of thermal and non-thermal loads

We first estimate hourly thermal loads 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.

5.1.1 Econometric specification

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.[4] 65 F is therefore also what we assume to be the so called bliss point. We are thus attempting to estimate the responsiveness of electricity demand to outdoor temperature due to residential space cooling

[4] This is, for example, how Cooling and Heating Degree Days are defined.

and heating, respectively. We define this incremental response as thermal load.[5]

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. 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 specified 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:

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

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

[5] As alluded to earlier, this specification assumes that water heating usage is not a part of our thermal load as here defined.

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.

The CDH, HDH and humidity variables are based on the outdoor temperature and relative humidity in 2016 at Chicago Midway Airport Climatological Data Station, retrieved from a public dataset provided by the National Centers for Environmental Information (NCEI).

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:

$$\widehat{ThermalLoad}_t = \hat{\beta}_{CDH,t}\, CDH_t \; + \hat{\beta}_{CDH^2}\, [CDH_t]^2 \tag{3}$$

and for hours with temperature <65 F:

$$\widehat{ThermalLoad}_t = \hat{\beta}_{HDH,t}\, HDH_t \; + \hat{\beta}_{HDH^2}\, [HDH_t]^2 \tag{4}$$

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

Our econometric specification thus estimates the responsiveness of customers' electricity usage to outdoor temperature. We thereby 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 because 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 (beyond the monthly fixed effects which captures any differences in the average monthly electricity price).

5.1.2 Econometric Results

In the figures below, we show the regression results for our random sample of 50 households (henceforth users).

Figures 1 and 2 show the regression-estimated average seasonal thermal load profile across the hours of the day for our sub-sample. The peaks with the green line representing the average summer thermal load shape and seen for most of the users are consistent with AC use which increases during the warmer hours of the day. The only user with a significant estimated winter thermal load is user 7 which indeed turns out to be the only electric space heating customer in this random sample of 50 users (note that the scales on the y-axis are different across the users and determined by the size of each user's average hourly thermal load).

Figure 1: Average Regression-Predicted Thermal Load across the Hours of the Day - Subsample 1

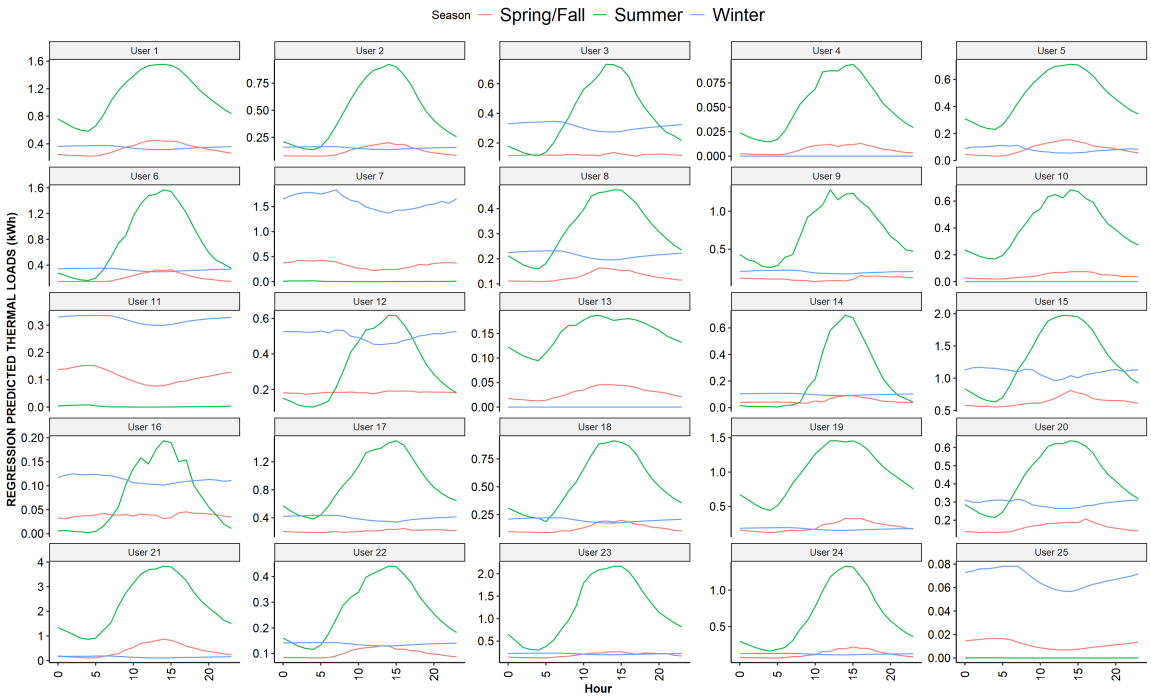
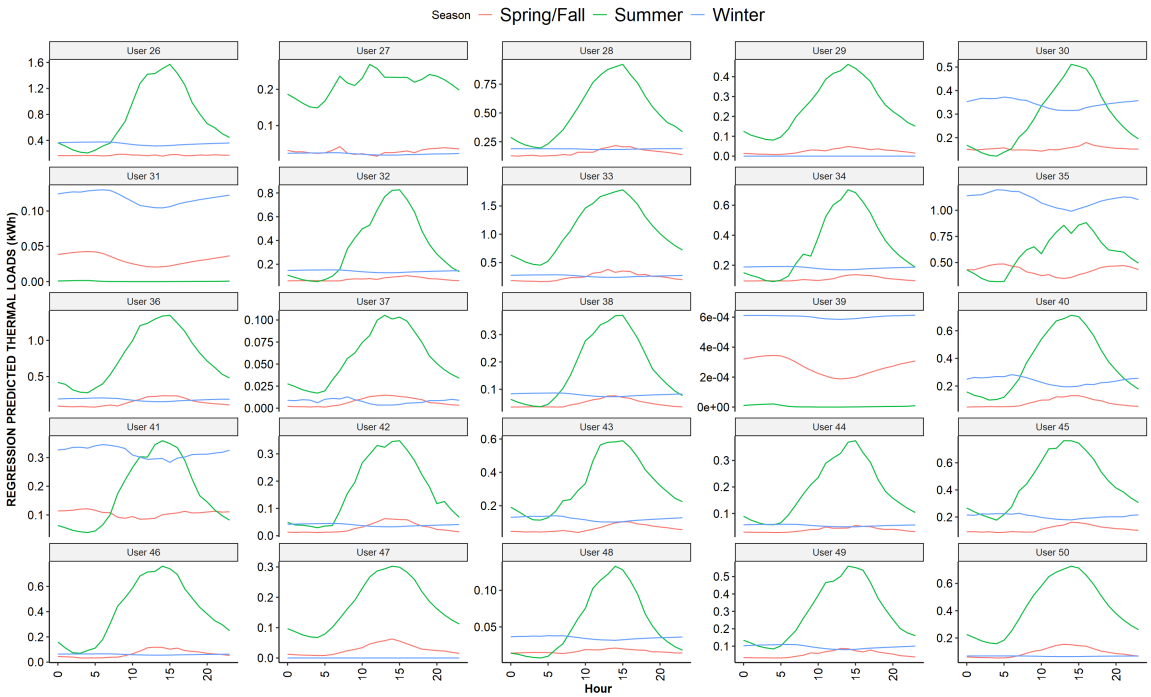


Figure 2: Average Regression-Predicted Thermal Load across the Hours of the Day - Subsample 2



Figures 3 and 4 show the average seasonal non-thermal load profile across the hours of the day based on the regressions. The hourly non-thermal loads were calculated by subtracting the thermal loads estimated with equations (3) and (4) from the household's total load in each hour. While the non-thermal load shapes vary across the users and for some users also by season, there is for many a fairly typical pattern of increased usage during the afternoon and evening when residential customers typically ramp up their usage of electric appliances.

Figure 3: Average Regression-Predicted Non-Thermal Load across the Hours of the Day - Subsample 1

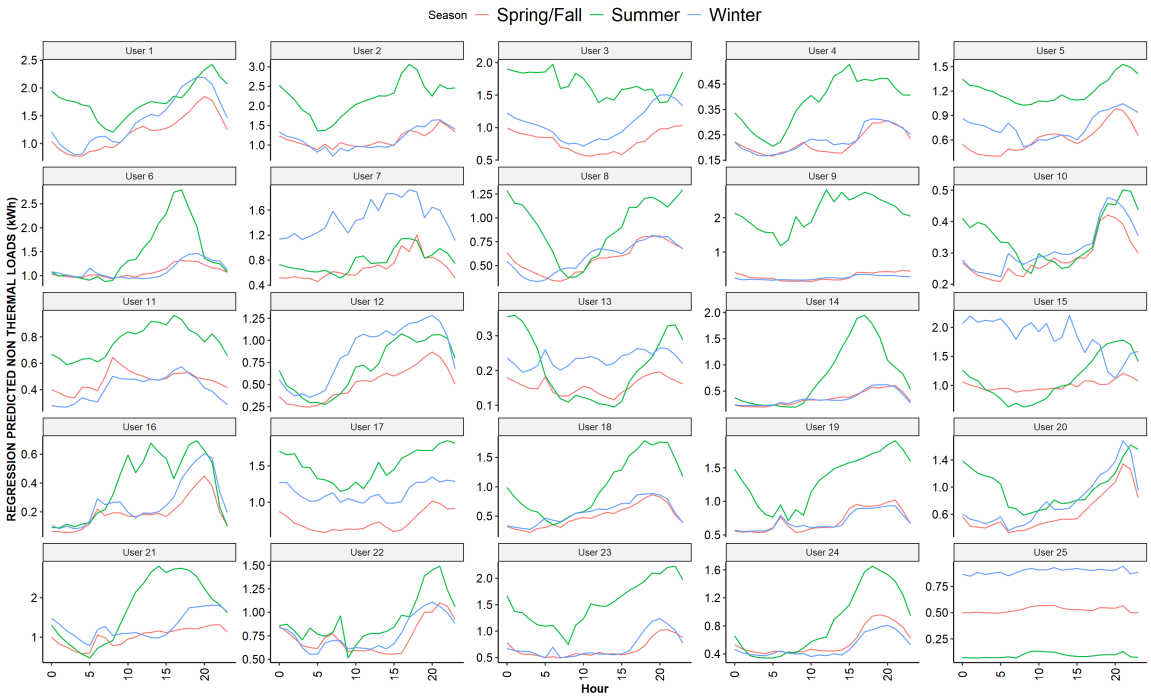
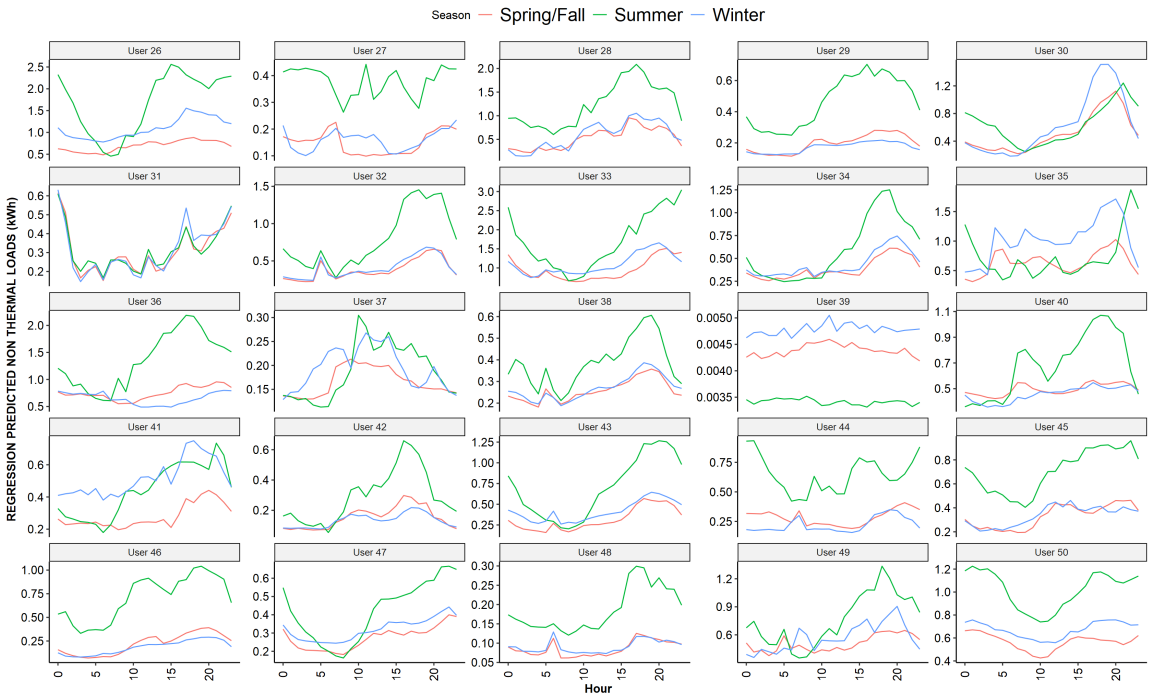


Figure 4: Average Regression-Predicted Non-Thermal Load across the Hours of the Day - Subsample 2



5.2 Calibration using the estimated thermal and non-thermal loads

This section describes our approach for calibrating the parameters of the updated DR-DRE model against the non-thermal loads and thermal loads estimated as described in the previous section.

5.2.1 Calibration of household preference parameters determining non-thermal load

The preference parameters we need to calibrate against the electric load data are the a_t values and the $\bar{u}$ values from expressions (1) and (2) for each household. Using the estimated non-thermal load for each hour, 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.[6]

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

$$a_t = \frac{price(t)\ [Non\,Thermal\,Load\,(t) - MinLoad]}{m - \sum_{t=1}^{T} price(t)\ MinLoad} = \frac{Non\,Thermal\,Load\,(t) - MinLoad}{\sum_{t=1}^{T} Non\,Thermal\,Load\,(t) - \sum_{t=1}^{T} MinLoad} \quad (5)$$

, where $price(t)$ is the electricity price in hour t. Because the 2016 electricity prices were time-invariant across the hours of the day for the vast majority of our sample of residential electricity users, 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.[7] To arrive at a reasonable number of parameters for each household (which could otherwise be 8784 parameters - one for each hour of the year) that represent the household's preference for non-thermal electricity use in each hour of the day, we calculate 48 average values to represent each hour of the day on weekends and weekdays, respectively. For example, we calculate the mean for all the values of a_t for 1 am on all the weekdays and weekends, respectively 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 calculate 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

[6] 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. Essentially, in this step, we assume the value of the second, thermal part of the utility function is equal to 0.

[7] We assume a value of the $Minload$ parameter equal to 0.0001.

the average level of daily household utility achieved with the loads we observe for 2016; essentially, the value of the $\bar{u}$ parameter 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 all users, we therefore choose to have two $\bar{u}$ values for weekday and weekends during the spring/fall, two values for weekday and weekend during the summer and two values for weekday and weekend durign 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 the second term in the utility is not directly determined by cooling load but instead by model-determined indoor temperatures and is further described in the next section.

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 load data under the business as usual scenario (BAU) scenario with flat volumetric charges in 2016, these $\bar{u}$ values define the minimum household utility level[8] which the DR-DRE solution needs to achieve when minimizing expenditure subject to a new set of electricity prices in our rate design scenarios.

5.2.2 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 (65 F). 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

[8] This implicitly represents an assumption that the elasticity of substitution between consumption of electricity services and other household consumption is zero.

the second term of the utility function, after which the simulation gave indoor temperature results. With the new indoor temperature value, we recalculated the value of the summer $\bar{u}$ value for each user based on the value of the first term as described in the previous section and the value of the second term as implied by the model-generated indoor temperatures and thereafter re-ran the model with the updated $\bar{u}$ value and so forth 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. As we can see there is very little difference across the users in the indoor temperature (illustrated with the one green line). 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

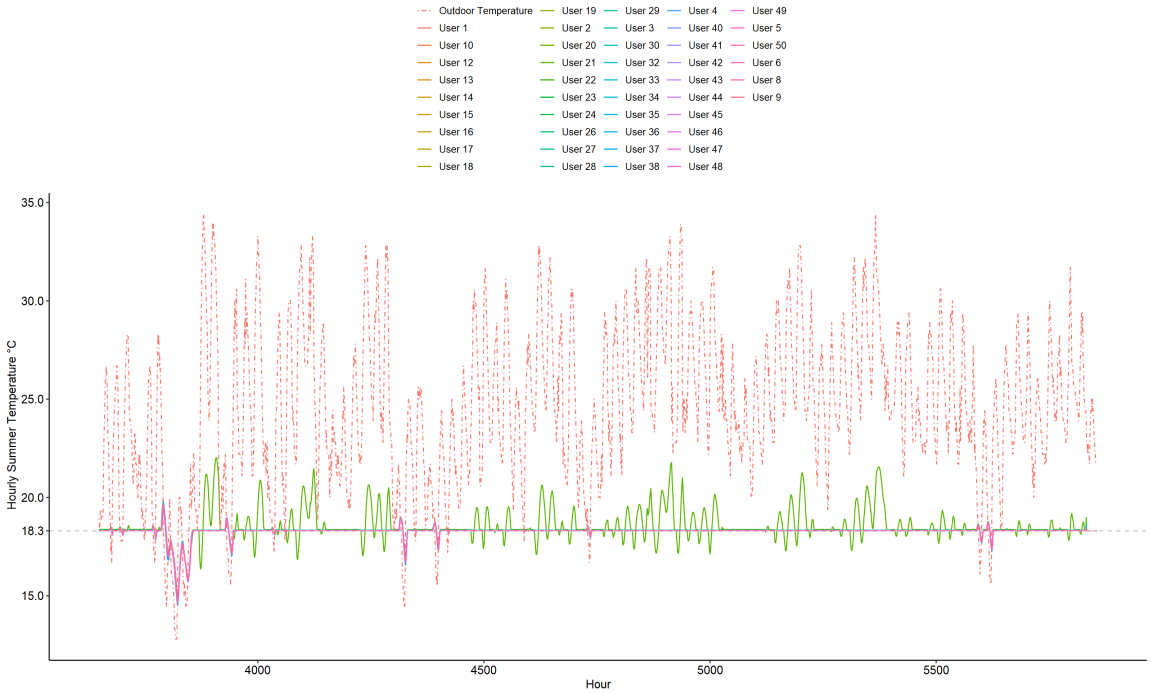
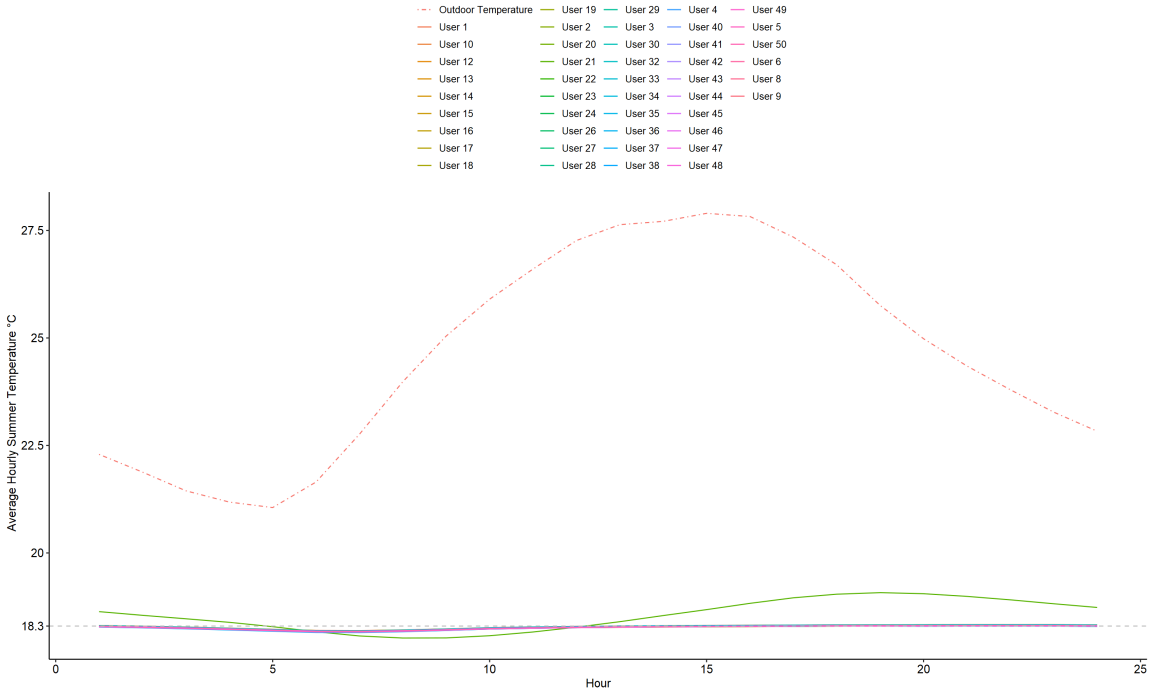


Figure 6 instead illustrates the summer average indoor temperature profiles across the hours of the day. As can be seen, there is very little variation in indoor temperature across users, most being within decimal points to the assumed bliss point of 18.3 degrees and therefore not individually visible, except for the outlier of user 21 (green line) which has a high AC load.

5.2.3 Calibration of parameters determining space cooling load

We next calibrate the model such that the total cooling loads simulated by DR-DRE during the summer months of June, July and August matches the users' regression-predicted cooling loads during those months in 2016. This calibration step allows us to introduce variation in the magnitude of model-simulated space cooling loads across all users.

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



The DR-DRE thermal module depends on multiple input parameters[9] 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 the building's ability to store thermal energy in the building materials).

This calibration step relies on adjusting the R and C 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 is based on the following 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.[10] Furthermore, based on typical appliance specifications we assumed for the AC unit a COP of 3 and a maximum cooling capacity of 4kWe.[11] We also assumed a thermal time constant (τ) of 10 hours based on the old age of the buildings in the Chicago area our load data are from. The τ defines the ability of a building to retain heat and depends on the building thermal parameters such that $\tau = R*C$. With these assumptions, DR-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 DR-DRE several times for each of the 50 randomly sampled users using multiple R values. We pick the mean value across the 50 users in order to get a total summer cooling load for each of the R values used in these simulation runs (the standard deviation of the model-generated cooling loads across the sample of users was so small that the mean was

[9] See Huntington [2016] for a description of these input values and associated references.

[10] This was the value at which DR-DRE could be made to replicate the regression estimated thermal and non-thermal loads, and was found through iteration as described in the previous section.

[11] 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.

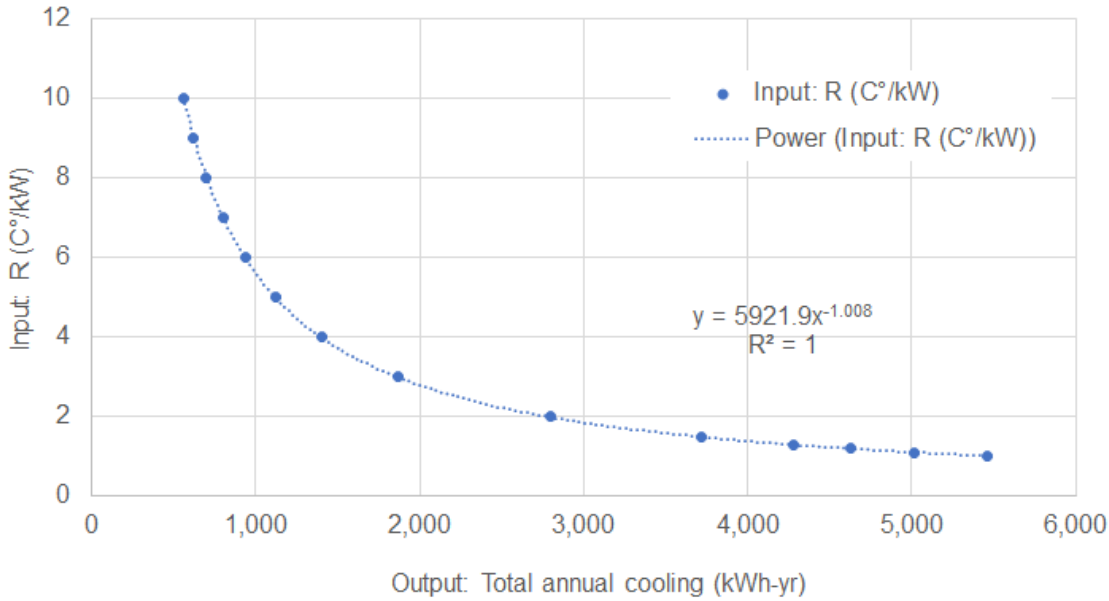
representative for the level of cooling load for all the users). Table 4 shows the simulation results.

Table 4: Input and Output Values for Constrained Space Cooling Runs Simulated in DR-DRE

Output: Total summer space cooling load (kWh-yr)	Input: R (C°/kW)	C = Tau/R (kWh/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 summer 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 (6) is derived to describe the relationship between R and DRE simulated space cooling load.

$$R = 5,921.9 * DREThermalLoad^{-1.008} \tag{6}$$

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

Finally, using this expression and paired with the regression-predicted summer space cooling loads for each user, we are able to determine a unique R value and (because we assumed a τ of 10 hours) also a unique C value for every user. Fortunately, when re-running DRE after having thus 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 were still valid.

The result of this exercise is that we can use the estimated relationship in (6) to determine

each user's R (and by implication C) values for their building characteristics based on the user's regression-predicted cooling loads during the 2016 summer months.

6 Calibration Results

For the electricity tariff scenarios in Spiller et al. [2020] and Unel et al. [2020], we chose to focus on single family households with no electric space heating as these 44,185 customers represent the vast majority in our AMI dataset as can be seen in section 4. Due to computational constraints, we needed to develop a more limited set of user profiles representative of these customers. 45 representative profiles were therefore developed using a clustering approach as further described in Esparza et al. [2020]. This section presents the results of using the approach described in section 5 to calibrate the parameters for these 45 representative customers. As these are non-electric space heating customers, there were no thermal winter electric load data to calibrate winter thermal parameters against. The assumptions we made with respect to these customers' space heating are instead further described in our Methodology report [ADD APPROPRIATE REFERENCE].

6.1 Creating representative load profiles

The clustering approach as described in Esparza et al. [2020] yielded 45 segmented groups of households by grouping the households based on key features of their load patterns (such as daily load shape variability, consumption volume, and peak times) that could affect DER adoption and response to different electricity tariffs/rate designs. For each cluster/group, we created a representative electric load profile of 8784 hourly load data points by, for each hour of the year, taking the average total load across all the households in the cluster.

6.2 Simulation of baseline thermal loads

Below, we present figures to illustrate DR-DRE's capability of simulating summer cooling loads that match the regression-estimated thermal load profile for each of our 45 representative

customers (i.e., after we for each representative customer performed the calibration steps described in section 5).

For each of the representative customers, Figures 8 and 9 plot the average DRE-generated hourly summer cooling loads and compares to the results from the thermal load regressions performed on the representative load profile. The figures show that, in general, the DRE-generated and regression-estimated thermal 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 cluster's characteristics and profiles. Overall, the shapes are in general similar and represent the expected peak around 3 and 4 pm for residential electricity users.

Figure 8: Average Summer Thermal Loads across the Hours of the Day Estimated by Regression vs Simulated by DR-DRE - for representative customer 1-20

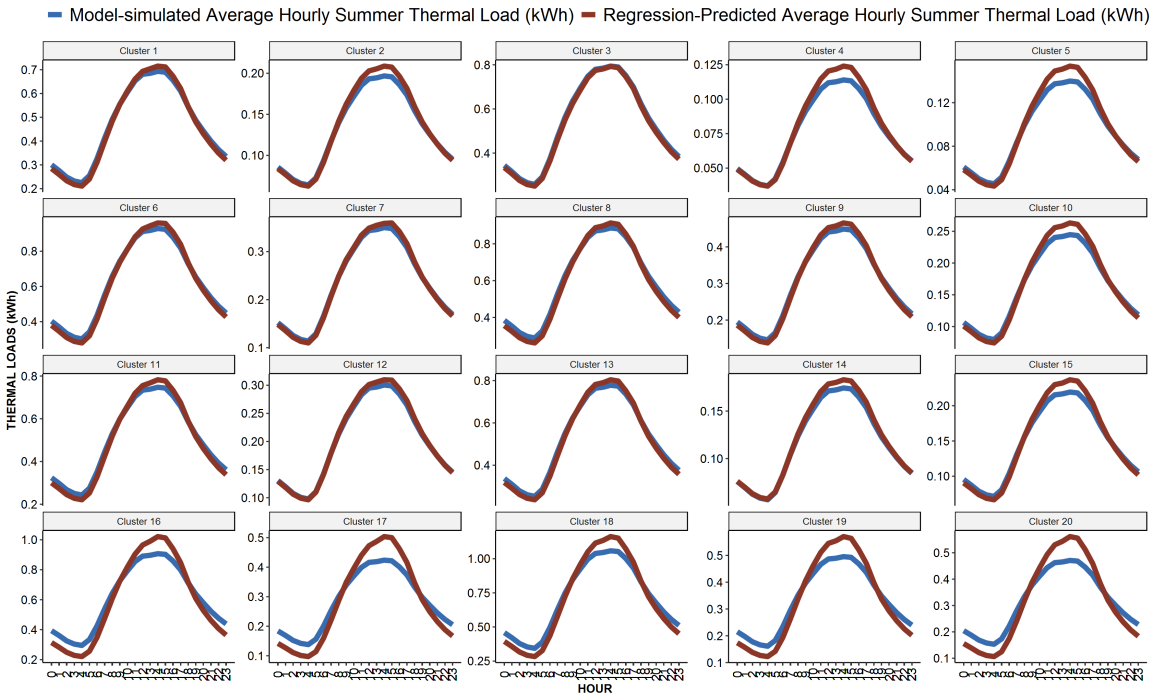
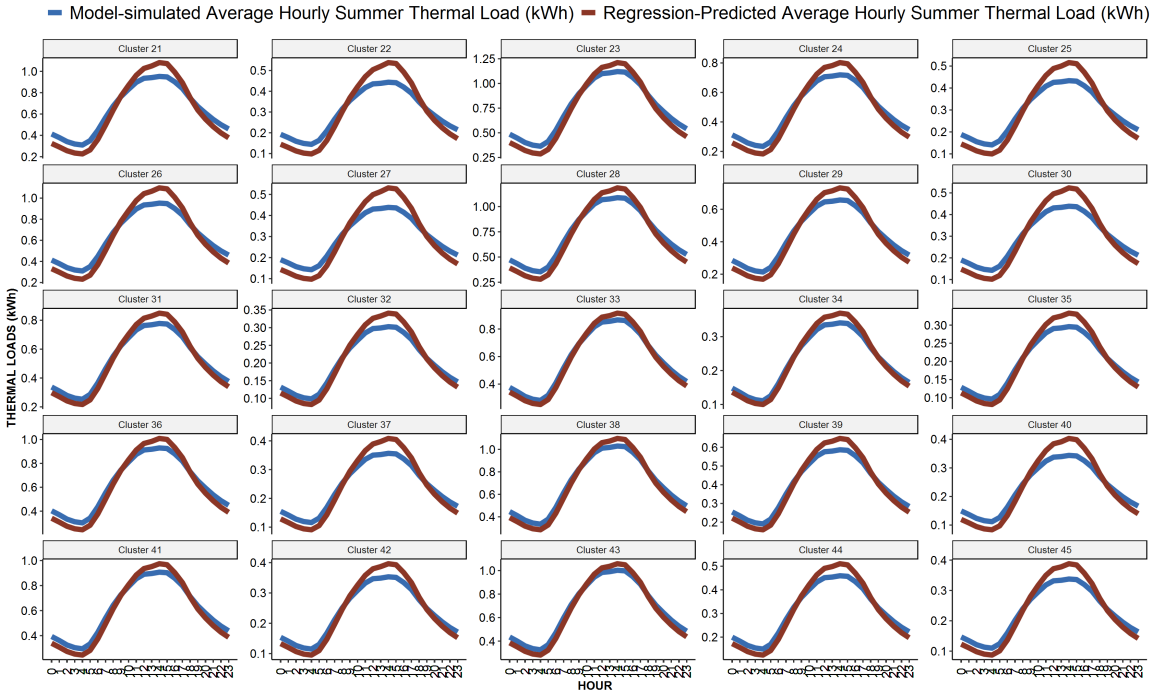


Figure 9: Average Summer Hourly Thermal Loads across the Hours of the Day Estimated by Regression vs Simulated by DR-DRE - for representative customer 21-45



The cooling loads estimated from the regression share a strong linear relationship with the cooling loads generated by DRE, as shown in table 5. The complete hourly series for the whole summer have a mean correlation of .98 across the 45 representative customers.

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

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
0.9610	0.9769	0.9846	0.9839	0.9931	0.9987

6.3 Simulation of baseline non-thermal loads

Below, we present figures to illustrate DR-DRE's capability of simulating non-thermal loads that match the regression-estimated non-thermal load profile for each of our 45 representative customers.[12]

For each representative customer, Figures 10 and 11 plot the average DRE-generated non-thermal load profiles and compares to the regression-estimated non-thermal load profiles. With only a few exceptions (such as cluster 4, 9 and 34) the profiles for each user are very similar, illustrating that the calibration approach performs well. The figures also show how different the load shapes are across the 45 clusters' representative profiles.[13]

[12] In line with section 5.1, we estimated non-thermal loads for each representative customer as the difference between the total load and the regression-predicted thermal load in each hour.

[13] Since the DR-DRE 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.Since this is the yearly average across the hours of the day, these profiles do not differentiate between weekday and weekend.

Figure 10: Average Non-Thermal Loads across the Hours of the Day Estimated by Regression Analysis vs Non-Thermal Loads Simulated by DR-DR - for representative customer 1-20.

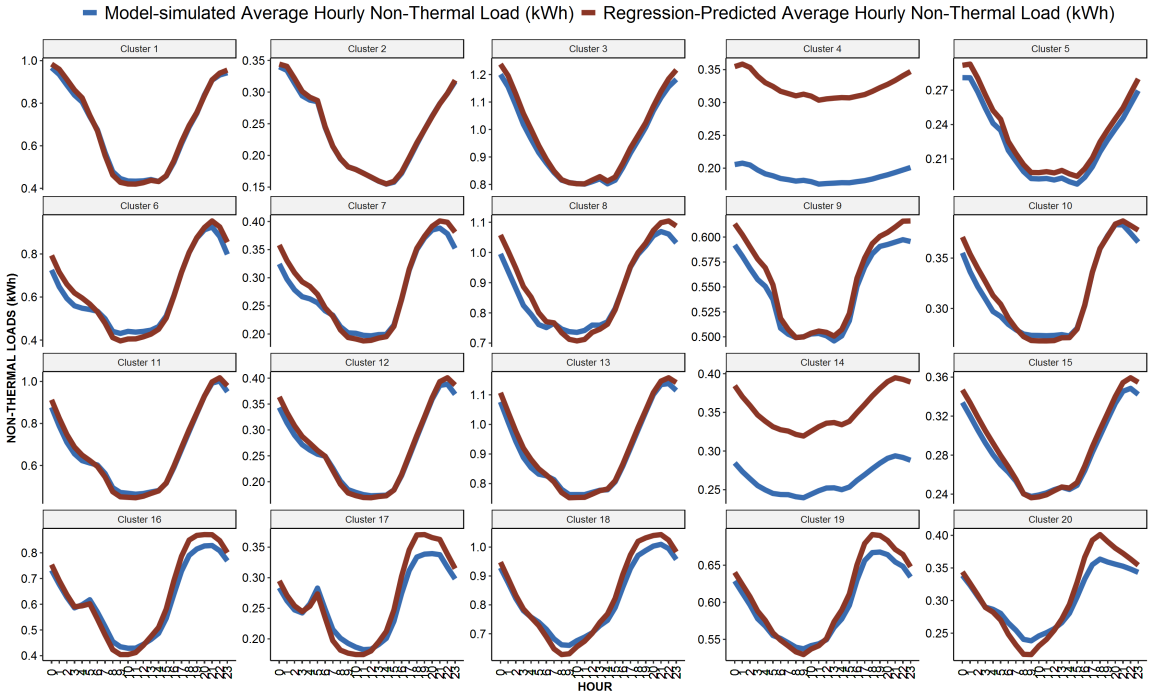


Figure 11: Average Non-Thermal Loads across the Hours of the Day Estimated by Regression Analysis vs Non-Thermal Loads Simulated by DR-DR - for representative customer 21-45.

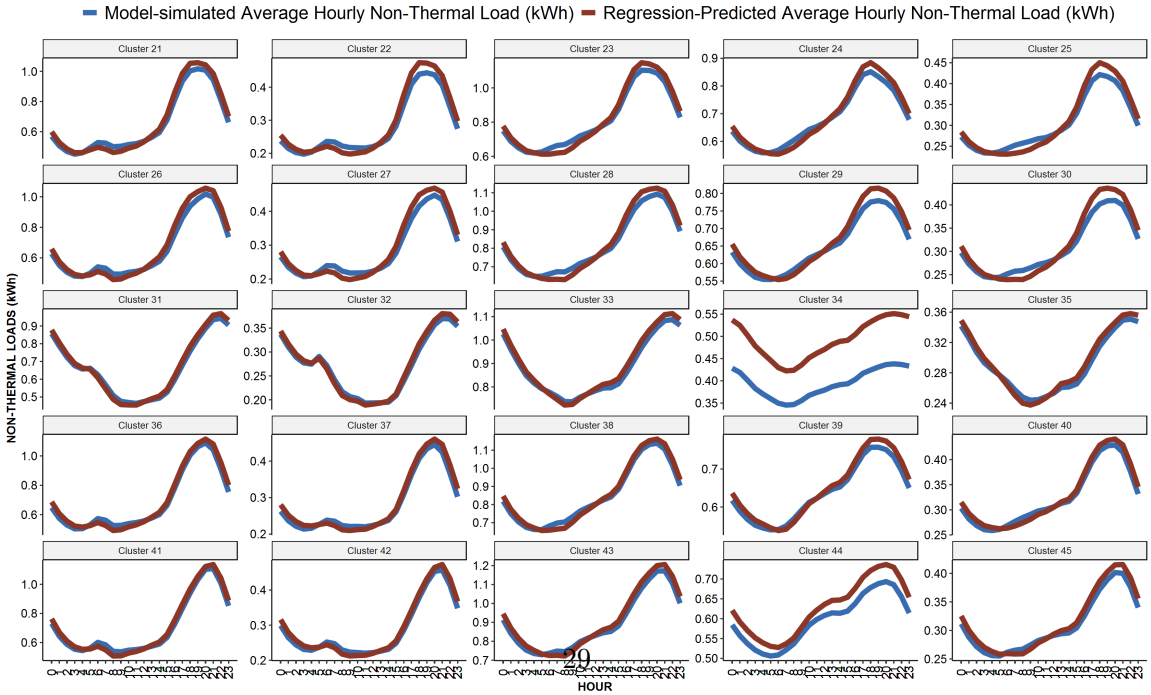


Table 6 below shows the distribution of the correlations across the 45 clusters. There are very high correlations between the averaged non-thermal loads predicted by the regression analysis and those simulated by DR-DRE, with a mean coefficient of 0.99 across users, thus demonstrating that the DR-DRE model does a good job of replicating the average non-thermal load shape as estimated by the regressions.

Table 6: 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
0.9793	0.9933	0.9951	0.9945	0.9969	0.9997

Unfortunately, we are not as well able to replicate the (non-averaged) hourly non-thermal load series over the whole year as shown in Table 7. This is explained by the fact that there is large seasonal variation in the non-thermal patterns across the clusters' representative customers and, 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 significant computing penalty for DR-DRE. Thus there is a trade-off between how well one specific user's hourly non-thermal profile can be replicated by increasing the number of parameters and how long it takes to run the model for all the users. Even so, we are able to achieve a relatively high correlations, around 70% or more.

Table 7: Correlation Between DR-DRE Simulated Hourly Non-Thermal Loads and Hourly Regression-Predicted Loads across 45 clusters

Min.	1st Quartile	Median	Mean	3rd Quartile	Max
0.0729	0.6767	0.7142	0.6901	0.7344	0.7709

7 Discussion

This paper describes how we improved the representation of end-user preferences in an existing electricity simulation model. With the approach presented here, 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 demonstrate the capabilities of the DR-DRE model for creating synthetic end-user profiles that can closely replicate observed load patterns.

Spiller et al. [2020] and Unel et al. [2020] use the calibrated 45 profiles presented in this paper as the baseline to run scenarios with new electricity tariffs and assess how end-users may respond to different electric rate designs by changing their electricity load and investing in DERs.

In future research, once computational constraints are further relaxed, one could consider running similar simulations on household specific load data rather than relying on representative households based on clustering methods.

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.

R. Esparza, E. Spiller, K. Mohlin, K. Tapia-Ahumada, and B. Unel. Creating representative profiles from clustered non-electric heating customers load profiles. Working paper, 2020.

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.

E. Spiller, R. Esparza, K. Mohlin, K. Tapia-Ahumada, and B. Unel. The role of electricity tariff design in distributed energy resource deployment. Working paper, 2020.

Richard Stone and Deryck Almond Rowe. *The measurement of consumers' expenditure and behaviour in the United Kingdom, 1920-1938*. CUP Archive, 1954.

B. Unel, R. Esparza, K. Mohlin, E. Spiller, and K. Tapia-Ahumada. Retail electricity rate design, distributed energy resources, and emissions. Working paper, 2020.