Applied Energy 377 (2025) 124583

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Revealing spatial and temporal patterns of residential cooling in Southern California through combined estimates of AC ownership and use

McKenna Peplinski, Stepp Mayes, Kelly T. Sanders *

University of Southern California, Department of Civil and Environmental Engineering, Los Angeles, CA 90089, USA

HIGHLIGHTS

- Three-part methodology to characterize patterns of residential AC ownership and use.
- Smart meter dataset of ∼200,000 households in Southern California.
- Estimated AC Ownership Rate of 79 % and average AC Operation rate of 8.3 %.
- Average AC Operation Rate ranges from 1 to 23 % of total hours across census tracts.
- Evaluate energy insecurity using Net AC Utilization, a function of ownership & use.

ARTICLE INFO

Keywords:
Air conditioning
Smart meters
Residential buildings
Southern California

ABSTRACT

Air conditioning (AC) is an important tool for combatting the adverse health effects of heat, but its use can also drive surges of high demand for electricity. To better understand these effects, there is a need for non-intrusive methods of estimating AC access and operation. In this study, we use a novel methodology to identify residential AC ownership rates using smart meter data from 200,000 customers in Southern California, and find that 79 % of all customers in the region have AC. In contrast to previous methods, we classify AC ownership using hourly, rather than daily, electricity consumption records and directly account for the potential presence of electric heating. We then adapt and apply an algorithm to determine in which hours these households operate their AC. We estimate that the average customer runs their AC during 8.3 % of all hours in the two-year study period, but census-tract level averages range from 1 to 23 % of all hours. Lastly, we combine our estimates of AC ownership and use to analyze cooling behavior spatially and temporally, and are able to identify pockets of high cooling demand, areas lacking in access or the ability to use their AC, and potential targets for cooling-related DR programs.

1. Introduction

Rising temperatures associated with global climate change and urban warming coupled with higher standards of living are set to drive huge increases in the electricity demand for cooling. By 2050, the global capacity for air conditioning (AC) is expected to triple through both new AC adoptions and increased use of existing units [1]. While it is prudent to ensure that people have sufficient access to cooling resources, especially as extreme temperatures threaten public health [2], doing so will have large implications for the power grid. Thus, a robust understanding of residential cooling demand is necessary to identify communities without adequate AC access and plan for future energy needs.

AC is a key adaption tool to protect populations from the from the health effects of climate change [3], especially as extreme heat events both intensify and become more frequent [4]. As such, equitable and resilient cooling access is a major objective for several stakeholders including utilities, public health officials, and energy advocates. Although there has been huge growth in AC adoption, there are still many communities and countries in warm climates with low rates of AC ownership [5,6]. Further, many households with AC are unable to meet their energy needs because of the rising costs of electricity. For example, in a 2020 survey of US household energy insecurity, 5 % of respondents cited that financial circumstances prevented them from using their AC and 11 % reported keeping their home at an unhealthy temperature to

* Corresponding author.
E-mail address: ktsanders@usc.edu (K.T. Sanders).

https://doi.org/10.1016/j.apenergy.2024.124583
Received 5 June 2024; Received in revised form 9 September 2024; Accepted 24 September 2024
Available online 1 October 2024

lower their electricity bill [7]. It is important to identify the communities that do not have adequate access to AC (either because of the complete lack of AC or underutilization of cooling appliances due to energy insecurity) to promote policies that will ensure vulnerable communities will not be in danger during extreme heat events.

Meeting the increased demand for cooling could exacerbate the challenge of managing peak loads across the power grid. The use of AC units and fans currently account for 20 % of electricity demand in buildings and 10 % of all global electricity consumption [1]. With rising temperatures and increasing AC adoption, this percentage is expected to grow, placing strain on the electric grid through increases in both the overall and peak demand [8–10]. Grid operators and utilities rely on accurate forecasts of electricity demand to ensure there is enough power on the grid at any given time [11]. As AC units account for a significant portion of electricity consumption in hot months, accurate estimates and projections of the cooling demand are critical.

The challenge of quantifying a region's demand for cooling is twofold. Highly accurate, high-resolution estimates of AC penetration are essential to determine the contribution of cooling to a region's electricity consumption, as well as make projections of how energy needs may change in the future. However, residential customers have widely varying patterns of demand due to different occupancy patterns, thermal comforts, building characteristics, and appliances [12–15]. Thus, merely knowing if a household has AC, or the number of households in a region with AC, is not enough to model the energy demand of the house or region itself. Instead, knowledge of how residential customers use their AC, in combination with who has AC, is critical to evaluate the electricity demand that is required for cooling.

Exploring patterns of AC ownership and use is difficult due to the shortage of data. Data regarding household appliances is rarely publicly available and information about AC ownership and usage has typically been gathered through state or federal surveys which are both financially expensive and time intensive [7,16,17]. Further, these efforts most often produce AC estimates at large spatial extents, such as statewide or regionally, that do not provide insight into local energy needs. More recently, studies have utilized large scale smart meter data records to study cooling demand, but these methods have shortcomings [6,18–21]. First, several of these studies ignore the impact of electric heaters on the electricity-temperature relationship by either excluding data at lower temperatures or assuming electric heaters are rarely present in the dataset, which can lead to misidentification of AC households [19–21]. Second, most studies utilize daily electricity data, which does not capture intraday patterns of electricity use, and therefore focus primarily on classifying if households have AC rather than analyzing how ACs are used [19,21–23].

In this study, we present a three-part framework to study spatial and temporal patterns of cooling demand in Southern California. In the first part of this study, we develop a novel methodology (referred to as the "AC Ownership Algorithm" for the remainder of the paper) that utilizes hourly smart meter electricity records to identify the presence of AC and electric heat appliances based on the relationship between electricity demand and outdoor temperature. In the second part of this study, we then adapt and apply a linear regression method (referred to as the "AC State Algorithm" for the remainder of the paper) to the identified AC households to make predictions about their hourly AC on/off state. (Note: even though we identify electric heating in this paper, doing so is only to estimate AC penetration more accurately. Hence, we focus our analysis on AC ownership and use characterization, and we do not attempt to characterize electric heating use). In the final step of the methodology, we aggregate and combine estimates of AC ownership and hourly AC states to better understand the regional cooling demand. Through this analysis we answer the following research questions:

1. How do AC penetration estimates produced with a model that utilizes hourly electricity data and considers both electric heating and cooling compare to previous estimates in the literature?
2. Can we use hourly electricity data to identify hours in which customers use their AC?
3. What are the aggregate trends in sub-daily cooling behavior across temporal, climatic, and spatial extents?
4. Can a region's residential cooling behavior be captured through combined estimates of regional AC ownership and patterns of AC consumption?

This framework improves upon previous studies because it can identify electric heating in homes, and hence, does not ignore or require the lack of electric heating to correctly identify AC. The use of hourly data also gives insight into the intraday patterns of households AC consumption, which can better inform grid system planning, energy equity policies, and demand side management.

2. Literature review

As smart meter installations expand and providers make data more accessible, researchers have used the electricity records to make inferences about residential electricity behavior. Specifically, studies have used the relationship between electricity demand and heat metrics to make estimates of which households in a dataset have AC [19–21,24]. These methods are based on the understanding that AC units will consume more energy to cool a space as the temperature increases above a certain threshold; thus, a positively correlated relationship between demand and temperature at higher temperatures will indicate a household with AC.

Several papers have employed a method that screens for whether a household's electricity demand has temperature dependence and determines whether a household has AC based on that dependence. For example, in the first step of a multi-part methodology, Dyson et al. regressed the daily electricity demand of a household against the ambient daily average temperature on days above 55 °F, calculated the slope of the linear model, and asserted that homes with a positive slope (i.e., electricity demand increases with temperature) had AC [20]. The method used in this study is a simple demonstration of the electricity temperature sensitivity concept that underlies many of the studies regarding AC identification and behavior [19,21,23–25].

One limitation of this methodology is that homes without AC that have a very slight temperature dependence (e.g., a home that uses fans during warmer temperatures) could be misclassified as having AC, since there is no minimum slope threshold. Chen et al. developed a more robust methodology to avoid misclassifying these homes that regressed daily average electricity demand against daily average outdoor temperature with a segmented linear regression model [18]. Then, a home was determined to have AC if a) the slope to the right of the stationary point temperature, or SPT (i.e., the outdoor temperature at which a home is expected to turn on their AC if they have it), was greater than zero and b) the sum of the slopes to the right and left of the SPT was greater than zero. The second criterion was included to ensure that homes with a negligible temperature dependence were not identified as having AC, but the rule assumes the household does not have electric heating or rarely uses it. While this is a reasonable assumption in California where a majority of homes are heated with natural gas, electric heating is more common in other regions and will become more common on a future grid with high electrification [26,27]. This methodology was used to make census-level estimates of residential AC ownership across the Southern California region and identify communities that would be most vulnerable to extreme heat and adapted in a later study to test whether humid heat metrics are better indicators of AC ownership [6,19,21].

A study by Elmallah et al. investigated access to both heating and cooling in Northern California through a dataset of ~60,000 households in PG&E territory, addressing the gap in the literature regarding electric heat [28]. In this study, the segmented linear regression model described by Chen et al. was adapted to detect electric or gas heating and electric cooling using both gas and electricity usage records. In contrast to the model used by Chen et al., which used one changepoint (referred to as the SPT) the researchers fit the data to three different linear models in which there were no changepoints, one changepoint, and two changepoints. The different models represent 1) a house without heating or cooling that has no temperature dependence, 2) a house with either heating or cooling that has temperature dependence at either low or high temperatures but not both, and 3) a house with both electric heating and cooling that has temperature dependence at both low and high temperatures. Then, Bayesian Information Criterion was used to select the best model for each household, informing whether the household heats or cools. In all, the study detected gas or electric heating in 68 % of households, and electric cooling in 40 % of households. This study is advantageous because it does not rely on the absence of electric heating to categorize homes, and further explores the distribution of both heating and cooling access.

While these studies were novel in their ability to detect cooling across large spatial extents at high resolutions, they only capture whether a household has an AC unit which is not enough to quantify the cooling demand of a region. Information pertaining to how households use their AC is critical to plan for electricity needs, but acquiring appliance-level data is challenging and thus research related to AC use is even further limited. Studies on household AC use typically analyze data obtained through surveys on household energy use [29–31] or smart meter trials and programs with sub-metered appliances [32–39]. For example, a study related to AC usage in Hong Kong collected questionnaires from ~554 residents which included questions about how many hours and in which months they turned on their AC at night as well as their temperature settings [29]. The results of the study provided insights into the cooling preferences of residents in Hong Kong, but studies that use survey data are only capable of capturing general trends in AC use and are likely imprecise as they rely on customers to accurately report their energy behavior.

Datasets consisting of sub-metered appliance electricity records are advantageous because they can produce a more exact quantification of the cooling patterns of the studied buildings. A study in Sydney analyzed the contribution of ACs to regional summer demand peaks using the Smart Grid Smart City (SGSC) data set which includes appliance level data from 808 homes and found that residential AC contributes up to 9 % percent of total peak demand [38]. However, the authors acknowledge that the size of the dataset is a limitation of the study and may not fully capture the variety of AC load profiles that exist in the study region. In general, a limitation of appliance monitoring datasets is that they consist of a small number of samples (e.g., less than 1000 homes). Thus, until large-scale, sub-metered electricity datasets become available, using appliance monitoring to draw inferences about the cooling behavior of an entire region is not feasible.

The granularity and size of smart meter datasets presents an opportunity to gain insight into patterns of cooling behavior within and across regions. The previously described AC identification studies used smart meter data records but aggregated the records to the daily level. While it has been shown that the correlation between daily electricity records and temperatures is stronger than the correlation between hourly electricity and temperatures [18], the coarse resolution conceals intraday patterns of AC use. Conversely, many studies have used higher resolution data and developed methods to non-intrusively disaggregate appliance level consumption from overall electricity demand, but isolating the AC load from smart meter records is challenging [40]. For example, one study implemented a three-stage load decomposition method that relied on the hourly electricity temperature relationship and building characteristics to separate the AC load from the household's total load [41]. The method was able to accurately estimate the AC load profiles of the households in the dataset, when compared with ground truth appliance-level data. However, researchers with large-scale smart meter datasets typically do not have access to the building characteristics that were utilized in this study.

In the second step of the study by Dyson et al., the authors used smart meter data from 30,000 customers in PG&E's service territory to identify the hours in which a household turned on their AC [20]. Similar to the studies that determine the presence of an AC unit based on temperature and electricity, the household's hourly electricity demand and outdoor temperature were fit to a linear regression model. However, this study utilized hourly measurements to analyze intraday electricity demand and make inferences about AC usage. Each pair of hourly electricity and temperature measurements were fit to either a temperature-independent model or a temperature dependent model with hourly and weekend/weekday fixed effects and reassigned iteratively until the model converged. The authors then calculated the impact that a 4-degree change in the AC setpoint would have on each household's power consumption and aggregated the results to estimate the demand response capacity of customers in PG&E's service territory. This study provided meaningful insight into the extent of grid services and flexibility that residential customers can provide, but the analysis of cooling behavior was limited.

Recent residential cooling demand studies have produced highly resolved estimates of AC ownership across large spatial extents, offering unprecedented insight into regional patterns of AC adoption. However, a majority of these studies rely on daily data to infer which households have AC, thus limiting the knowledge of patterns of AC use. Conversely, studies pertaining to the patterns of AC have thus far utilized small dataset samples that cannot be extrapolated to understand regional cooling usage. Therefore, a research gap exists in the literature as the heterogeneity of AC consumption across spatial, temporal, and climatic extents has not been well explored. In this body of work, we use a large scale hourly smart meter dataset to make highly resolved estimates of household AC ownership and use patterns across the study region of Southern California. The approach taken in our work provides novelty and improves upon previous methods by directly modeling electric heating, analyzing smart meter data at the hourly (rather than daily) level, and combining estimates of AC ownership and hourly operation to characterize cooling demand across a demographically and geographically diverse region.

3. Methodology

In this section, we describe the three-part framework that we develop to characterize cooling behavior in Southern California using smart meter data from ~200,000 residential customers. Before implementing the framework, we carry out several filtering methods and outlier checks to ensure that our dataset only contains households with a sufficient amount of data, and that erroneous values are removed for each separate household. In step 1 of our framework, we use a novel AC identification methodology (the AC Ownership Algorithm) to determine which households have AC. We compare our AC ownership results with the results from Chen et al. [19] and survey data for the same study area [16,42] to analyze the impact that household location and technologies may have on AC identification for each method. In step 2, we employ an AC state model (the AC State Algorithm) to determine in which hours the AC households (determined in step 1) have their AC on. In step 3, we combine our estimates of AC penetration and operation to describe the cooling demand of the study region. This series of steps is summarized in Fig. 1. In this three-part framework, we define and calculate three different metrics that capture different aspects of a region's cooling demand:

- AC Penetration Rate: The estimated percentage of homes in a defined region with AC.
- AC Operation Rate: The estimated fraction of hours out of a defined set of hours (e.g., two-year study period or all hour 12s in a year) for which the AC is active. The rate can be calculated for a single household or as an average of all households in a defined region.

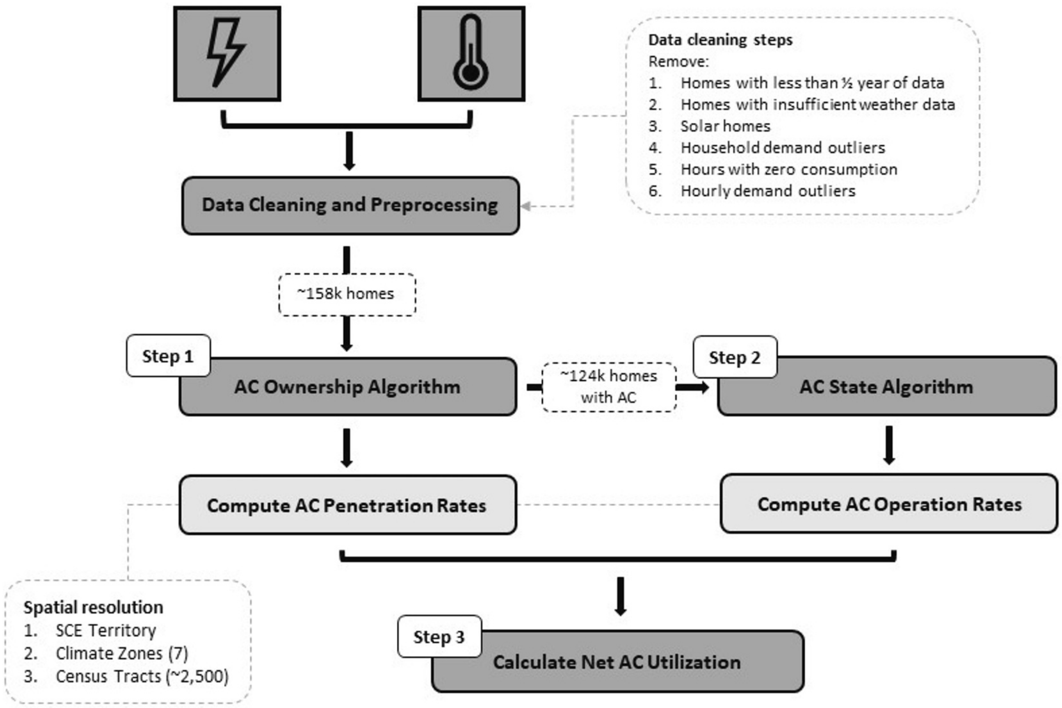

Fig. 1. Overview of methodology for finding AC Penetration Rates, AC Operation Rates, and Net AC Utilization.

- Net AC Utilization: The product of a defined region's AC Penetration Rate and AC Operation Rate.

3.1. Dataset information and preprocessing

The dataset used in this analysis consists of smart meter electricity records measured in 15-min intervals at the household level for 2015 and 2016. The data was provided by Southern California Edison (SCE), an investor-owned utility, and contains data from roughly 200,000 distinct customers identified by SCE as being single-family households. The customers were selected at random to be statistically representative of the 4.5 million households located in Greater Los Angeles at 99 % confidence level. The street address for each customer was also provided, allowing for detailed spatial analysis (e.g., AC use at the census tract level). These dwellings span over ~2500 census tracts and 7 building climate zones, as defined by the California Energy Commission [43], in the Southern California area. As this data is highly confidential, the smart meter records were stored on a high-security data account (HSDA) provided by the University of Southern California to meet the security requirements of SCE.

Prior to applying the AC Ownership and AC State Algorithms, we perform outlier analyses on the aggregate hourly and daily electricity data. A large portion of this outlier analysis follows the steps performed by Chen [19] and Peplinski, et al. [21]. The goal of the preprocessing step is to remove all homes for which there is insufficient smart meter data and remove smart meter records that indicate missing or highly abnormal behavior. We aim to curate a dataset that is both representative of the region and makes it possible to clearly establish the relationship between electricity and temperature at the household level.

First, homes with fewer than 20 kWh of average annual electricity consumption, which is approximately the daily electricity demand of an average California home, are removed as it is likely these homes are uninhabited [44]. Additionally, homes with consumption falling more than three standard deviations above the mean annual electricity consumption are removed as outliers. Next, we filter all homes that are suspected to have solar panels on site to avoid the inconsistencies created by net metering, discussed in greater detail in the work by Chen [19] (we estimate that less than 2 % of homes in our dataset have solar panels). For the remaining homes, we aggregate the 15-min smart meter data to the hourly level and drop all hours for which the electricity consumption is zero. For a smart meter to give a reading of zero across an hour, the home would have to either be disconnected from the grid due to long-term vacancy or power failure, or possess solar panels that cause a meter read of zero due to net-metering. In either case, the hours in question would not reflect the customer's typical consumption patterns, which may interfere with the analysis of AC ownership and use. Note that temporary vacancy would be highly unlikely to give a meter read of zero due to plug loads like refrigerators.

Next, we match each individual household to weather stations within a 20-mile radius, using data from 102 weather stations within three different land-based weather station systems [45–47]. For each household and each hour, the temperature of the nearest weather station that has a temperature reading in that hour is assigned to the household. If no weather station within 20 miles has a temperature reading, the hour in question is removed from the household's data due to an inability to establish an electricity-temperature relationship. Next, to eliminate hours with extreme levels of electricity consumption, we bin electricity data into ten temperature quantiles and remove hours for which consumption exceeds 5 standard deviations above the mean within said quantile. This eliminates hours with highly irregular electricity consumption, caused by an unexpected load, that would distort the relationship between electricity demand and ambient temperature. After performing this hourly filtering, we drop any homes for which less than 4380 hourly records remain (one half of a year) to ensure sufficient data to perform the AC Ownership and State Algorithms. At the end of this outlier removal process, we retain ~160,000 households from 2439 census tracts and four counties across Southern California Edison's service territory.

3.2. AC Ownership Algorithm and computation of AC Penetration Rate (Step 1 in Fig. 1)

We determine whether each household in the filtered data set has an

AC unit and/or electric heater by examining the relationship between hourly electricity consumption and hourly outdoor temperature. For households with an AC, we expect that there is a positive correlation between hourly electricity consumption and hourly temperature above a certain temperature threshold, but this relationship is dependent on the operational status of the AC in a specific hour. For example, a household with an AC may turn it off when unoccupied, so high-temperature hours will only display temperature dependence within the subset of hours for which the AC unit was running. Similarly, homes with electric heating should display temperature dependence at temperatures below a specific temperature threshold, but only for the fraction of hours for which the electric heater was in use. Given that electricity consumption depends on many loads that are not related to temperature (e.g., cooking, entertainment, household chores) and will therefore depend on individual user behavior, we must account for this temperature-independent electricity consumption before analyzing the temperature-dependent AC and electric heating loads. We do this by subtracting an estimate of the typical temperature-independent load for each hour (that is, the expected electricity consumption of non-AC or electric heating loads). For each home, we group the electricity data by hour of the day and day type (weekend vs weekday) and find the 25th percentile of electricity consumption for each group. We then subtract the corresponding 25th percentile value from each hourly electricity record (note: this leads to some hourly electricity consumption records being negative, as shown in Fig. 2). We use the 25th percentile, rather than the 50th percentile, as an estimate of the temperature-independent consumption to account for the fact that some hours will feature a significant amount of AC and/or electric heating use that skews the distribution.

We then fit each user's adjusted data to four models that relate electricity consumption to temperature with each model representing an AC and electric heating technology combination. Examples of homes that demonstrate good fits for each of the above models are shown in Fig. 2. We refer to temperatures at which heating or cooling behaviors may change as stationary point temperatures (SPT). For example, the cooling SPT is the temperature above which there is a possibility of AC use.

- Model 1: All data is fit to one horizontal line, implying that electricity consumption is independent of temperature. This model corresponds to no electric heating or AC (top left quadrant).
- Model 2: A portion of the data is fit to one line representing the temperature-independent portion of the load, and, at temperatures above a cooling SPT, a portion is fit to an additional line representing the hours that demonstrated a temperature-dependent load due to AC usage. This model corresponds to a residence with AC but no electric heating (top right quadrant).
- Model 3: A portion of the data is fit to one line representing the temperature-independent portion and, at temperatures below a heating SPT, a portion is fit to an additional line representing the hours that demonstrated a temperature-dependent load due to

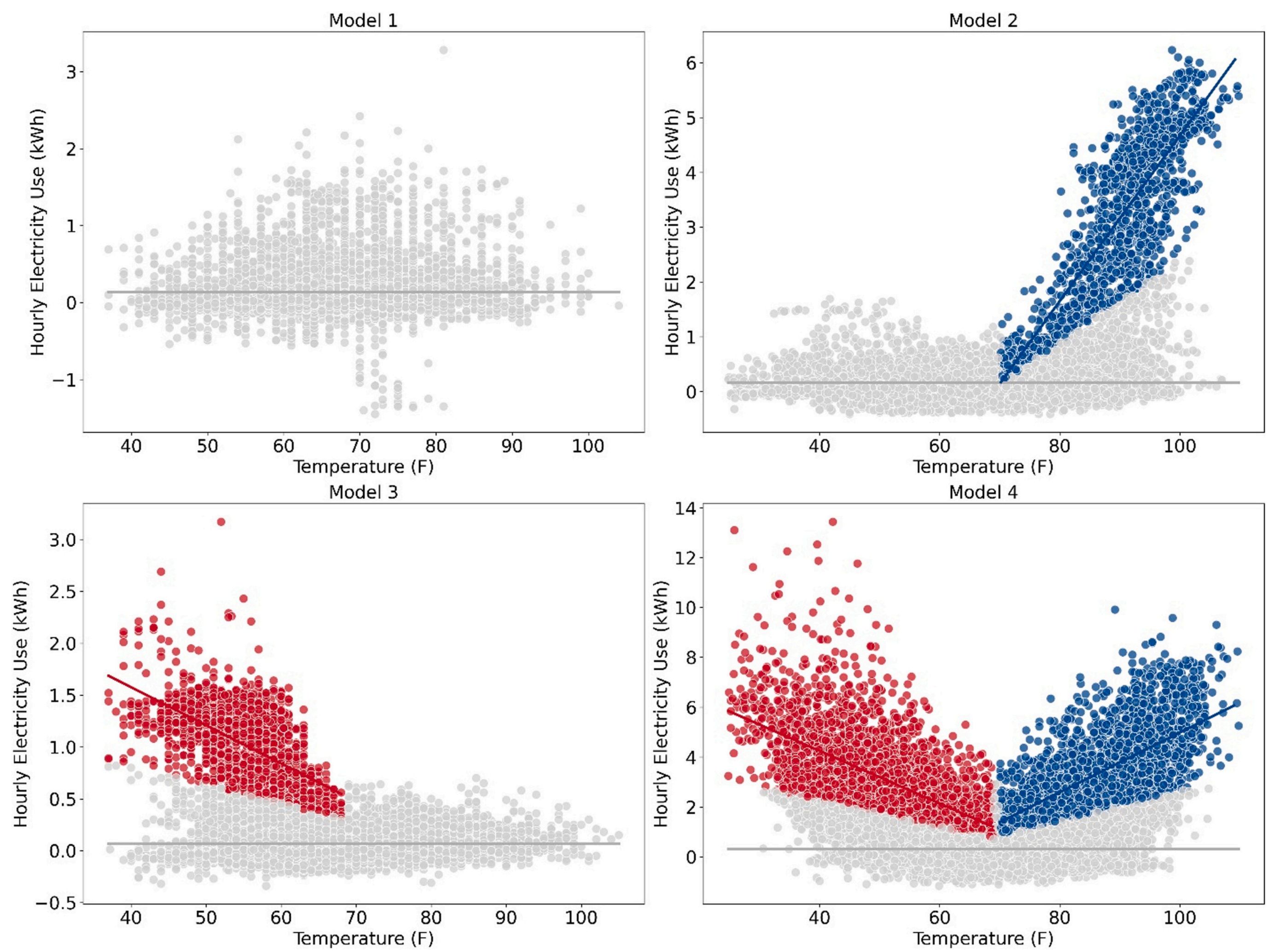

Fig. 2. Hourly electricity consumption versus hourly ambient temperature for four example homes in Southern California over the two-year period. Each plot depicts one of the four AC and electric heating (EH) technology combinations: Model 1) no AC or EH, Model 2) AC no EH, Model 3) EH no AC, and Model 4) AC and EH.

electric heating usage. This model corresponds to a residence with electric heating but no AC (bottom left quadrant).

- Model 4: A portion of the data is fit to one line representing the temperature-independent portion of the load and the remaining data is fit to one of two temperature-dependent lines, with one line for temperatures below the heating SPT and one for temperatures above the cooling SPT. This model corresponds to a residence with electric heating and AC (bottom right quadrant).

For all the above models, every individual datapoint (i.e., hour) is fit to exactly one line. For Model 1, there is a single temperature-independent line that has a constant y-value equal to the mean electricity consumption of all points. However, for Models 2–4, we determine in which hours the electricity demand exhibits temperature dependence and the lines of best fit for temperature dependent and temperature-independent hours using a version of the expectation-maximization (EM) algorithm. The EM algorithm involves iteratively classifying datapoints to groups and then fitting models of those groups until a condition is met.

In Model 2, we assume that for temperatures above the cooling SPT there is a possibility that the AC will be running and therefore that these hours can demonstrate temperature dependent or temperature-independent electricity consumption. To begin the EM algorithm, we first assume that all hours with temperature above the 70th percentile and electricity consumption above the 70th percentile of this subset of hours are temperature dependent, and all other hours are temperature independent (though the results of this algorithm were not noticeably sensitive to different initial seedings). The temperature-independent line is then defined by the mean electricity consumption for all points assigned to it, and the slope of the temperature-dependent line is calculated via a non-negative linear regression of electricity consumption on temperature for all hours assigned to it. All hours above the cooling SPT are then reassigned to the two lines depending on error minimization, and the models are refit with the newly assigned hours. This process continues iteratively until fewer than 1 % of eligible points switch line assignment or until fewer than 1 % of the total hours are assigned to the temperature-dependent line. We test potential cooling SPTs of integers ranging from 60 to 100 °F to cover a large range of potential cooling preferences and select the SPT that minimizes the total error. Model 3 proceeds identically to Model 2, but the classification of points occurs at temperatures below the heating SPT, and the search space for the heating SPT ranges from 40 to 70 °F.

For Model 4, a household's data is split into two portions based on the midpoint of the heating and cooling SPTs found by Models 2 and 3, and then the algorithm described above is repeated for each portion of data with the midpoint serving as the lowest possible cooling SPT and the highest possible heating SPT. The temperature-independent line is again set as the mean of all points not assigned to the temperature-dependent lines, regardless of temperature.

For each of the four models, the model error is determined by the mean-squared error for the lines of best fit multiplied by the number of lines fitted (one line for Model 1, two for Models 2 and 3, and three for Model 4). The multiplier on the mean-squared error penalizes more complex models that would otherwise generally have lower error (similar to error terms used in information criterion analysis [48]). We fit all four of the models to each home, and then select the model that minimizes this custom error function. Homes for which Models 2 or 4 were selected are considered to have AC, and homes for which Models 3 or 4 were selected are considered to have electric heating. This algorithm is designed to capture the general relationship between electricity consumption and temperature of specific households, which gives insight to their space conditioning technologies, and not to minimize model error or most-accurately describe their heating or cooling demand.

To characterize AC ownership across our region, we match each household to a census tract and a California building climate zone using shapefiles from the US Census Bureau [49] and California Energy Commission [43]. For each respective census tract and climate zone, we calculate the AC Penetration Rate by dividing the number of homes in the area identified as having AC by the total number of homes in the region present in our dataset. We compare the results of this methodology to the results found by Chen et al. [19] at the census tract level for Southern California and our aggregated results to survey data collected in the region. Following the filtering steps, the remaining records are statistically representative of 1534 census tracts.

3.3. AC State Algorithm and computation of AC Operation Rate (Step 2 in Fig. 1)

For the subset of households designated as having AC, we proceed with a more fine-tuned algorithm to determine the cooling SPT and the hours during which the AC is on. While the AC Ownership Algorithm (described in Section 3.2) aimed to establish general electricity-temperature relationships for the purpose of identifying the presence of electric heating and cooling technologies, here we use the AC State Algorithm, adapted from Dyson et al., to establish specific cooling behaviors and parameters [20]. This includes the cooling SPT (i.e., the temperature at which people begin to turn their AC on) and a more precise prediction of which hours feature AC activity. The AC State Algorithm classifies every hour as being "AC on" or "AC off" even though during an "AC on" hour the AC may not be running continuously throughout the hour.

In this method, each home is fit to a multiple-linear regression model that regresses electricity consumption on temperature and dummy variables that represent the interactions of day type (weekend vs weekday), hour of day, and a binary variable that classifies each hour as a high-temperature or low-temperature hour. The temperature dependent portion of the model is again conditional on the state of the heating and cooling technologies and is only defined for specific temperature ranges.

$$E_t = \boldsymbol{D}_{t\in \boldsymbol{h},\boldsymbol{w},\boldsymbol{s}} + H_t(\beta_1 \times (SPT_H - T_t) + i) + C_t(\beta_2 \times (T_t - SPT_C)) + j) \quad (1)$$

In Eq. 1, the electricity consumption during a one-hour time period t (E_t) is determined by the AC state (C_t), the electric heating state (H_t), and a vector of dummy variables ($\boldsymbol{D}_{\boldsymbol{h},\boldsymbol{w},\boldsymbol{s}}$) that specify the fixed impact of the combination of the hour of the day ($\boldsymbol{h}$) and the day type ($\boldsymbol{w}$, weekday vs weekend), which are further split into high-temperature vs low-temperature hours ($\boldsymbol{s}$) with a threshold of 60 °F. These dummy variables account for non-temperature-dependent loads, such as cooking and entertainment-related electricity consumption, that occur with different frequency depending on the time of the week and the weather (for example, people are likely to be home consuming some electricity at 8 pm on the weekdays during colder weather). If the AC is classified as on during time t, the electricity consumption depends on the electricity-temperature sensitivity for cooling (β_2) multiplied by the difference between the temperature and the cooling SPT (SPT_C), and an AC intercept (j). Similarly, at temperatures below the heating SPT we assume that electric heating could be on, and that electricity consumption therefore depends on the electricity-temperature sensitivity for heating (β_1) and a separate heating intercept (i). For homes that were classified as having AC but no electric heat, the heating state of all hours was set to zero. The heating and cooling intercepts can be interpreted as the minimum additional electricity consumed when the AC or electric heating is on and the electricity-temperature sensitivities can be interpreted as the increase in electricity consumption that occurs as outdoor temperature increases when the AC is on, or as temperature decreases when the electric heating is on.

For each hour, we again determine the AC state (on/off) and electric heating state (on/off) via the EM algorithm that was used in the AC Ownership Algorithm; we iteratively find lines of best fit for each state, and then reassign hours based on minimizing the prediction error. We use the same initial seeding from Section 3.2 for AC-on hours, and again terminate this algorithm when fewer than 1 % of eligible points switch AC state or when fewer than 1 % of total points are classified as AC on. Fig. 3 illustrates the results of this model for an example home with a SPT of 81 °F. We show results for four hours of the day during the week with the hourly AC and non-AC points indicated. Note that across all hours the AC intercept and temperature sensitivity are constant, which assumes that AC consumption is linearly dependent on changes in temperature regardless of time of day or current temperature (provided the temperature is above the SPT). With hour of the day included as a variable in the regression, an 8 am datapoint may be classified as AC on despite having a lower electricity consumption than a 4 pm datapoint that is classified as AC off because the 8 am datapoint represents unusually high electricity consumption for that time of day and day type.

To estimate the cooling SPT, we look for a temperature that both reduces error and increases the probability of correctly classifying the state of the AC. A lower cooling SPT generally reduces the error term because more of the data is fit to two lines instead of one. Conversely, higher cooling SPTs generally lead to more confident predictions of the AC state, since at higher temperatures there is typically a higher fraction of hours classified as AC on and a larger magnitude difference between the electricity consumption of an AC-on versus AC-off hour. We balance these two objectives through an error term that combines the prediction likelihood and the probability of AC being on, which is defined as the fraction of hours with temperature above the cooling SPT that are classified as AC hours. We test each cooling SPT between 60 and 100 °F and find one value for each household that minimizes the error term. More discussion of the SPT selection method can be found in [20]. We used a fixed heating SPT of 60 °F for the electric heating system because we are not interested in identifying specific electric heating behaviors.

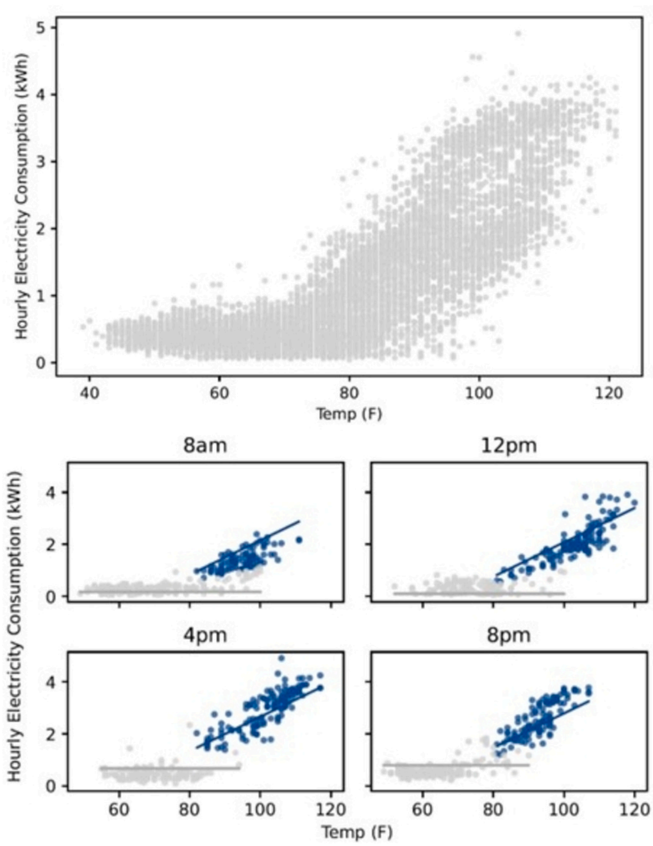

Fig. 3. Top: Scatterplot of hourly electricity consumption and temperature for the two-year period for one household. Bottom: Results of the AC State Algorithm for four different weekday hours over the two-year period. For this home, we determined a cooling SPT of 81 °F, hence only hours with temperature above 81 °F can be classified as AC on.

We note that it is necessary to include this temperature dependence below 60 °F to avoid the errors at low temperature hours dominating the total error and therefore skewing the cooling SPT selection process. With the optimal cooling SPT selected, we can make a final classification of the hours during which a household's AC is on.

We then determine a household's AC Operation Rate, which is the fraction of hours that a home has its AC on (the number of hours classified as AC on divided by the total hours). Recall that an "AC on" hour is an hour that demonstrates clear temperature dependence, and the classification does not capture the number or length of AC cycles that occur during the hour. Following the same method of matching households to census tracts and climate zones as was used for AC Penetration Rate, we also find the average AC Operation Rate for a region by taking the mean of the AC Operation Rate for each household in the region. The household and average AC Operation Rates can be calculated for the entire study period or a subset of time (e.g., the AC Operation Rate in all hour 12s).

3.4. *Calculation of Net AC Utilization (Step 3 in Fig. 1)*

Finally, we use our regional estimates of AC Penetration Rate and AC Operation Rate to calculate the Net AC Utilization for a region as shown in Eq. 3.

$$Net\_AC\_Utilitzation_R = AC\_Penetration\_Rate_R \times AC\_Operation\_Rate_R \quad (3)$$

Eq. 3 defines the Net AC Utilization of a region *R* as the product of the region's AC Penetration and Operation Rates. We find Net AC Utilization for the entire study region, as well as for each census tract and climate zone within the study region. Net AC Utilization is directly proportional to the number of per-household "AC on" hours in a region and better describes the AC use in a region than estimates of AC ownership or state in isolation.

4. Results and discussion

4.1. *Comparison of AC penetration rates with other studies*

Across the entire study region, AC was detected in 79 % of households. In the California Residential Appliance Saturation Survey (RASS), 75 % and 86 % of customers surveyed in SCE territory reported having central or room AC in the years 2009 and 2019, respectively. Our region-wide estimate is in alignment with the survey results considering that the smart meter records analyzed in this study (2015–2016) fell in between the survey years, [16,42].

This study's estimate of the region's overall AC Penetration Rate is significantly higher than the value found in Chen et al. (69 %) [19]. There are multiple explanations that account for the difference. First, the methodology used in this study to classify AC households does not make assumptions regarding the electric heating status, and thus, is more likely to correctly identify homes that have and use both electric heating and cooling systems. Second, we expect this methodology to better capture households that use their AC infrequently and/or have other electric loads that contribute significantly to total demand, diluting the electricity-temperature signature at the daily level. This theory is in part validated by the breakdown of central versus room AC units reported in the RASS (58 %/18 % in 2009 and 68 %/18 % in 2019), indicating that the daily methodology utilized by Chen et al. might have accurately captured the central conditioners but failed to identify the room conditioners with smaller loads.

To gain an understanding of spatial differences in the results from this study and Chen et al. [19], the difference in AC Penetration Rate estimates in each census tract was plotted on a choropleth map shown in Fig. 4. The areas in blue were estimated to have higher AC Penetration Rates when the proposed hourly method was used in place of the previous daily method, while areas shown in red were estimated to have

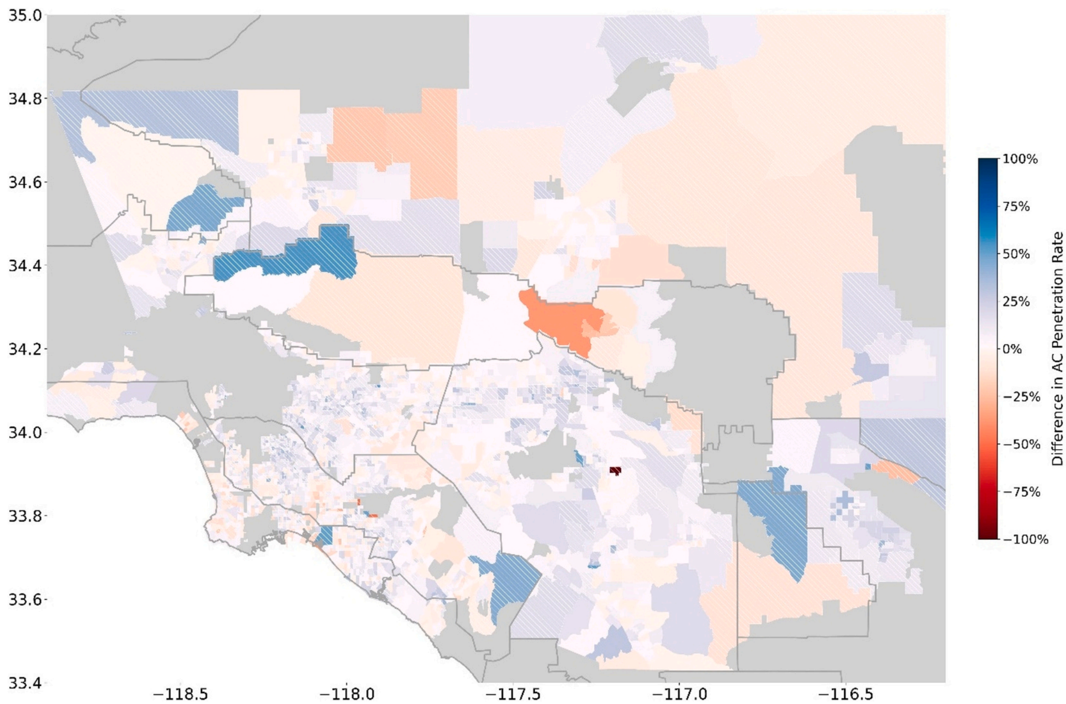

Fig. 4. Choropleth maps depicting the difference between census tract level AC Penetration Rates estimated with the hourly method proposed method in this study and the daily method developed by Chen et al. [19]. Generally, the AC Penetration Rate computed with the new method is higher (blue) than when the previous method was used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lower AC Penetration Rates. This study's method of detecting AC found a higher penetration in the majority of census tracts across the region. We note that the households studied in this analysis were not evenly distributed across census tracts, and thus some of the census tracts that show large differences between methods in Fig. 4 are the result of having a small number of homes in that specific census tract. (Census tracts that are not statistically represented are indicated in Fig. 4 with cross hatching.)

Through this methodology, we also estimated that 25 % of households have electric heating. In the 2009 RASS, only 4 % and 1 % of customers reported an electric heater as their primary and auxiliary space heating appliance. The percentages of primary and auxiliary space heating appliances increased to 17 % and 6 % in the 2019 RASS. Since our smart meter records fall in between the survey years, the methodology used in this study likely overestimates the portion of electric heaters present in Southern California. One explanation for the discrepancy in values is that households may be supplementing their natural gas heating with electric room space heaters that were not surveyed in RASS. Elmallah et al. detected electric heating in 27 % of homes, which was higher than the value reported for RASS in some of climate zones located in their dataset; similarly, the authors pointed to the use of room space heaters as an explanation [28]. It is important to note that Southern California has more cooling degree days (CDDs) than heating degree days (HDDs) [50], and demand for space conditioning is driven by cooling needs rather than heating needs.

4.2. Tracking temporal patterns of AC operation rate

One of the major advantages of the method described in this study is the ability to track patterns of AC operation. While the AC Penetration Rate is an important metric to characterize who has access to AC in a community and inform where the power grid may experience spikes in demand, information about how people use their AC is also necessary to quantify cooling demand. Here, we analyze variations in customer AC Operation Rate, including how often and when their AC unit is on, and explore how these behaviors create regional differences in cooling behavior.

The results of this study found the average customer with AC has an AC Operation Rate of 8.3 % calculated across the entire two-year study period. The bar chart shown in Fig. 5 depicts how AC Operation Rates vary across and within the climate zones in SCE's territory. In the cooler, coaster climate zones (e.g., climate zones 6 and 8) the AC Operation Rate across the full study period is generally lower than for customers in the hot, desert climate zones (e.g., climate zones 14 and 15). For example, in climate zone 15 which is characterized by a hot, desert climate, only 5 % of customers have an AC Operation Rate less than 3 %, compared to 23 % of customers in the coastal climate zone 6.

In addition to knowing how often utility customers use their AC, we can capture the timing of when customers use their AC and how that varies across the region. The heat maps in Fig. 6 show the average AC Operation Rate in each hour and month combination for each of the study region's seven climate zones. Across all climate zones, the AC Operation Rate is higher in the afternoon and early evening, as well as in the hot, summer months. In climate zones that experience relatively cool temperatures (e.g., climate zones 6 and 8) the range of hours and months with notable AC Operation Rates is smaller, and the AC Operation Rate itself is, in those time periods, generally lower than in the hotter, desert climate zones, such as 14 and 15.

4.3. Spatial trends in AC penetration rate, AC operation rate, and net AC utilization

To observe how trends in cooling behavior vary across the study region, study results were aggregated to the census tract level. In Fig. 7, panel a) depicts AC Penetration Rates, lending insight into which areas have higher rates of AC ownership. Panel b) displays AC Operation Rates, which measure how often the average customer in each census tract used their AC during the study period. In general, the cooler, coastal and mountainous regions have lower rates of ownership and use their AC less frequently than the hotter, inland and desert regions (also shown in Section 4.2).

While AC Penetration Rates and AC Operation Rates separately

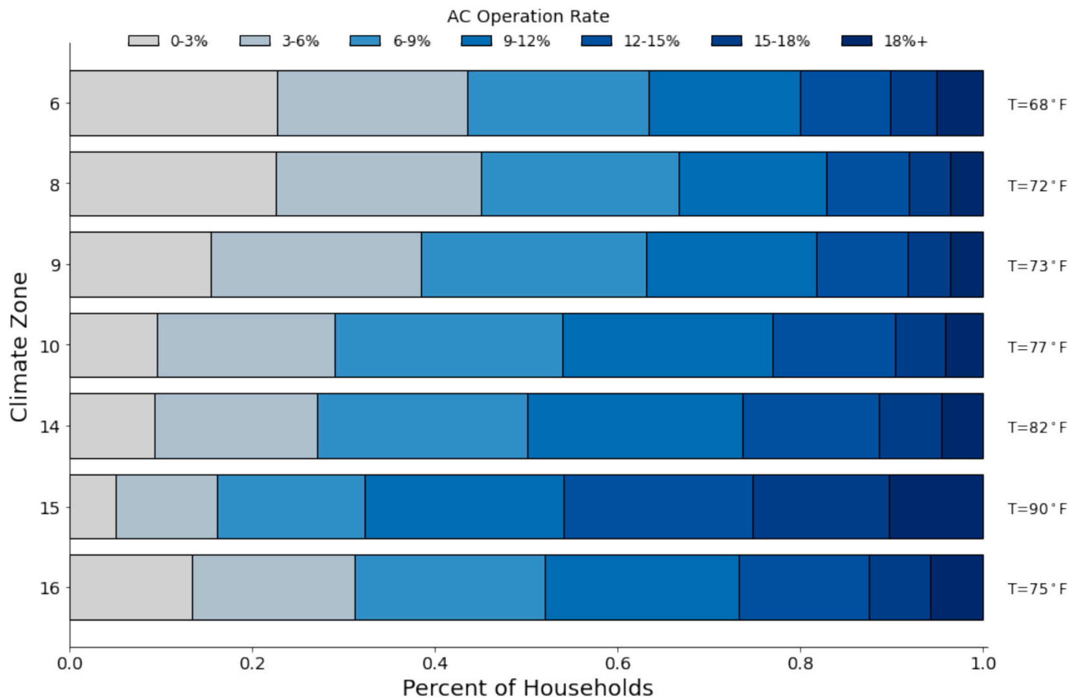

Fig. 5. Stacked bar chart showing the breakdown of AC Operation Rates over the study period for each climate zone. Each bin represents an AC Operation Rate range, with darker shades of blue indicating a higher AC Operation Rate (e.g., AC is classified on for more hours). The summer mean temperature for each climate zone is shown to the right of each bar. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

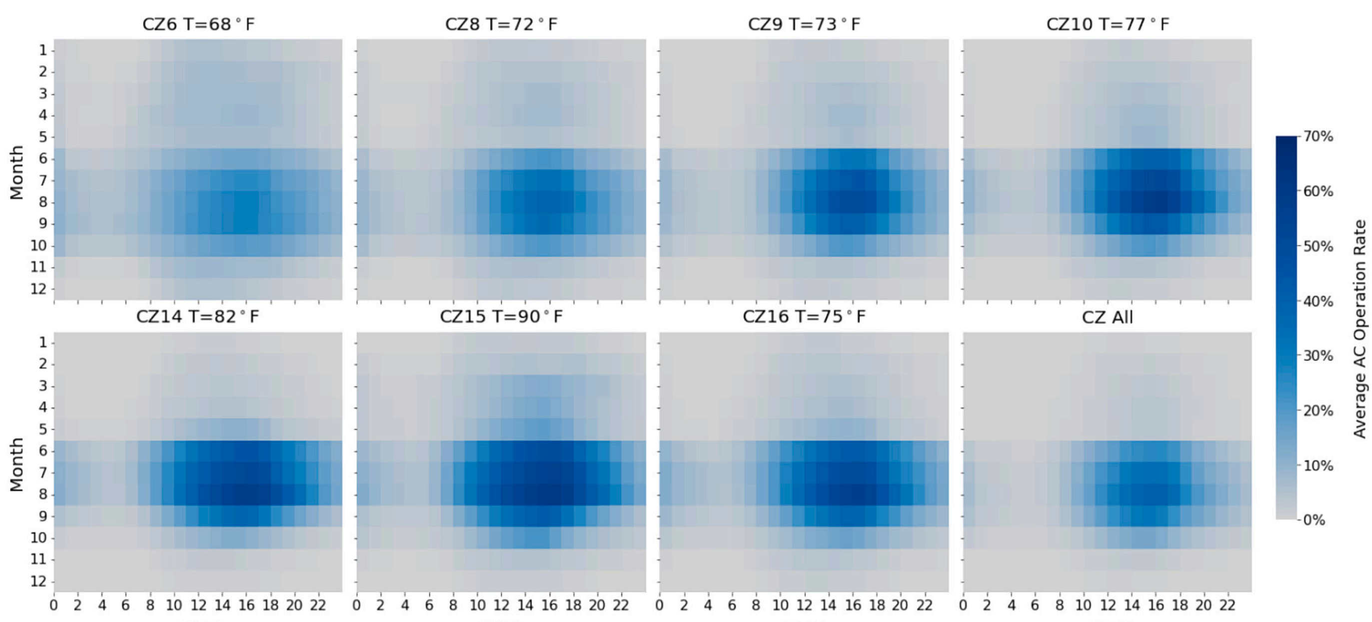

Fig. 6. Heat map depicting the average AC Operation Rate of each day and month of the year combination for a-g) each climate zone and h) full study region. The AC Operation Rate is averaged across all pertinent customers that were identified as having AC. The summer mean temperature for each climate zone is shown above each subplot.

provide important information about the cooling demand of a community, we can better estimate the locations that are likely to have high cooling demand by combining these factors into one metric. Thus, Net AC Utilization, which accounts for both the percentage of households in a specified area that have AC and how often those customers have their AC on was computed for the entire study region by census tract, with the results shown in Fig. 7, panel c). The Net AC Utilization of a census tract is directly proportional to the expected number of hours of AC use that an average household selected from our data in that census tract would have and thus is useful for evaluating local cooling need and the location of demand surges during extreme heat events. In Fig. 7, we report Net AC Utilization by decile because there is not a clear physical meaning of the metric as a percentage value (in contrast to AC Penetration and AC Operation Rates).

In general, the regional patterns are consistent across each of the panels shown in Fig. 7, meaning areas with higher AC Penetration Rates

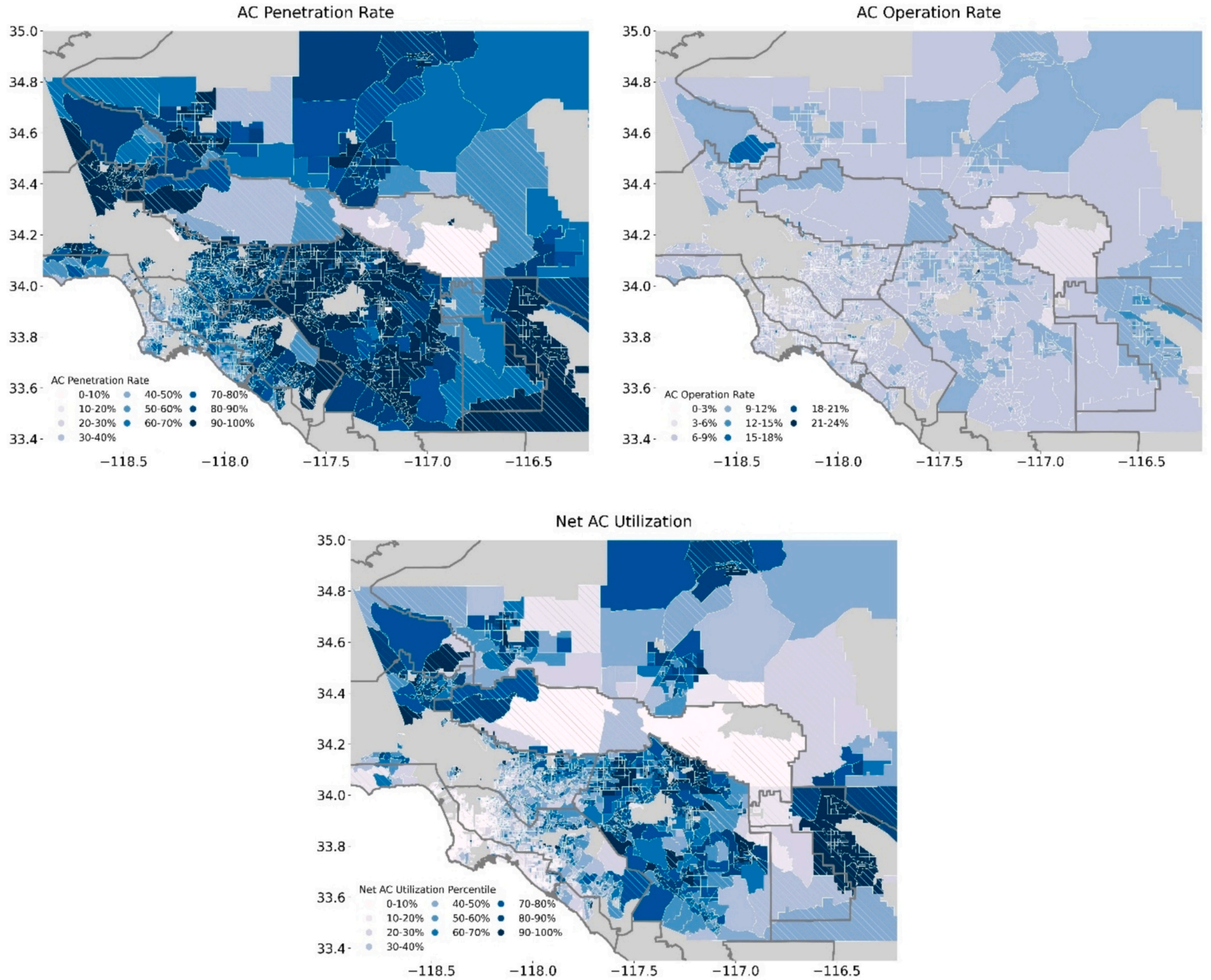

Fig. 7. Choropleth maps depicting the a) AC Penetration Rate, b) AC Operation Rate, and c) Net AC Utilization (binned by decile) computed at the census tract level.

also have higher AC Operation Rates and Net AC Utilization. Although the regional trends are consistent, there are still census tracts where the AC Penetration Rate is relatively high, but the AC Operation Rate is relatively low (and vice versa), which demonstrates the limitation of relying on AC Penetration Rates alone when evaluating cooling demand.

If we compare the AC Penetration Rates and Net Utilization Rates, we can see how incorporating the AC Operation Rates impacts our evaluation of cooling demand. Table 1 provides the percentage of census tracts at each quantile of AC Penetration Rate that fall into each quantile of Net AC Utilization. For example, of the census tracts in the 20–40 % percentile of AC Penetration, it is more likely that they fall into a lower percentile of Net AC utilization than remain in the 20–40 % percentile range. This could be explained by the fact that these census tracts experience cool enough temperatures that they rarely need to use their AC, or that they are lower-income census tracts within that quantile that are more conscious of their electricity consumption.

A second interesting insight is that while most census tracts with high AC Penetration Rates also have high AC Operation Rates, roughly 23 % of census tracts in the top quantile of AC Penetration Rate shift into the bottom two Net AC Utilization quantiles. This could be explained by high-income census tracts that own ACs despite living in relatively cooler climates, thus not requiring cooling often, or low-income census tracts in hot regions where households forgo cooling to lower electricity

Table 1
A transition matrix summarizing the AC Penetration Rates and Net AC Utilization percentile ranks of the census tracts in the study region, where 0–20 % indicates the lowest and 80–100 % indicates the highest AC Penetration Rate/Net AC Utilization quantile. Each value represents the percent of census tracts that originally fell in each AC Penetration Rate quantile (denoted by row) that shift into the specified Net AC Utilization quantile (denoted by column), effectively showing the impact that including AC Operation Rates has on the cooling demand evaluation.

		Net AC Utilization Percentiles				
		0–20 %	20–40 %	40–60 %	60–80 %	80–100 %
AC Penetration Rate Percentiles	0–20 %	49 %	26 %	15 %	7 %	3 %
	20–40 %	29 %	25 %	21 %	15 %	10 %
	40–60 %	18 %	21 %	22 %	20 %	18 %
	60–80 %	13 %	18 %	21 %	24 %	24 %
	80–100 %	9 %	14 %	19 %	25 %	33 %

costs despite high temperatures. The results of this section suggest that AC Penetration and AC Operation Rates are not always tightly correlated and warrants a further analysis of what factors cause diverging results in some regions, as those populations may either be underserved or consume a disproportionate amount of electricity making them a target for grid flexibility efforts.

4.4. Net AC utilization considering climate

While Net AC Utilization provides a useful metric of existing cooling demand in a region, we are also interested in the relationship between a household's theoretical need for cooling and their actual AC behaviors. We approximate a single household's theoretical cooling need by aggregating their hourly temperatures to the daily level and calculating their annual CDDs. In Fig. 8, we plot the mean household Net AC Utilization against the mean household CDDs at the census tract level.

We see that for a given number of CDDs, there is a large variety in the degree of Net AC Utilization across census tracts. This is of particular note for census tracts with a high number of CDDs, and thus a high theoretical cooling need, but a low Net AC Utilization. For example, there are 41 census tracts that rank above the 80th percentile of CDDs but fall below the 50th percentile of Net AC Utilization. These census tracts may be experiencing energy insecurity due to poor access to AC or lack the financial resources needed to use the AC that they do have (although there are confounding factors unrelated to enery insecurity, such as AC efficiency and a building's thermal properties, that can influence AC use). Additional analysis is needed to determine if these census tracts are particularly vulnerable to extreme heat. Lastly, a small number of census tracts display high Net AC Utilization despite relatively low theoretical cooling need, which may represent an opportunity for targeted demand response programs.

5. Conclusion

In this three-part framework, we first developed a novel methodology for identifying the presence of AC from household-level smart meter data and used the model to compute regional AC Penetration Rates. Unlike previous methods, our novel model used hourly, rather than daily, electricity consumption data and directly modeled electric heating, which was a confounding or ignored variable in several previous studies. We believe our focus on hourly data allowed us to better identify homes with a variety of AC types and with intermittent AC use and find that our results align well with survey data from similar years in the same region. In the second part of this study, we predicted the hourly AC state at the household level using the AC State Algorithm and aggregated the results to observe trends in AC Operation Rates across spatial, temporal, and climatic ranges. Finally, we combined AC Penetration and AC Operation Rates to calculate each census tract's Net AC Utilization and better characterize regional residential cooling behavior.

Unsurprisingly, we find higher rates of AC Operation Rates in the middle of the day and afternoon of summer months. We also find that some census tracts have surprisingly low Net AC Utilization when compared to adjacent areas and when compared to the amount we would expect for an area with significant climatic need for cooling. This phenomenon may be explained by the demographic or economic traits of the census tract (which is beyond the bounds of this analysis). Regardless of the cause, these areas would likely benefit from programs designed to increase AC access and/or address energy insecurity. In future work, we plan on conducting a more rigorous analysis of the factors that drive disparities in the cooling demand. For areas that already have high AC Penetration Rates and AC Operation Rates, these census-level estimates increase our understanding of where surges in demand are likely to occur during extreme heat events and high temperatures which is useful information for utilities and grid planners.

The authors would like to acknowledge several limitations of this study. First, there is no ground truth data of AC ownership or operation at the household level with which to validate our results, thus we cannot determine the accuracy of our algorithms that were used to determine the AC Penetration and AC Operation Rates. Furthermore, comparisons between methods also cannot speak to whether one method is more or less accurate for our dataset. Instead, we focus on comparing our AC Penetration estimates with relevant survey data for the same region. This study would also benefit from a sensitivity analysis that examined how algorithm specifications such as the range of potential STP temperatures, the error formula used, and the estimate of temperature-independent load impact the AC Penetration and AC Operation Rates. Unfortunately, the extensive runtime of the algorithms on our computational resources makes parametric analysis impractical.

While our dataset contains nearly 160,000 homes after filtering, the large spatial extent of the data spreads these homes across many census tracts and creates a large range in the number of homes per census tract. As a result, the samples of homes in this dataset are only statistically representative for ∼63 % of the census tracts. We believe our general method of relating electricity consumption and ambient temperature at

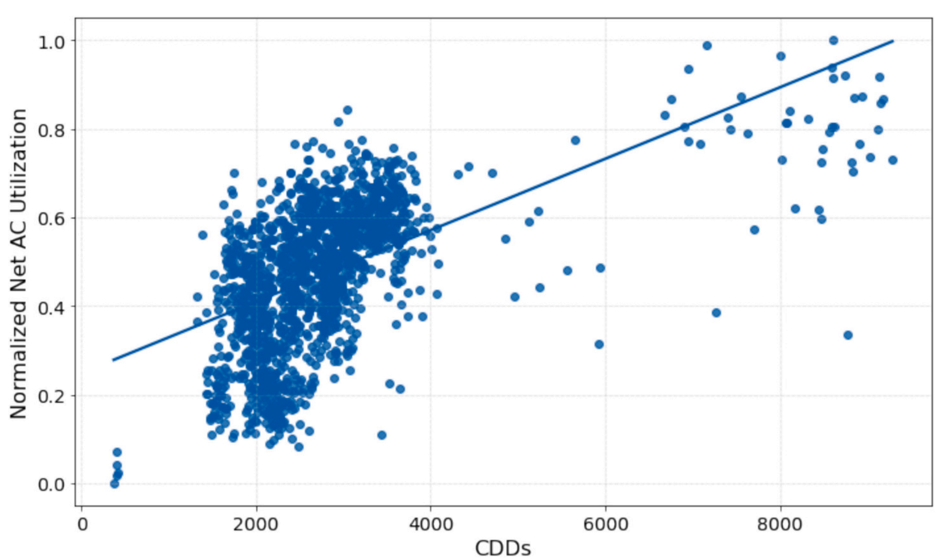

Fig. 8. Scatter plot of normalized Net AC Utilization versus CDD experienced during the study period averaged by census tract. Census tracts that are not statistically represented by the households in our dataset are not included.

the hourly level with models that represent distinct electric space conditioning technologies and usage patterns can be extrapolated to other regions. However, in other regions, the different climatic factors and relative frequencies of a variety of space heating and cooling technologies may require modifications to the methodology presented here. For example, studies of other areas may find that a humid heat metric is more closely related to AC ownership and use than temperature alone. As large smart meter datasets become more widely available, these algorithms should be repeated on a variety of climate zones and populations. Lastly, in this study we discuss Net AC Utilization as a way to characterize AC behavior, but we acknowledge that a more complete study of cooling demand would consider the magnitude of AC electricity consumption, which is beyond the bounds of this analysis.

CRediT authorship contribution statement

McKenna Peplinski: Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Stepp Mayes:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Kelly T. Sanders:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Kelly T Sanders reports financial support was provided by National Science Foundation. Kelly T Sanders reports financial support was provided by Jet Propulsion Laboratory. Please note that the views herein do not necessarily reflect the views of any current or past employer.

If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

[1] IEA. The future of cooling, Paris [Online]. Available: https://www.iea.org/reports/the-future-of-cooling; 2018.
[2] Luber G, McGeehin M. Climate change and extreme heat events. Am J Prev Med 2008;35(5). https://doi.org/10.1016/j.amepre.2008.08.021.
[3] Barreca A, Clay K, Deschenes O, Greenstone M, Shapiro JS. Adapting to climate change: the remarkable decline in the US temperature-mortality relationship over the twentieth century. J Polit Econ 2016;124(1). https://doi.org/10.1086/684582.
[4] Meehl GA, Tebaldi C. More intense, more frequent, and longer lasting heat waves in the 21st century. Science (80-) 2004;305(5686). https://doi.org/10.1126/science.1098704.
[5] Mastrucci A, Byers E, Pachauri S, Rao ND. Improving the SDG energy poverty targets: residential cooling needs in the global south. Energ Build 2019;186. https://doi.org/10.1016/j.enbuild.2019.01.015.
[6] Chen M, Ban-Weiss GA, Sanders KT. Utilizing smart-meter data to project impacts of urban warming on residential electricity use for vulnerable populations in Southern California. Environ Res Lett 2020;15(6). https://doi.org/10.1088/1748-9326/ab6fbe.
[7] EIA (Energy Information Administration). 2020 Residential Energy Consumption Survey (RECS) [Online]. Available: https://www.eia.gov/consumption/residential/data/2020/; 2020.
[8] Pietro Colelli F, Wing IS, De Cian E. Air-conditioning adoption and electricity demand highlight climate change mitigation–adaptation tradeoffs. Sci Rep 2023;13(1). https://doi.org/10.1038/s41598-023-31469-z.
[9] Auffhammer M, Baylis P, Hausman CH. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. Proc Natl Acad Sci USA 2017;114(8). https://doi.org/10.1073/pnas.1613193114.
[10] Waite M, Cohen E, Torbey H, Piccirilli M, Tian Y, Modi V. Global trends in urban electricity demands for cooling and heating. Energy 2017;127. https://doi.org/10.1016/j.energy.2017.03.095.
[11] Hahn H, Meyer-Nieberg S, Pickl S. Electric load forecasting methods: tools for decision making. Eur J Oper Res 2009;199(3). https://doi.org/10.1016/j.ejor.2009.01.062.
[12] McLoughlin F, Duffy A, Conlon M. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. Energ Build 2012;48. https://doi.org/10.1016/j.enbuild.2012.01.037.
[13] Akbari S, Haghighat F. Occupancy and occupant activity drivers of energy consumption in residential buildings. Energ Build 2021;250. https://doi.org/10.1016/j.enbuild.2021.111303.
[14] Wang Z, et al. Individual difference in thermal comfort: a literature review. Build Environ 2018;138. https://doi.org/10.1016/j.buildenv.2018.04.040.
[15] Bartusch C, Odlare M, Wallin F, Wester L. Exploring variance in residential electricity consumption: household features and building properties. Appl Energy Apr. 2012;92:637–43. https://doi.org/10.1016/j.apenergy.2011.04.034.
[16] California Energy Commission (CEC). 2019 California Residential Appliance Saturation Survey (RASS) [Online]. Available: https://www.energy.ca.gov/data-reports/surveys/2019-residential-appliance-saturation-study; 2019.
[17] United States Census Bureau. 2021 American Housing Survey (AHS) [Online]. Available: https://www.census.gov/programs-surveys/ahs.html; 2019.
[18] Chen M, Ban-Weiss GA, Sanders KT. The role of household level electricity data in improving estimates of the impacts of climate on building electricity use. Energ Build 2018;180:146–58. https://doi.org/10.1016/j.enbuild.2018.09.012.
[19] Chen M, Sanders KT, Ban-Weiss GA. A new method utilizing smart meter data for identifying the existence of air conditioning in residential homes. Environ Res Lett 2019;14(9). https://doi.org/10.1088/1748-9326/ab35a8.
[20] Dyson MEH, Borgeson SD, Tabone MD, Callaway DS. Using smart meter data to estimate demand response potential, with application to solar energy integration. Energy Policy 2014;73. https://doi.org/10.1016/j.enpol.2014.05.053.
[21] Peplinski M, Kalmus P, Sanders KT. Investigating whether the inclusion of humid heat metrics improves estimates of AC penetration rates: a case study of Southern California. Environ Res Lett 2023;18(10):104054. https://doi.org/10.1088/1748-9326/acfb96.
[22] Xu X, Wang W, Hong T, Chen J. Incorporating machine learning with building network analysis to predict multi-building energy use. Energ Build 2019;186:80–97. https://doi.org/10.1016/j.enbuild.2019.01.002.
[23] Westermann P, Deb C, Schlueter A, Evins R. Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data. Appl Energy 2020;264. https://doi.org/10.1016/j.apenergy.2020.114715.
[24] Porteiro R, Nesmachnow S, Moreno-Bernal P, Torres-Aguilar CE. Computational intelligence for residential electricity consumption assessment: detecting air conditioner use in households. Sustain Energy Technol Assess 2023;58. https://doi.org/10.1016/j.seta.2023.103319.
[25] Perez KX, Cetin K, Baldea M, Edgar TF. Development and analysis of residential change-point models from smart meter data. Energ Build Mar. 2017;139:351–9. https://doi.org/10.1016/j.enbuild.2016.12.084.
[26] U.S. households' heating equipment choices are diverse and vary by climate region. U.S. Energy Information Administration; 2017. https://www.eia.gov/todayinenergy/detail.php?id=30672.
[27] Deason J, Borgeson M. Electrification of buildings: potential, challenges, and outlook. Curr Sustain Renew Energy Rep 2019;6(4). https://doi.org/10.1007/s40518-019-00143-2.
[28] Elmallah S, Montañés CC, Callaway D. Who heats and cools? Access to residential heating and cooling in Northern California and implications for energy transitions. Energy Policy 2024;191:114169.
[29] Lin Z, Deng S. A questionnaire survey on sleeping thermal environment and bedroom air conditioning in high-rise residences in Hong Kong. Energ Build 2006;38(11). https://doi.org/10.1016/j.enbuild.2006.04.004.
[30] Chua KJ, Chou SK. Energy performance of residential buildings in Singapore. Energy 2010;35(2). https://doi.org/10.1016/j.energy.2009.10.039.
[31] Meng X, Gao Y, Hou C, Yuan F. Questionnaire survey on the summer air-conditioning use behaviour of occupants in residences and office buildings of China. Indoor Built Environ 2019;28(5). https://doi.org/10.1177/1420326X18793699.
[32] An J, Yan D, Hong T. Clustering and statistical analyses of air-conditioning intensity and use patterns in residential buildings. Energ Build 2018;174. https://doi.org/10.1016/j.enbuild.2018.06.035.
[33] Xia D, Lou S, Huang Y, Zhao Y, Li DHW, Zhou X. A study on occupant behaviour related to air-conditioning usage in residential buildings. Energ Build 2019;203. https://doi.org/10.1016/j.enbuild.2019.109446.
[34] Wang Y, Wu J, Xie F, Liu C, Li H, Cheng J. Survey of residential air-conditioning-unit usage behavior under South China climatic conditions. 2011. https://doi.org/10.1109/ICEICE.2011.5776982.
[35] Zhu M, Huang Y, Wang SN, Zheng X, Wei C. Characteristics and patterns of residential energy consumption for space cooling in China: evidence from appliance-level data. Energy 2023;265. https://doi.org/10.1016/j.energy.2022.126395.
[36] Schweiker M, Shukuya M. Comparison of theoretical and statistical models of air-conditioning-unit usage behaviour in a residential setting under Japanese climatic conditions. Build Environ 2009;44(10). https://doi.org/10.1016/j.buildenv.2009.03.004.
[37] Ramapragada P, Tejaswini D, Garg V, Mathur J, Gupta R. Investigation on air conditioning load patterns and electricity consumption of typical residential buildings in tropical wet and dry climate in India. Energy Inform 2022;5. https://doi.org/10.1186/s42162-022-00228-1.

[38] Malik A, Haghdadi N, MacGill I, Ravishankar J. Appliance level data analysis of summer demand reduction potential from residential air conditioner control. Appl Energy 2019;235. https://doi.org/10.1016/j.apenergy.2018.11.010.

[39] Zaki SA, Hagishima A, Fukami R, Fadhilah N. Development of a model for generating air-conditioner operation schedules in Malaysia. Build Environ 2017; 122. https://doi.org/10.1016/j.buildenv.2017.06.023.

[40] Dash S, Sahoo NC. Electric energy disaggregation via non-intrusive load monitoring: a state-of-the-art systematic review. Electr Power Syst Res 2022;213. https://doi.org/10.1016/j.epsr.2022.108673.

[41] Qi N, et al. Smart meter data-driven evaluation of operational demand response potential of residential air conditioning loads. Appl Energy Dec. 2020;279:115708. https://doi.org/10.1016/J.APENERGY.2020.115708.

[42] Palmgren C, Stevens N, Goldberg M, Bames R, Rothkin K. 2009 California residential appliance saturation survey. 2009.

[43] California Energy Commision 2015 California Building Climate Zone Areas [Online]. Available: http://energy.ca.gov/maps/renewable/building_climate_zones.html; 2015.

[44] Household Energy Use in California [Online]. Available: https://www.eia.gov/consumption/residential/reports/2009/state_briefs/pdf/ca.pdf; 2009.

[45] California Irrigation Management Information Sytem (CIMIS). CIMIS station reports [Online]. Available: https://cimis.water.ca.gov/Stations.aspx; 2024.

[46] National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI). NOAA NCEI Local Climatological Data (LCD) [Online]. Available: https://www.ncei.noaa.gov/cdo-web/datatools/lcd; 2024.

[47] Environmental Protection Agency (EPA). EPA Air Quality System (AQS) [Online]. Available: https://aqs.epa.gov/aqsweb/airdata/download_files.html; 2024.

[48] Stoica P, Selen Y. Model-order selection: a review of information criterion rules. IEEE Signal Process Mag 2004;21(4).

[49] US Census Bureau 2016 cartographic boundary shapefiles-census tracts [Online]. Available: https://census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html; 2024.

[50] Petri Y, Caldeira K. Impacts of global warming on residential heating and cooling degree-days in the United States. Sci Rep 2015;5. https://doi.org/10.1038/srep12427.