Energy Strategy Reviews 49 (2023) 101175

Contents lists available at ScienceDirect

Energy Strategy Reviews

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

A data-driven approach for the disaggregation of building-sector heating and cooling loads from hourly utility load data

Yinbo Hu [a,*], Michael Waite [a,c], Evan Patz [a], Bainan Xia [b], Yixing Xu [b], Daniel Olsen [b], Naveen Gopan [a], Vijay Modi [a]

[a] Department of Mechanical Engineering, Columbia University, 220 S.W. Mudd Building, 500 West 120th Street, New York, NY, 10027, USA
[b] Breakthrough Energy, USA
[c] The American Council for an Energy-Efficient Economy, Washington, DC, USA

ARTICLE INFO

Handling editor: Xi Lu

Keywords:
Decarbonization
Electrification
Load disaggregation
Building thermal load

ABSTRACT

Electrification of space heating in buildings, currently dominated by on-site fossil fuel use, will be an essential element of decarbonization. Some electrification of heating is already underway, and although state-by-state adoption is highly heterogeneous, the associated impact on the grid is already being felt. The same has been true for air-conditioning except adoption levels are generally higher. As we expect rapid adoption of electrification, a model that can seamlessly disaggregate the existing electric load into that for heating, cooling and non-thermal uses is essential. We develop a model that captures the building thermal response to quantities such as building floor areas that change over years and weather that changes through the day and through the year. Complexity of occupancy, thermostat settings, diversity in building envelopes or technologies deployed are not explicitly represented, but their effects are captured as hourly changes in the response. The model can then be used to estimate the non-thermal dependent loads. Once such a disaggregation is available, it can be used to estimate new load profiles as changes in floor area or electrified loads or weather occurrences.

The model is validated against actual hourly utility loads by re-aggregating the simulated hourly loads for a single year for all load zones in NYISO (boundary aligns with New York), ERCOT (covering most of Texas), CAISO (covering most of California) and some other individual balancing authorities of California (BANC, TIDC, IID, LADWP and WALC) with mean absolute percentage errors (MAPEs) across all between 3.0% and 6.0%. The obtained model parameters are further tested by backcasting hourly load for the past 10-year period without degradation in errors, which suggest the model is promising for forecasting in the long-term. While our results bridge the gap between building level energy simulation and building stock energy prediction, all source data for the present study are extracted from open-access datasets. The model is available as an open-source tool that can be easily applied to any spatial resolution at any geographical locations, as long as load profiles and building census data are available.

1. Introduction

The US building sector is responsible for 40% of the country's energy consumption [1], and 30% of GHG emissions [2]. Within this sector, nearly half of site energy is for space heating, cooling and ventilation [3,4]. In order to comply with pressing decarbonization goals, space heating must be electrified at a much faster rate and inefficient space cooling devices need to be upgraded as well [5,6]. This will significantly change the current shape of load and thus create challenges on resources allocation for power grids [7,8]. Thus, governments, associated stakeholders and power utilities need to better understand not only the current amount of energy consumed by each end-use in building sector, but also what future effects electrification and energy efficiency measures might have [9]. Understanding the relationship between long-term electricity load growth (and the evolution of the load profiles) and weather, electrification and building floor areas is therefore essential for utilities and independent system operators (ISO). Such relationships allow utilities to develop policies for future grid planning and allow for the co-design of incentives that encourage customers to reduce their site fossil-fuel use and utility-side incentives for the distribution grid capabilities [10,11].

* Corresponding author.
E-mail address: yh3184@columbia.edu (Y. Hu).

https://doi.org/10.1016/j.esr.2023.101175
Received 4 November 2022; Received in revised form 29 December 2022; Accepted 27 March 2023
Available online 22 August 2023

Nomenclature

A_z total building floor area of BA z
$DF_{z,t}$ darkness fraction of BA z at hour t
$E_{z,t}$ hourly grid load for BA z at hour t
$\widetilde{E}^{base}_{z,t}$ model estimated temperature-neutral 'base' electricity load of BA z at hour t
$\widetilde{E}^{clg}_{z,t}$ model estimated cooling electricity load of BA z at hour t
$\widetilde{E}^{htg}_{z,t}$ model estimated heating electricity load of BA z at hour t
$I^{htg}_{z,dh(t)}$ intercept of electricity heating regression model
$I^{clg}_{z,dh(t)}$ intercept of electricity cooling regression model
p^{htg}_{z} fraction of floor area with electric heating in BA z
p^{clg}_{z} fraction of floor area with electric cooling in BA z
$S^{htg}_{z,dh(t)}$ temperature-dependent heating electricity response coefficient for BA z at hour of day t
$S^{df}_{z,dh(t)}$ darkness fraction-dependent heating electricity response coefficient for BA z at hour of day t
$S^{clg,db}_{z,dh(t)}$ temperature-dependent cooling electricity response coefficient for BA z at hour of day t
$S^{clg,wb.diff}_{z,dh(t)}$ modified wet bulb temperature-dependent cooling electricity response coefficient for BA z at hour of day t
$s^{htg}_{z,dh(t)}$ specific temperature-dependent heating electricity response coefficient for unit building floor area using electric heating in BA z at hour of day t
$s^{clg}_{z,dh(t)}$ specific temperature-dependent cooling electricity response coefficient for unit building floor area using electric cooling in BA z at hour of day t
$T_{z,t}$ temperature of BA z at hour t
$T^{wb}_{z,t}$ wet bulb temperature of BA z at hour t
$T^{wb.diff}_{z,t}$ modified wet bulb temperature of BA z at hour t
$T_{bph,z,dh(t)}$ breakpoint temperature of electricity heating
$T_{bpc,z,dh(t)}$ breakpoint temperature of electricity cooling

Abbreviation

NYIS-ZONA New York Independent System Operator (NYIS), West
NYIS-ZONB New York Independent System Operator (NYIS), Genesee
NYIS-ZONC New York Independent System Operator (NYIS), Central
NYIS-ZOND New York Independent System Operator (NYIS), North
NYIS-ZONE New York Independent System Operator (NYIS), Mohawk Valley
NYIS-ZONF New York Independent System Operator (NYIS), Capital
NYIS-ZONG New York Independent System Operator (NYIS), Hudson Valley
NYIS-ZONH New York Independent System Operator (NYIS), Millwood
NYIS-ZONI New York Independent System Operator (NYIS), Dunwoodie
NYIS-ZONJ New York Independent System Operator (NYIS), New York City
NYIS-ZONK New York Independent System Operator (NYIS), Long Island
ERCO-C: Electric Reliability Council of Texas, Inc. (ERCO), Coast
ERCO-E: Electric Reliability Council of Texas, Inc. (ERCO), East
ERCO-FW Electric Reliability Council of Texas, Inc. (ERCO), Far West
ERCO-N: Electric Reliability Council of Texas, Inc. (ERCO), North
ERCO-NC Electric Reliability Council of Texas, Inc. (ERCO), North Central
ERCO-S: Electric Reliability Council of Texas, Inc. (ERCO), South
ERCO-SC Electric Reliability Council of Texas, Inc. (ERCO), South Central
ERCO-W: Electric Reliability Council of Texas, Inc. (ERCO), West
CISO-PGAE California Independent System Operator (CISO), Pacific Gas and Electric
CISO-SCE California Independent System Operator (CISO), Southern California Edison
CISO-SDGE California Independent System Operator (CISO), San Diego Gas and Electric
CISO-VEA California Independent System Operator (CISO), Valley Electric Association
IID Imperial Irrigation District
WALC Western Area Power Administration - Desert Southwest Region
LADWP Los Angeles Department of Water and Power
TIDC Turlock Irrigation District
BANC Balancing Authority of Northern California

Two distinct approaches are typically used to estimate grid loads: top-down or bottom-up. Bottom-up models typically take advantage of the accuracy of physics-based energy dynamic models of individual buildings, and then aggregate those simulated energy consumption results to higher spatial regimes [12–15]. These bottom-up approaches estimate building energy consumption based on building construction layouts, envelope specifications and local weather conditions [16–18]. The predictions from these studies must rely on establishing a representative set of building diversity and calibrating models on a case-by-case basis to the building stock in a region. Substantial efforts are therefore required to guarantee the repeatability when applying the models to different geographical scales and locations. One of the most representative works is the demand side grid model that NREL developed, in which around 900,000 typical buildings were simulated then weighted to represent the entire building stock of the US [19]. Nevertheless, the hourly accuracy of the aggregated load at state level is reported to be 20%.

A top-down approach, on the other hand, predicts aggregated building stock energy consumption over a well-defined geographical area, attempting to capture variations in weather, energy prices and building stock throughout the given area. The areas we consider are balancing authorities (BA) and load zones. As the most studied meteorological factor, ambient temperature dependency of electricity load is often characterized by a 'U' or 'V' shape. Hourly electricity demand generally increases as temperatures rise beyond some threshold, or drop below a threshold, since some heating is electric despite most of the contiguous US still primarily heating with fossil fuels. Thus, many have proposed a V-shaped segmented regression model to explain the relationship between energy demand and temperature [20,21]. Subsequently, some studies have proposed to have a comfort temperature zone between two thresholds, where electricity demand has no temperature-dependence [22–25]. This temperature-independence assumption is unlikely to capture effects aggregated across buildings with different heating and cooling behavior, but it correctly identifies a transition between heating- and cooling-dominated loads.

Regression models also have the advantage of isolating effects of different variables with coefficients that reflect load response to each variable and allow one to study perturbation of those variables due to anticipated system changes. While temperature is a primary factor that drives electricity consumption [26], building stock energy demand responds to countless other factors. For example, humidity, solar radiation, wind speed, building envelope thermal efficiency, diurnal

occupancy and internal load patterns are commonly studied [27–30]. However, multiple linear regression (MLR) models that takes all of these predictors typically have multicollinearity problems raised by correlated predictors [26,29]. Hence some studies resort to sensitivity analysis [31]. The results of those studies show that the influence of these variables on electricity load varies with locations. For example, humidity has greater impact on space cooling energy consumption in hot and humid climates [32]. While solar radiation has a larger impact on building loads for places with a dry and high solar radiation climate [33]. Socioeconomic indicators can also impact certain countries and certain historic periods [34]. Our study utilizes two new variables that are built upon some of the well-studied factors, and are less location-wise sensitive for regression models.

One of the drawbacks of top-down studies is that they have limited ability to forecast future loads if the prediction horizon involves significant changes from status quo (e.g. high growth in building floor area and electrification) [35]. Predicting load growth under such energy supply-demand scenarios becomes questionable except for the business-as-usual (BAU) case [9,36]. Without validated disaggregation that is locally specific to building stock characteristics, the ability to estimate load profile changes over a decadal time horizon is not evaluated in existing literature. This is even more important when building stock varies across climatically diverse geographies. With the increased availability of high-quality meteorological observation datasets, high-resolution grid monitoring, and household stock characteristics at census tract level that are published for the entire US, a building energy disaggregation approach is inevitably attractive for understanding load sub-components and subsequently end-use demand forecasting. Such an approach should be applicable to a range of utility or balancing area footprints with acceptable error margins.

Other data-driven methods have been adopted as well and shown promising results. For example, classic time-series models [37]; machine learning based models like support vector machine (SVM) [38], neural networks [28,39], genetic algorithms [40]; and semi- or non-parametric methods like smoothing splines [41] and fuzzy [42] models. However, challenges exist for some of such models in either interpretability, over-fitting or repeatability for various locations.

Most recently, top-down and bottom-up methods have begun to merge into hybrid methods that are meant to analyze the energy performance of neighborhoods or cities [9]. However, research gaps between the two approaches arise from differences in spatial scale or temporal resolution or attempts at providing physical interpretation to results. Consequently, these results inherently differ in spatial, temporal scale and the model framework, making them impossible to compare and cross validate.

In an attempt to bridge these gaps, the present study introduces a segmented multiple regression model to study balancing authority (BA) and sub-BA level electric loads' dependencies on ambient temperature, time-of-day, weekday/weekend, wet-bulb temperature and the presence of daylight with hourly resolution. The methodology is then applied to 22 load zones of three representative BAs of the US – NYISO, ERCOT, CAISO, as well as 5 other BAs that operate parts of California's power grid. The overall geographical area covered includes all of New York (NY), and the majorities of Texas (TX) and California (CA). The re-aggregation of the disaggregated results is then compared with the actual load every hour of the year, which exhibits high level of agreement.

2. Methods

2.1. General description

The disaggregation of electricity usage for heating and cooling from other electrical loads is performed for a BA or a load zone where hourly electrical load data are available. There are three different temperature ranges considered. In the temperature range below T_{bpc} (as shown in Fig. 1), we identify electric loads that increase with decreasing temperatures as heating loads, and loads that increase with temperatures in the range above T_{bph}, as cooling loads. The temperature range in between T_{bpc} and T_{bph} is referred to as the transition zone, where some floor area is presumably heated, some cooled and some neither heated nor cooled.

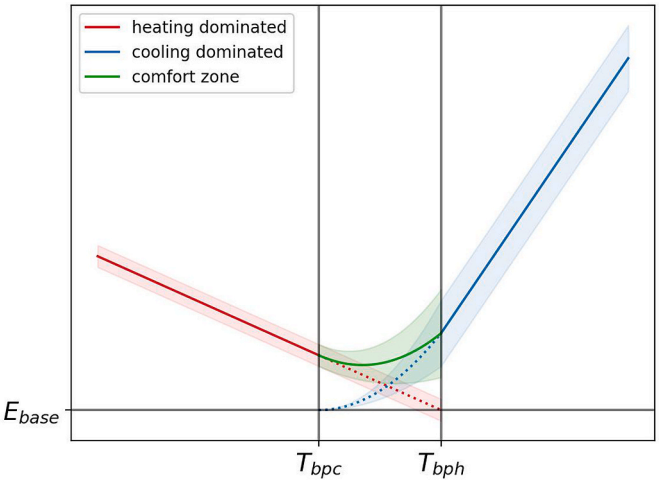

Fig. 1. Scheme plot of segmented regression model on hourly BA load and temperature.

In the heating and cooling dominated ranges, we assume a linear dependence of load on temperature and fit a least-square estimator to determine associated temperature-dependent heating and cooling response intensity coefficients. The temperature-dependent load usage at the intermediate 'transition zone' is deterministically estimated as the sum of the linear extrapolated heating load and the extrapolated cooling load by a second-order polynomial function. In addition to hourly temperatures, we also consider effects of humidity on cooling by using the deviation of wet bulb temperature (WBT) from the mean DBT-WBT relationship and the time fractional presence of daylight for each hour. The intervals in Fig. 1 is intended to show the range of load prediction when considering these two variables. Correlations between loads and each predictor variables for the case study regions are shown in Tables S2–S5.

The temperature-dependent electricity response coefficients are obtained at each hour of the day (day-hour), with separate coefficients for weekdays (strictly workdays) and weekends (strictly includes holidays), resulting in 48 coefficients for each load zone, and in units of energy each hour per unit temperature difference from the corresponding threshold temperature. These coefficients are then normalized by building floor area that is equipped with electrical appliances specific to either heating or cooling. This permits comparison of the results across individual regions in terms of electricity usage and forecasting load growth over long periods of time. We obtain building stock floor area information from four open access datasets – Hazus General Building Stock (GBS) [43], American Community Survey (ACS) 5-year estimates [44], Residential Energy Consumption Survey (RECS) [3] and Commercial Buildings Energy Consumption Survey (CBECS) [4].

We apply the method described by Waite and Modi [7] and synthesize the above datasets to estimate the total residential and commercial building floor area in the region and the fractions of floor area that are heated or cooled by electricity. A flow chart of data collection and usages in modelling process is shown in Fig. 2, while additional details are presented in supplementary file.

2.2. Mathematical formulation

For each one of the 24 h on a weekday or weekend day throughout

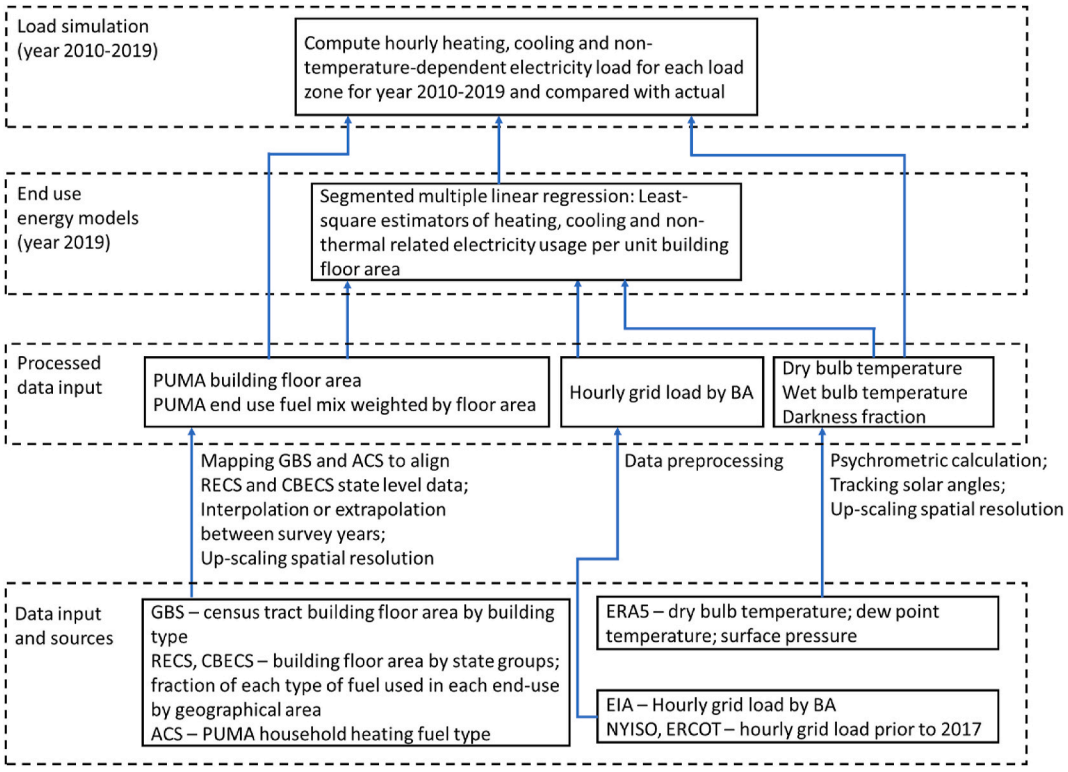

Fig. 2. Flow chart of modelling data processing and usages in modeling process.

the year, a segmented linear regression model is fitted to estimate heating electricity response for all hours with temperature lower than the threshold temperature T_{bpc} (Eq. (1)) and cooling electricity response for all hours with temperature higher than the breakpoint temperature T_{bph} (Eq. (2)). To determine 'base' load, the load that would be presented in the absence of any heating or cooling loads, the obtained heating response is linearly extrapolated to the heating breakpoint temperature T_{bph}. This process was found to assure a consistent temperature-neutral 'base' load for both the heating and cooling regression, while not altering the buildings' thermal response to weather conditions that are extracted from the data. We initially set T_{bpc} to 10 °C and T_{bph} to 18.3 °C and allow these two threshold temperatures to self-adjust up and down for locations with extremely warm or cold climates to ensure an adequate number of data points. Existing literature has shown that warmer climates tend to have higher breakpoint temperature for the best model fitting [45].

$$E_{z,t} = I^{htg}_{z,dh(t)} - S^{htg}_{z,dh(t)} T_{z,t} + S^{df}_{z,dh(t)} DF_{z,t}, \text{if } T_{z,t} < T_{bpc,z,dh(t)} \quad (1)$$

$$E_{z,t} - I^{htg}_{z,dh(t)} + S^{htg}_{z,dh(t)} T_{bph,z,dh(t)} - S^{df}_{z,dh(t)} DF_{z,t} = I^{clg}_{z,dh(t)} - S^{clg,db}_{z,dh(t)} T_{z,t} + S^{clg,wb.diff}_{z,dh(t)} T^{wb.diff}_{z,t}, \text{if } T_{z,t} > T_{bph,z,dh(t)} \quad (2)$$

Where, $E_{z,t}$ is the electricity load of load zone (balancing area), z, at time step, t. $I^{htg}_{z,dh(t)}$ and $I^{clg}_{z,dh(t)}$ are intercepts of heating and cooling regression models. $S^{htg}_{z,dh(t)}$ and $S^{clg,db}_{z,dh(t)}$ are the temperature-dependent heating electricity response coefficient and cooling electricity response coefficient, respectively. Both are in units of MW/^oC. $dh(t)$ denotes the day-hour on either a weekday or weekend day (for 48 total day-hours). $T_{z,t}$ is the dry bulb temperature. The model takes two more variables other than ambient temperature to improve accuracy – darkness fraction ($DF_{z,t}$) and modified wet bulb temperature ($T^{wb.diff}_{z,t}$).

Prior studies have established that the electricity consumption at clock times in the morning and late afternoon is affected by whether that those hours have light and hence solar gain (e.g. in the summer) or not (e.g. in the winter) [46,47]. A recent study found the presence of sunlight can influence grid load by up to 5% for hours around sunrise in the morning and 15% for hours around sunset in the evening [48]. To identify lighting load in those hours from heating load, the darkness fraction (DF) time series is simulated for each load zone to quantify the temporal ratio of daylight presences for each hour of a year. $S^{df}_{z,dh(t)}$ is the assumed linear correlation between darkness fraction and grid load.

Wet-bulb temperature (WBT) affects latent cooling load, particularly during humid summer conditions [49]. In this study, a baseline of WBT is established first by fitting hourly WBT with dry-bulb temperature with a second order polynomial function (Eq. (4)). This provides a dry bulb temperature-dependent performance along the mean DBT-WBT relationship. The model then considers the effect of the deviation of the WBT from this mean-line behavior (DWBT), $T^{wb.diff}_{z,t}$ (Eq. (3)), which, like other variables in the regression, is assumed to have a linear relationship with grid load. $S^{clg,wb.diff}_{z,dh(t)}$ is the corresponding model coefficient and has the unit of MW/^oC. In this way, WBT-dependent energy usage is established as a deviation from the mean DBT-WBT relationship, which allows for robustness against outlier WBTs, easy model interpretation and the extraction of a single dominant temperature-dependent cooling behavior.

$$T^{wb.diff}_{z,t} = T^{wb}_{z,t} - wb_{func}(T_{z,t}) \quad (3)$$

Where $wb_{func}(T_{z,t})$ is assumed to be the least square estimator of a polynomial function of DBT as shown in Eq. (4) below. The parameters a_z, b_z, c_z in Eq. (4) are obtained from such a fit to hourly dry bulb temperature and wet bulb temperature.

$$wb_{func}(T_{z,t}) = a_z(T_{z,t})^2 + b_z T_{z,t} + c_z, T_{z,t} > T_{bph} \quad (4)$$

The heating coefficients $S^{htg}_{z,dh(t)}$ and cooling coefficients $S^{clg,db}_{z,dh(t)}$ that are estimated from the above regression model are then normalized by total floor area that is heated or cooled by electricity within the load zone

boundary to obtain the response of building electric loads for thermal end uses to temperature per unit floor area $s^{htg}_{z,dh(t)}$ and $s^{clg,db}_{z,dh(t)}$ (Eq. (5)). These specific building thermal response intensities are in units of $Btu_e/(h \bullet m^2 \bullet^\circ C)$, which allows comparison between heating and cooling, and across geographical regions and energy sources.

$$s^{htg}_{z,dh(t)} = \frac{S^{htg}_{z,dh(t)}}{A_z p^{htg}_z};\ s^{clg}_{z,dh(t)} = \frac{S^{clg}_{z,dh(t)}}{A_z p^{clg}_z} \tag{5}$$

Where A_z is the total building floor area of load zone z, $p^{htg}_{elec,z}$ and $p^{clg}_{elec,z}$ are the fractions of floor area with electric heating and cooling.

Estimated hourly heating, cooling and 'base' load are then calculated from obtained coefficients as shown by Eqs. (6)–(8). While we do not explicitly fit for hours with temperature between T_{bpc} and T_{bph}, we calculate the temperature-dependent electricity usage in this range later by applying two deterministic diminishing behaviors to heating and cooling load. We assume heating load is decreasing linearly, whereas cooling load is decreasing quadratically as temperatures move away from respective breakpoint temperature in this range (Eq. (7)). Note that the "+" outside a bracket returns the value inside the bracket or zero, whichever is larger.

$$\widetilde{E}^{htg}_{z,t} = S^{htg}_{z,dh(t)} \left(T_{bph,z,dh(t)} - T_{z,t}\right)^+ \tag{6}$$

$$\widetilde{E}^{clg}_{z,t} = \begin{cases} I^{clg}_{z,dh(t)} + S^{clg,db}_{z,dh(t)} T_{z,t} + S^{clg,wb.diff}_{z,dh(t)} T^{wb.diff}_{z,t}, \text{if } T_{z,t} \geq T_{bph,z,dh(t)} \\ \left(I^{clg}_{z,dh(t)} + S^{clg,db}_{z,dh(t)} T_{bph,z,dh(t)} + S^{clg,wb.diff}_{z,dh(t)} T^{wb.diff}_{z,t}\right)\theta^2, \text{if } T_{z,t} < T_{bph,z,dh(t)} \end{cases} \tag{7}$$

where $\theta = \frac{T_{z,t} - T_{bpc,z,dh(t)}}{T_{bph,z,dh(t)} - T_{bpc,z,dh(t)}}$.

$$\widetilde{E}^{base}_{z,t} = I^{htg}_{z,dh(t)} - S^{htg}_{z,dh(t)} T_{bph,z,dh(t)} + S^{df}_{z,dh(t)} DF_{z,t} \tag{8}$$

3. Study areas

We validate our approach for all load zones in three representative BAs – NYISO, ERCOT and CAISO (excluding CAISO-VEA which is entirely outside California), and five other BAs that operate some parts of the California grid – BANC, LADWP, IID, TIDC and WALC. The selected area covers all of New York (NY), and most of Texas (TX) and California (CA), while avoiding the complication raised by load zones that cross state boundaries. The modeled regions are shown in Fig. 3. For simplicity, we use "CAISO" in the following sections of this paper to refer to the studied area in CA, including the other balancing authorities, as shown in Fig. 3(c). With these studied load zone areas, our results cover 4 out of 7 major climate zones in the contiguous US. The characteristics of the population weighted average temperature for load zones are shown in Table 1.

For the consistency of meteorologically driven time series and energy consumption, and to ensure that the joint distribution over these time series is adequately characterized, building floor areas and electricity penetrations by end-use are interpolated to the year 2019, while locational weather data and energy consumption profiles are retrieved for 2019 data directly from ERA5 dataset [50] and EIA grid monitor [51] respectively.

4. Model accuracy

The modelling approach allows us to estimate the hourly electric heating, cooling and remaining load (here called the 'base' load) for all hours of the year for any load zones. This allows us to obtain a model-generated estimate of the hourly load of a load zone, and we compare that with the actual zonal demand. This is first carried out for each of the load zones for year 2019. Table 2 shows the mean absolute percentage error (MAPE) over 8760 annual hours and relative error in annual peak load for all zones under study. The model captures load variabilities within 6% of error for all load zones in NYISO and ERCOT. In CAISO, 4%–7% error is typically measured. We notice a major source of uncertainty comes from particular zones that have low building floor area (see Table S1 for floor areas and fraction for each BA). This issue was also discussed in a recent NREL report [52].

Fig. 4 shows the fitting accuracy of daily peak time (with logit x-axis scale). For most of the load zones, the time errors between simulated and actual daily peak time of grid load are 0–2 h for most of the days during a year. Accurately capturing the daily peak load is beneficial to infrastructure capacity and operational planning.

State aggregation of predicted hourly loads compared to actual are shown in Fig. 5. For most of the hours, prediction errors are within 20%, while the MAPEs of all three states fall within 4%. The model is especially accurate for high demand periods, which is a desired feature, since predicting peak load is of great importance for power grid planning and operations [53]. Accurately predicting the peak load – and hourly load more generally – is also one of the necessary conditions to guarantee convincing result for any further studies on reducing buildings greenhouse gas emission that will be built upon this model.

Hourly energy prediction breakdowns for peak demand days of both heating and cooling seasons are shown in Fig. 6 for each state. Results show that the hourly error rates of model prediction of peak load are generally within 6% in total load for both seasons. The model is also

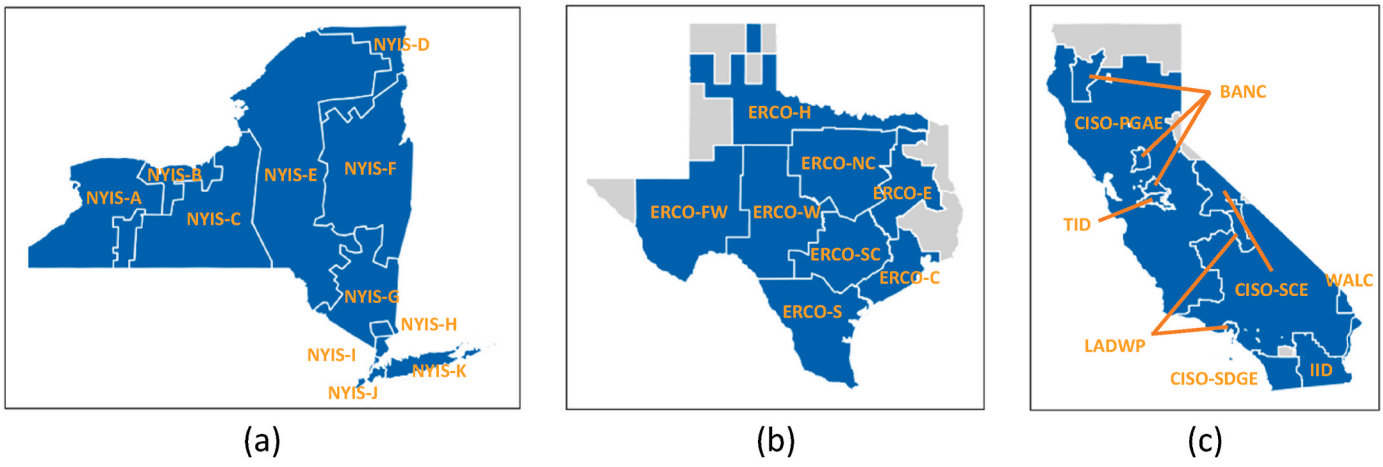

Fig. 3. BA and load zone maps of the three representative regions for this study. (a) NYISO, (b) ERCOT, and (c) CAISO (excluding CISO-VEA), BANC, LADWP, IID, TIDC, WALC operated area in California.

Table 1
Characteristics of load zone electricity loads and climate conditions.

Load zone	Annual average	Coldest month			Hottest month		
	Average load (MW)	Average DBT (°C)	99% coldest DBT (°C)	Peak load (MW)	Average DBT (°C)	99% hottest DBT (°C)	Peak load (MW)
NYIS-ZONA	1697	−4.9	−18.8	2192	22.4	28.7	2620
NYIS-ZONB	1109	−4.6	−17.5	1603	22.5	29.5	1926
NYIS-ZONC	1804	−5.6	−19.5	2711	22.5	29.8	2705
NYIS-ZOND	550	−9.7	−24.1	774	21.8	29.9	608
NYIS-ZONE	898	−7.4	−21.8	1418	22.0	29.5	1395
NYIS-ZONF	1350	−6.0	−21.3	2065	23.2	31.4	2301
NYIS-ZONG	1092	−2.5	−18.5	1617	24.3	33.0	2243
NYIS-ZONH	321	−1.4	−16.8	533	25.1	33.6	659
NYIS-ZONI	682	−0.6	−15.3	941	25.8	33.9	1392
NYIS-ZONJ	5936	0.2	−14.3	7755	26.2	34.2	10,802
NYIS-ZONK	2345	0.9	−12.8	3390	25.2	31.4	5437
ERCO-C	12,354	11.8	0.8	14,157	29.6	36.6	21,256
ERCO-E	1485	9.0	−1.9	2131	30.2	37.7	2554
ERCO-FW	3435	7.9	−3.6	3555	30.9	41.0	4308
ERCO-N	854	6.0	−4.0	1168	29.8	38.0	1476
ERCO-NC	13,819	7.6	−1.9	19,501	30.4	37.6	25,494
ERCO-S	3608	15.1	4.9	4726	31.4	38.0	6040
ERCO-SC	6970	10.6	0.5	9198	30.9	38.3	12,785
ERCO-W	1292	8.8	−1.0	1701	31.3	39.3	2117
CISO-PGAE	11,316	8.8	1.7	13,514	22.4	32.9	21,039
CISO-SCE	11,371	9.6	2.4	13,265	24.1	32.0	20,934
CISO-SDGE	2197	11.4	5.1	3035	21.9	28.3	3637
IID	419	12.0	4.1	415	33.8	43.7	1067
WALC	1087	9.9	1.3	1331	32.5	41.9	1919
LADWP	3064	10.1	2.7	3851	23.8	32.3	5609
TIDC	308	9.6	0.7	320	28.0	40.3	643
BANC	1949	9.3	0.6	2288	26.1	39.6	4428

Table 2
Model prediction accuracy of all BAs that are under study. MAPE: mean absolute percentage error of predicted load of 8760 annual hours.

BA	MAPE	Relative error for peak load	BA	MAPE	Relative error for peak load
NYIS-ZONA	3.10%	3.05%	ERCO-N	4.28%	2.37%
NYIS-ZONB	3.50%	5.62%	ERCO-NC	5.00%	0.59%
NYIS-ZONC	3.48%	2.02%	ERCO-S	5.11%	4.89%
NYIS-ZOND	3.57%	1.83%	ERCO-SC	4.84%	0.73%
NYIS-ZONE	5.57%	4.97%	ERCO-W	3.77%	0.67%
NYIS-ZONF	4.97%	0.57%	CISO-PGAE	6.83%	1.43%
NYIS-ZONG	4.19%	1.13%	CISO-SCE	3.78%	4.29%
NYIS-ZONH	6.02%	0.18%	CISO-SDGE	5.99%	6.77%
NYIS-ZONI	3.82%	2.84%	IID	5.30%	8.08%
NYIS-ZONJ	3.10%	2.86%	WALC	10.9%	2.87%
NYIS-ZONK	4.47%	2.15%	LADWP	4.00%	3.44%
ERCO-C	3.74%	1.73%	TIDC	4.67%	0.71%
ERCO-E	4.52%	1.13%	BANC	5.28%	5.21%
ERCO-FW	4.33%	2.93%			

capable of capturing the bi-peak behavior of load in winter months for all of the three states. Seasonal average load prediction can be found in Fig. S1.

The hourly predictions of our model are then compared to a short list of existing models that we reviewed. As shown in Table 3, our model is able to predict in high accuracy, and at the same time, without loss of generality and repeatability. All sources of data in this study are open access, which allows the model to be fully open-source.

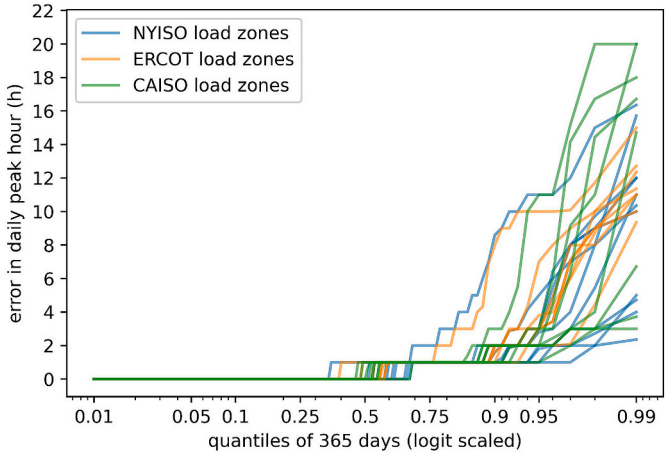

Fig. 4. Error distribution for all BAs under study in predicting daily peak load time. X-axis is in logit scale. We assume a day has bi-peak load profile if the second highest load hour is not adjacent to peak hour. And If a day has two comparable peaks, the minimum period between actual peaks and the simulated peak is recorded.

5. Building thermal response to temperature and long-term load forecasting

5.1. Electric load response for temperature-dependent end use

The disaggregated intensities of electric heating and cooling response to temperature are shown at a BA-level in Fig. 7. As discussed in Section 2 (Eqs. (1), (2) and (5)), all coefficients for temperature-dependent usages are normalized with building floor area that is heated or cooled by electricity, providing us results in temperature and area normalized units of $\mathrm{Btu_e}/(\mathrm{h} \bullet \mathrm{m}^2 \bullet^{\circ} C)$. Heating and cooling coefficients without taking DF and DWBT are also shown to illustrate the effect of these two supplementary features. For readability consideration, all coefficients have been averaged over weekdays and weekends. The

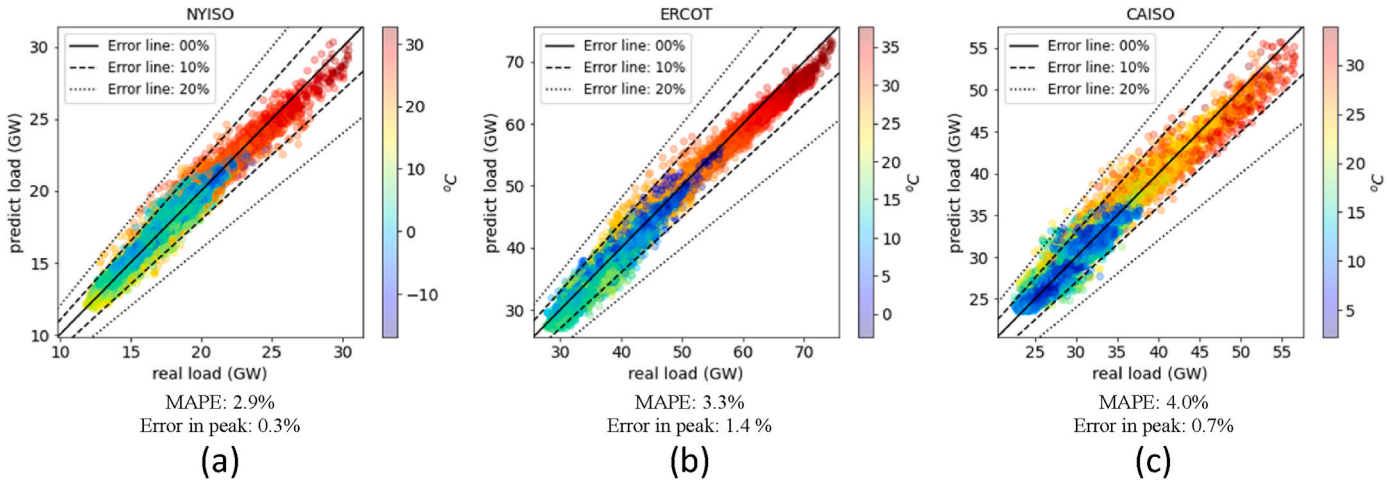

Fig. 5. Model simulated load compared to actual hourly load for BAs. (a) NYISO, (b) ERCOT, (c) CAISO.

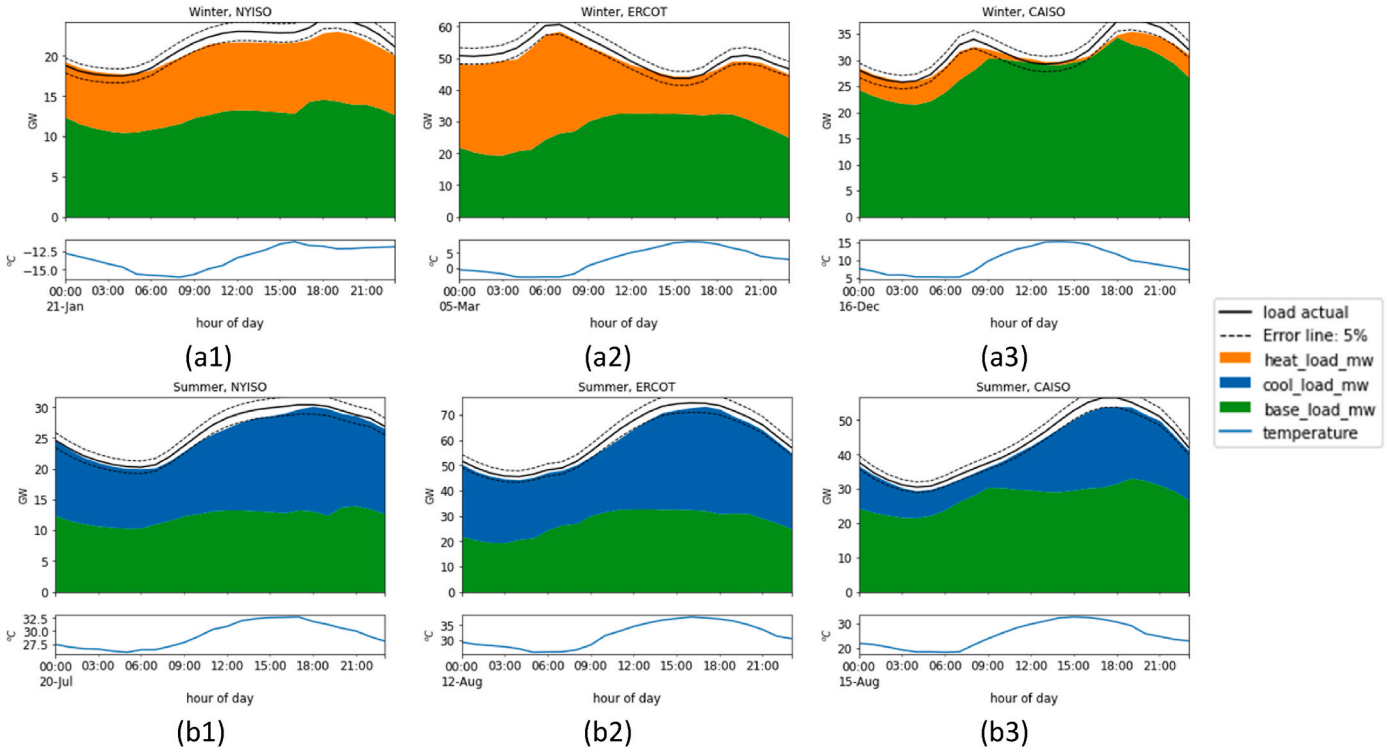

Fig. 6. Simulated load breakdowns on the peak load day of heating and cooling season for BAs and comparison with actual load profile. (a1)-(a3) heating season peak load day profiles for NYISO, ERCOT and CAISO respectively. (b1)-(b3) cooling season peak load day profiles for NYISO, ERCOT and CAISO respectively. Population weighted hourly temperatures for that day are shown in the lower plot.

confidence intervals are calculated by combining variances of the load zone regression models, while the propagation of uncertainty from source data and the synthesis of building floor area are not considered.

Although the regression model obtains results for each hour independently of other hours of a day, the coefficients show a nearly continuous behavior. The normalized heating coefficients are generally higher than those for cooling, which indicates the presence of lower efficiency electric resistance heating compared to air-conditioning. The cooling values in ERCOT ranging from 2.4 to 3.3 $Btu_e/(h \bullet m^2 \bullet^\circ C)$, higher than those for cooling in NYISO ranging from 1.7 to 2.6 $Btu_e/(h \bullet m^2 \bullet^\circ C)$. One explanation could be that building envelopes could be less efficient (e.g. not as well insulated and more air exchanges) in Texas. Or there might be a larger fraction of whole-house central cooling systems in Texas as compared to places like New York with larger fraction of window units that are more likely to be turned off when the occupant is away. But it is difficult to speculate exactly why, since these responses rely on estimates of what floor area is cooled and when. Given the Numerous competing factors, a detailed qualitative analysis on this behavior is out of the scope of this study.

Larger uncertainties are seen for electricity heating response estimators, especially for CAISO. One explanation is that California's relatively warmer winter temperatures and the low electrification rate of space heating compared to space cooling. According to our review of the ACS, RECS and CEBCS datasets (see Table S1), air conditioner possession rates are uniformly high among the three BA areas, whereas the ratios of electric heated building floor area are small for CAISO and NYISO,

Table 3
Comparison between our model and related models on modeling method and prediction accuracies.

developer	Modelling method	Modeled region	temporal framework	Selected areas for comparison	MAPE	Relative error in peak load
This study	Top-down	All BAs in NYISO, ERCOT and CAISO	Hourly	NYISO	2.9%	0.3%
				CAISO	3.3%	1.4%
				ERCOT	4.0%	0.7%
NREL [19,54]	Bottom-up	County level – entire contiguous US	Hourly	NY	14.2%	1.1%
				CA	31.8%	30.5%
				TX	16.5%	10.3%
Pineau P. O. et al. [55]	Bottom-up	New York state	Hourly	NY	5.8%	1.3%
Sailor D. J., Vasireddy C [56].	Top-down	Houston, Los Angles, Seattle	Monthly	Houston	7.6%	
				Los Angles	17.8%	
				Seattle	12.2%	
Tung N.X. et al. [57]	Top-down	Vietnam	Hourly		~5% [a]	
Wang Y. and Bielicki J. M [24].	Top-down	2 load zones in northeastern US	Hourly		~3.5%	

a A '~' before model error means the paper doesn't explicitly show the errors, and the authors make a rough estimation based on the plots of the literature.

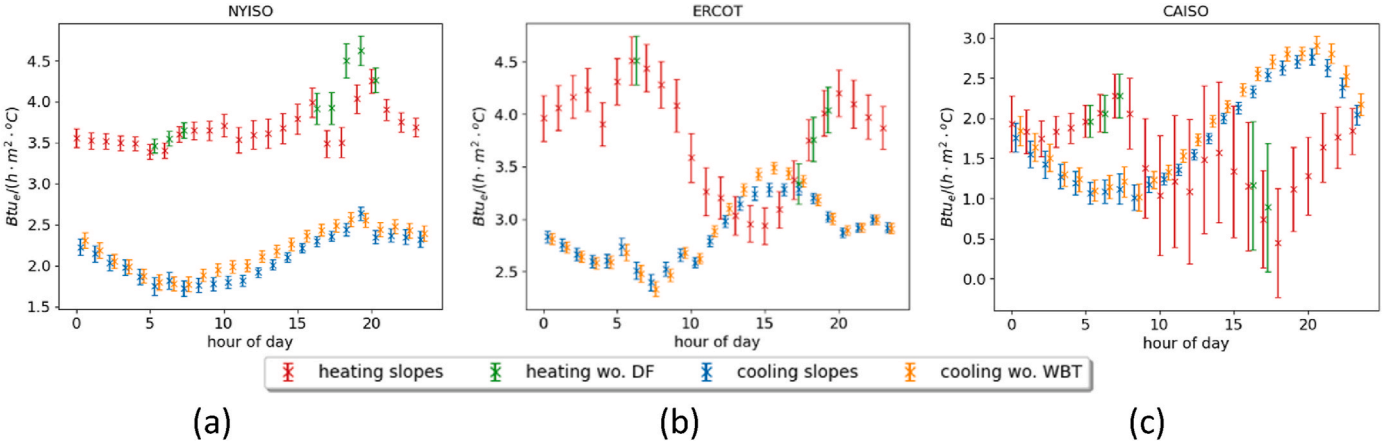

Fig. 7. Hourly coefficients of electricity response to ambient temperature for (a) NYISO, (b) ERCOT, (c) CAISO. Coefficients are averaged by weekdays and weekends, weighted by the numbers of days that are weekdays and weekends. Error bars show confidence interval of the individual regression models. Propagation of uncertainty from source data are not included.

around 22% and 12% on average respectively. The warmer winters and the low utilization rate of electric heating make electricity heating response to temperature too weak to be confidently extracted from the grid load. However, because of the mild temperatures and low penetration of electric heating, this does not have a major effect on overall load prediction in CAISO.

Both CAISO and ERCOT heating coefficients show daily patterns that roughly mirror those of the cooling coefficients. For example, the heating coefficients of ERCOT peaks in the early morning and at evening, while the cooling dips during those periods, and vice versa. This might be explained by the larger solar gain and larger elevations in afternoon temperatures during winter months for the two BAs, which can be observed in Fig. 6. Another observation is that the daily average heating slope of ERCOT is 34% higher than cooling, while this number

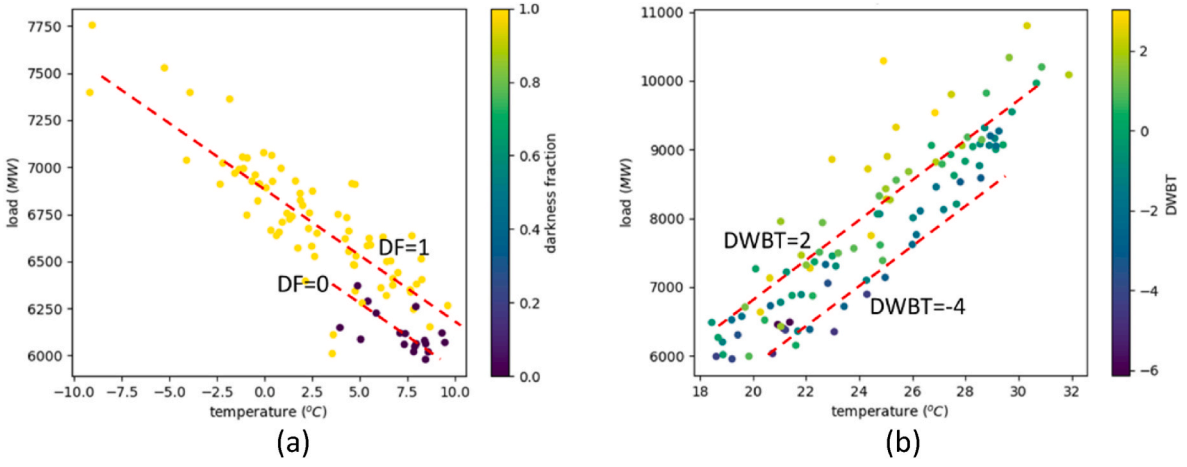

Fig. 8. The effect of (a) DF and (b) DWBT on the disaggregated heating and cooling load on hour 18 in NYIS-ZONJ.

for NYISO is around 76%. This is in accordance with the fact that TX has installed more heat pumps than NY, where less than 1% of the overall households have heat pump installed [58]. NYISO heating slopes seem to be shaped after people's activity schedules rather than the effect of daily temperature variation.

5.2. The effect of darkness fraction and wet bulb temperature

DF and DWBT (see Eq. (1) and (2) and Eq. (3) and (4) for the definition of DWBT) have an impact on the disaggregation results of load, which can be seen from Fig. 8. Fig. 8(a) shows an example of the linear model of the heating-induced temperature range for 6pm (hour 18) on weekdays. The effect of the darkness fraction is intuitive in that winter's artificial lighting is needed earlier than summer as daylight periods are shorter. Therefore, the extra nighttime illumination demand in winter must be disaggregated from the heating demand when fitting one regression model to the same hour for a whole year. Similarly, DWBT is an indicator of the amount of latent cooling load during humid summers, which is illustrated in Fig. 8(b). It's straightforward to find that for hours with WBTs higher than the mean relationship between DBT and WBT (i.e. positive DWBT), the overall demands tend to be higher, and vice versa. Model improvement considering DF and WBT can be found in Supplementary Fig. S2.

5.3. Potential for long-term load forecasting

To evaluate if the model fitted to input data for a single year (2019) is capable of forecasting over a long-term period, the model is tested by simulating grid load for a past 10-year period, from 2010 to 2019. As the usage of HVAC systems and the pool of existing appliances has been generally unchanged over the past decades, the disaggregated heating and cooling loads are assumed to vary only with the change of building stock floor area and the fraction of space that is either cooled or heated by electricity. Thus, the prediction intakes annual weather condition records and the interpolated building floor area that consumes electricity for heating or cooling for each year (Eqs. (6)–(8)). Load zone level prediction error can be found in Table 4. CAISO load data prior to 2019 is not available to the authors, so it is not included. NYISO and ERCOT MAPEs and relative errors in peak load fall within 5% for any year in the decadal horizon, except one outlier. Although MAPE generally increases as the gap between the prediction year and the model-fitting year (2019) increases, this is a very promising result for the use of this model to maintain reasonable accuracy in predicting loads in future years.

5.4. Robustness under extreme weather conditions

The model shows robust results for predicting demand under extreme weather conditions. Fig. 9 shows hourly load prediction breakdowns and temperatures for one of the hottest or coldest two-day windows for each BA in the past 10 years. Results show good alignment between the estimated grid loads from the model and utility records, with hourly error rate under 5% for most of the hours shown. Successfully predicting demand under extraordinary conditions is vital for utilities to plan for a stable, resilient and robust power system.

Table 4
Model testing error for past 10 years. Note, the model is fitted on year 2019 data, and applied to each of the preceding year with temperature records and interpolated building floor area and electricity penetration as model input. Also note, CAISO is not shown because CA BAs and load zones grid data are not available to the authors other than year 2018 and 2019.

	MAPE (%)		Error in peak (%)	
	NYISO	ERCOT	NYISO	ERCOT
2019	2.9	3.3	0.3	1.4
2018	3.1	3.4	1.8	2.6
2017	3.0	3.9	0.6	0.2
2016	3.1	3.9	3.3	3.2
2015	3.7	4.0	2.7	0.6
2014	4.3	4.0	2.9	3.3
2013	4.7	4.3	5.5	0.4
2012	4.6	4.3	2.4	1.3
2011	4.6	4.3	3.9	2.6
2010	4.8	4.2	1.3	5.0

6. Conclusion

In this study, we develop an open-source segmented multiple regression model to disaggregate building thermal response from electric load data with high temporal resolution over utility load-zone scale geographic areas. In addition to ambient temperature, time of day, weekend or weekday, and conditioned floor area, model fit is improved using the deviation of WBT from a mean DWT-WBT relationship and the presence of daylight. We apply our model to 27 load zones or BAs across three U.S. states, each of which has unique climatic and load characteristics. We validate model accuracy by comparing modeled hourly energy usages to publicly available datasets at multiple spatial aggregation levels.

Results show that our model is generally able to predict hourly grid load at BA aggregation within 5% MAPE, 1.5% relative error in peak load, and 2-h uncertainty in time at which the daily peak occurs. The maximum reaggregated load error for any hour remains under 20% for all three BAs. By comparing with existing literature, we show that the model's hourly load prediction accuracy and repeatability is as good as, if not better than, the most state-of-the-art models of its kind. These results hold across the study regions of NYISO, CAISO and ERCOT.

The model, along with coefficients obtained from a single year's data, is applied to backcast ten years of historical loads, using historical weather data, and a linear interpolation that applied to electrically conditioned building floor area. The comparison between actual hourly loads and estimates is found to be good. This validates the underlying assumption that the normalized heating and cooling coefficients remain invariant, at least in the short term. This suggests that the model can be used to predict loads into the future with acceptable error even as floor area evolves and the adoption of heating electrification accelerates. Thus, the model can be applied to estimate load profiles under various growth scenarios and evaluate demand side measures to reduce distribution grid investments and improve grid reliability. Such study is especially important for future grids with high penetration of intermittent resources.

With the model validated and tested, the daily profiles of disaggregated heating and cooling coefficients are presented. The electric cooling response is 2–4 $Btu_e/(h \bullet m^2 \bullet^\circ C)$ in most load zones. For ERCOT and NYISO, the electric heating responses range from 3 to 5 $Btu_e/(h \bullet m^2 \bullet^\circ C)$. The heating coefficients of CAISO have higher uncertainty; however, the total reaggregated load error remains low because of low heating demand and low penetration of electric heating in CAISO. Therefore, our model returns reliable load prediction, except when one is specifically interested in the disaggregated heating demand in moderate climates with minimal electric heating penetration. With these results available, as well as the building stock floor area datasets we presented in this study, projections of future demand can be easily obtained on a more-refined spatial resolution (e.g. Public Use Microdata Areas (PUMA)). With this unprecedented combination of temporal, geographic, and sectoral detail and coverage, this study informs not only transmission but distribution system planning to fulfill future demand scenarios, which is typically not available from existing studies [7].

The presented method of load disaggregation and forecasting provides a valuable input for future studies on predicting building load evolution and for integration into capacity expansion and grid operation models. However, the model is not designed to capture scenarios, policies or incentives that could significantly alter space occupancy

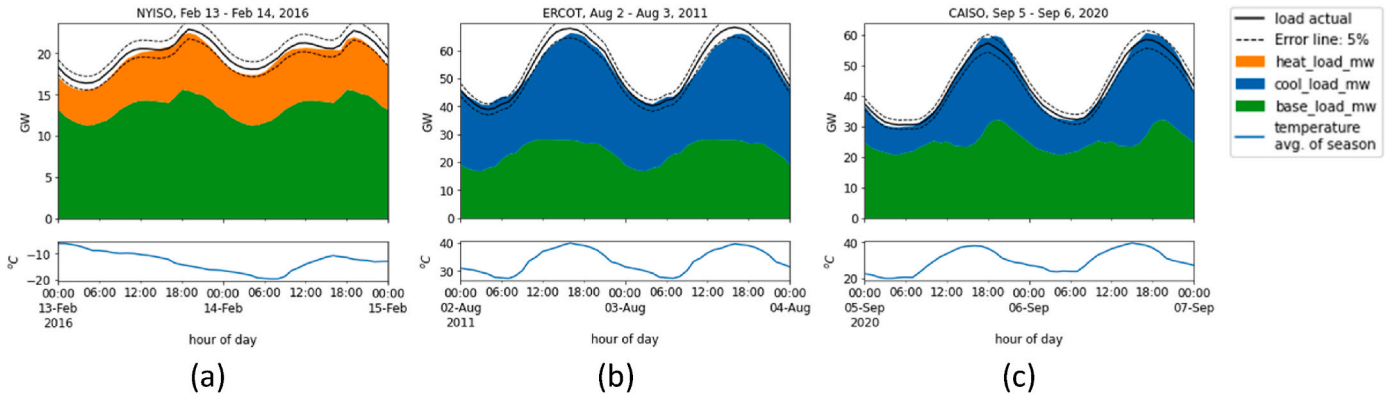

Fig. 9. Model testing with extreme weather conditions in 10-year period. Each two-day window is shown at (a) One of the coldest temperature records in New York's history on Feb.13 to Feb. 14, 2016. (b) An unprecedented heat wave across Texas and Southern Plains of US on Aug.2 to Aug. 3, 2011. (c) One of the hottest temperature records in California on Sep.5 to Sep. 6, 2020.

behavior, as happened with stay at home directives with COVID or significant improvements to building envelopes. Future work could focus on identifying residential and commercial breakdowns of the thermal response. Another extension of the present work could be to project growth in electricity demand from "electrifying" current fossil fuel-based space and water heating. These studies can be performed under various growth assumptions for housing, conditioned floor area, heat pump deployment, improved efficiency, flexible loads (e.g. demand response), renewable energy and storage capacity expansion, and greenhouse gas emission reduction targets.

Credit author statement

Yinbo Hu: Methodology, Software, Validation, Writing – original draft. Michael Waite: Conceptualization, Methodology, Software, Writing – review & editing. Evan Patz: Methodology, Software, Writing – review & editing. Bainan Xia: Writing – review & editing, Data curation. Yixing Xu: Writing – review & editing. Daniel Olsen: Writing – review & editing. Naveen Gopan: Writing – review & editing. Vijay Modi: Supervision, Conceptualization, Methodology, Writing – review & editing

Declaration of competing interest

The authors 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

Various data and code used in this study are available at https://doi.org/10.7916/k2k4-p867. This dataset provides: 1) data and code used to disaggregate utility loads for year 2019. 2) data for backcasting hourly load for the past 10-year period.

Acknowledgement

This work was funded in part via a collaboration with Breakthrough Energy. The authors would like to thank Dan Livengood for very useful insights and suggestions into this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.esr.2023.101175.

References

[1] U.S. Energy Information Administration (EIA), Use of energy explained: energy use in homes. https://www.eia.gov/energyexplained/use-of-energy/homes.php.
[2] K. Cleary, K. Palmer, Federal climate policy 106: the buildings sector. https://www.rff.org/publications/explainers/federal-climate-policy-106-the-buildings-sector/#:~:text=The%20buildings%20sector%2C%20which%20includes,heating%2C%20cooling%2C%20and%20cooking, 2021.
[3] U.S. Energy Information Administration (EIA), Residential energy consumption Survey (RECS). https://www.eia.gov/consumption/residential/.
[4] U.S. Energy Information Administration (EIA), Commercial building energy consumption Survey (CBECS). https://www.eia.gov/consumption/commercial/.
[5] T. Conlon, M. Waite, Y. Wu, V. Modi, Assessing trade-offs among electrification and grid decarbonization in a clean energy transition: application to New York State, Energy 249 (2022), 123787.
[6] Y. Wang, J. Wang, W. He, Development of efficient, flexible and affordable heat pumps for supporting heat and power decarbonisation in the UK and beyond: review and perspectives, Renew. Sustain. Energy Rev. 154 (2022), 111747.
[7] M. Waite, V. Modi, Electricity load implications of space heating decarbonization pathways, Joule 4 (2020) 376–394.
[8] J.J. Buonocore, P. Salimifard, Z. Magavi, J.G. Allen, Inefficient building electrification will require massive buildout of renewable energy and seasonal energy storage, Sci. Rep. 12 (2022) 1–9.
[9] C.F. Reinhart, C.C. Davila, Urban building energy modeling–A review of a nascent field, Build. Environ. 97 (2016) 196–202.
[10] H. Cho, Y. Goude, X. Brossat, Q. Yao, Modeling and forecasting daily electricity load curves: a hybrid approach, J. Am. Stat. Assoc. 108 (2013) 7–21.
[11] M. Imani, Electrical load-temperature CNN for residential load forecasting, Energy 227 (2021), 120480.
[12] Y.J. Huang, J. Brodrick, A bottom-up engineering estimate of the aggregate heating and cooling loads of the entire US building stock. Proceeding of the ACEEE Summer Study on Energy Efficiency in Buildings, 2000. https://escholarship.org/uc/item/5hv2t661.
[13] S.K. Firth, K.J. Lomas, A. Wright, Targeting household energy-efficiency measures using sensitivity analysis, Build. Res. Inf. 38 (2010) 25–41.
[14] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Build. Environ. 45 (2010) 1683–1697.
[15] S. Heiple, D.J. Sailor, Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles, Energy Build. 40 (2008) 1426–1436.
[16] K.P. Hallinan, J.K. Kissock, R.J. Brecha, A. Mitchell, Targeting Residential Energy Reduction for City Utilities Using Historical Electrical Utility Data and Readily Available Building Data, 2011.
[17] A.S. Hopkins, A. Lekov, J. Lutz, G. Rosenquist, L. Gu, Simulating a Nationally Representative Housing Sample Using EnergyPlus, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2011.
[18] R. Hendron, C. Engebrecht, Building america House Simulation Protocols (Revised), National Renewable Energy Lab.(NREL), Golden, CO (United States), 2010.
[19] E. Hale, H. Horsey, B. Johnson, M. Muratori, E. Wilson, B. Borlaug, C. Christensen, A. Farthing, D. Hettinger, A. Parker, The Demand-Side Grid (Dsgrid) Model Documentation, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2018.
[20] U.S. Energy Information Administration (EIA), Office of integrated analysis and forecasting, impacts of temperature variation on energy demand in buildings, in: Annual Energy Outlook 2005, 2005, pp. 55–58.
[21] N. Shorr, R.G. Najjar, A. Amato, S. Graham, Household heating and cooling energy use in the northeast USA: comparing the effects of climate change with those of purposive behaviors, Clim. Res. 39 (2009) 19–30.

[22] M. Bessec, J. Fouquau, The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach, Energy Econ. 30 (2008) 2705–2721.
[23] M. Waite, E. Cohen, H. Torbey, M. Piccirilli, Y. Tian, V. Modi, Global trends in urban electricity demands for cooling and heating, Energy 127 (2017) 786–802.
[24] Y. Wang, J.M. Bielicki, Acclimation and the response of hourly electricity loads to meteorological variables, Energy 142 (2018) 473–485.
[25] J. Moral-Carcedo, J. Vicéns-Otero, Modelling the non-linear response of Spanish electricity demand to temperature variations, Energy Econ. 27 (2005) 477–494.
[26] J.K. Kissock, T.A. Reddy, D.E. Claridge, Ambient-temperature Regression Analysis for Estimating Retrofit Savings in Commercial Buildings, 1998.
[27] T. Ihara, Y. Genchi, T. Sato, K. Yamaguchi, Y. Endo, City-block-scale sensitivity of electricity consumption to air temperature and air humidity in business districts of Tokyo, Japan, Energy 33 (2008) 1634–1645.
[28] S.R. Mohandes, X. Zhang, A. Mahdiyar, A comprehensive review on the application of artificial neural networks in building energy analysis, Neurocomputing 340 (2019) 55–75.
[29] B. Yildiz, J.I. Bilbao, A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, Renew. Sustain. Energy Rev. 73 (2017) 1104–1122.
[30] F. Apadula, A. Bassini, A. Elli, S. Scapin, Relationships between meteorological variables and monthly electricity demand, Appl. Energy 98 (2012) 346–356.
[31] J.C. Lam, K.K. Wan, D. Liu, C. Tsang, Multiple regression models for energy use in air-conditioned office buildings in different climates, Energy Convers. Manag. 51 (2010) 2692–2697.
[32] N. Nassif, Regression models for estimating monthly energy consumptions in schools in hot and humid climates, Build. Eng. 118 (2012).
[33] T. Reddy, D. Claridge, Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption, Energy Build. 21 (1994) 35–44.
[34] B. Lin, H. Liu, China's building energy efficiency and urbanization, Energy Build. 86 (2015) 356–365.
[35] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (2009) 1819–1835.
[36] W. Ma, S. Fang, G. Liu, R. Zhou, Modeling of district load forecasting for distributed energy system, Appl. Energy 204 (2017) 181–205.
[37] N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability, IEEE Trans. Power Syst. 16 (2001) 498–505.
[38] S. Paudel, P.H. Nguyen, W.L. Kling, M. Elmitri, B. Lacarrière, O. Le Corre, Support vector machine in prediction of building energy demand using pseudo dynamic approach, in: Proceedings of ECOS 2015-The 28th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, 2015.
[39] S.L. Wong, K.K. Wan, T.N. Lam, Artificial neural networks for energy analysis of office buildings with daylighting, Appl. Energy 87 (2010) 551–557.
[40] K. Li, H. Su, Forecasting building energy consumption with hybrid genetic algorithm-hierarchical adaptive network-based fuzzy inference system, Energy Build. 42 (2010) 2070–2076.
[41] E. Gupta, Global warming and electricity demand in the rapidly growing city of Delhi: a semi-parametric variable coefficient approach, Energy Econ. 34 (2012) 1407–1421.
[42] R. Nadimi, F. Ghaderi, A hybrid TSK-FR model to study short-term variations of the electricity demand versus the temperature changes, Expert Syst. Appl. 36 (2009) 1765–1772.
[43] U.S. Federal, Emergency management agency (FEMA), Hazus general building stock database, in: Hazus 5.1, 2019. https://msc.fema.gov/portal/resources/hazus.
[44] U.S. Census Bureau, 2015-2019 American community Survey 5-year estimates. https://www.census.gov/programs-surveys/acs, 2020.
[45] M.A. Brown, M. Cox, B. Staver, P. Baer, Modeling climate-driven changes in US buildings energy demand, Climatic Change 134 (2016) 29–44.
[46] M.J. Kotchen, L.E. Grant, Does daylight saving time save energy? Evidence from a natural experiment in Indiana, Rev. Econ. Stat. 93 (2011) 1172–1185.
[47] J. Moral-Carcedo, J. Pérez-García, Time of day effects of temperature and daylight on short term electricity load, Energy 174 (2019) 169–183.
[48] M. López, Daylight effect on the electricity demand in Spain and assessment of Daylight Saving Time policies, Energy Pol. 140 (2020), 111419.
[49] K.K. Wan, D.H. Li, D. Liu, J.C. Lam, Future trends of building heating and cooling loads and energy consumption in different climates, Build. Environ. 46 (2011) 223–234.
[50] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, J.-N. Thépaut, ERA5 hourly data on single levels from 1959 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (2018).
[51] U.S. Energy Information Administration (EIA), Hourly electric grid monitor. https://www.eia.gov/electricity/gridmonitor/dashboard/electric_overview/US48/US48.
[52] E.J. Wilson, A. Parker, A. Fontanini, E. Present, J.L. Reyna, R. Adhikari, C. Bianchi, C. CaraDonna, M. Dahlhausen, J. Kim, Section 5.1.3 how much uncertainty is there in the results, in: End-Use Load Profiles for the US Building Stock: Methodology and Results of Model Calibration, Validation, and Uncertainty Quantification, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2022, pp. 346–347.
[53] Y. Jang, E. Byon, E. Jahani, K. Cetin, On the long-term density prediction of peak electricity load with demand side management in buildings, Energy Build. 228 (2020), 110450.
[54] E. Hale, H. Horsey, B. Johnson, M. Muratori, E. Wilson, B. Borlaug, C. Christensen, A. Farthing, D. Hettinger, A. Parker, J. Robertson, M. Rossol, G. Stephen, E. Wood, B. Vairamohan, Demand-Side Grid Model (Dsgrid) Data from the Electrification Futures Project, EFS), 2018, https://doi.org/10.25984/1823248.
[55] P.-O. Pineau, P.-O. Caron-Perigny, G. Tarel, A. Borelle, L. Pollux, Aggregate load profile and decarbonization: impacts of actionable demand drivers in New York, Energy Strategy Rev. 42 (2022), 100868.
[56] D.J. Sailor, C. Vasireddy, Correcting aggregate energy consumption data to account for variability in local weather, Environ. Model. Software 21 (2006) 733–738.
[57] N.X. Tung, N.Q. Dat, T.N. Thang, V.K. Solanki, N.T.N. Anh, Analysis of Temperature-Sensitive on Short-Term Electricity Load Forecasting, 2020 IEEE-HYDCON, IEEE, 2020, pp. 1–7.
[58] U.S. Census Bureau, American housing Survey (AHS). https://www.census.gov/programs-surveys/ahs.html, 2019.