Applied Energy 235 (2019) 1561–1582

Contents lists available at ScienceDirect

Applied Energy

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

Evaluating regional climate-electricity demand nexus: A composite Bayesian predictive framework

Sayanti Mukherjee[a,*], C.R. Vineeth[b], Roshanak Nateghi[b,c]

[a] Department of Industrial and Systems Engineering, 411 Bell Hall, University at Buffalo, The State University of New York (SUNY), Buffalo, NY 14260, USA
[b] School of Industrial Engineering, 315 N. Grant Street, Purdue University, West Lafayette, IN 47907, USA
[c] Environmental and Ecological Engineering, 610 Purdue Mall, Purdue University, West Lafayette, IN 47907, USA

HIGHLIGHTS

- Climate-sensitivity of load is asymmetric; high demand is more sensitive than moderate demand.
- Dew point temperature is the key predictor of high- and moderate-intensity demands.
- Residential high-intensity demand is more sensitive to precipitation than moderate demand.
- Higher wind speeds decrease both the high- and moderate-intensity end-use demands.
- Moderate intensity demand is more price-sensitive than high intensity demand.

ARTICLE INFO

Keywords:
Climate-demand nexus
Electricity demand modeling
Bayesian learning
Predictive modeling
Grid reliability
Dew point temperature

ABSTRACT

Climatic variations significantly influence the shape of end-use electricity demand curves. Although the climate sensitivity of end-use electricity demand is well-established, projecting medium- and long-term future demand remains a significant challenge—mostly due to a multitude of uncertainties involved in the modeling process. In this paper, we leveraged a state-of-the-art Bayesian approach to develop rigorously validated regional prediction models of the climate–demand nexus conditioned on the intensity level of demand. The prediction models were developed for the residential and commercial sectors for the top eight energy-intensive states in the U.S. A key contribution of this work was to illustrate the asymmetry in the sensitivity of load to climate. More specifically, our results demonstrated a greater sensitivity of the high-intensity end-use demand to climate variability as compared to the moderate-intensity end-use demand. In addition, our results helped identify mean dew point temperature as the key predictor of the climate-sensitive portion of both residential and commercial electricity demands, irrespective of the demand intensity levels. Wind speed was identified as the second most important predictor of the high-intensity (i.e., ≥3rd quartile) end-use demand, while electricity price was found to be the key predictor of the moderate-intensity (i.e., < 3rd quartile) end-use demand. The influence of precipitation on the residential and commercial sectors' moderate end-use demand was found to be more variable. Precipitation was found to influence the commercial sector's electricity demand more significantly compared to the residential sector's demand.

1. Introduction

Projecting medium- and long-term electricity demand is a challenging task due to the uncertainties associated with both climatic and non-climatic factors such as regional climate and geographical characteristics, infrastructure types, policy incentives, technological advancements, socio-economic factors, and population shifts. Many of the existing energy-economy models can facilitate electricity demand projections under policy change and technological uncertainties; however, such models are unable to adequately account for the influence of climate variability on the demand [1]. Previous research established various weather and climate variables such as dew point temperature, maximum daily temperature, wind speed and precipitation as key predictors of climate-induced energy demand [2–4]. In addition, climatic factors such as ambient temperature and humidity have been shown to play an important role in determining the quantity of electricity consumed by a household or a commercial building [5–7] or an industrial firm [8]. In fact, a major portion of the climate-influenced

* Corresponding author.
E-mail address: sayantim@buffalo.edu (S. Mukherjee).

https://doi.org/10.1016/j.apenergy.2018.10.119
Received 18 April 2018; Received in revised form 6 October 2018; Accepted 28 October 2018
Available online 27 November 2018

electricity demand is attributed to space conditioning such as heating or cooling needs per the season [4,9]. Although the significant influence of climate variability and climate change on the shape of seasonal demand curves have been established, evaluating regional climate-demand nexus still remains a challenging task. End-use demand varies widely across different regions in the U.S., with some states projected to experience decreased loads under future warming scenarios, while others are projected to experience an increase in end-use demands under similar scenarios [5]. The ability to credibly characterize the climate-induced shifts in end-use demand is critical for devising appropriate policy decisions at the state and federal levels as well as the regional electricity markets.

While sectoral electricity demands show significant differences in their patterns and climate sensitivity, aggregate electricity demand has been found to be rather insensitive to climate, mostly because of the composition effects [8]. Thus, we developed prediction models for climate-sensitive load in the residential and commercial sectors in the most energy-intensive regions of the country [5]. We focused on the residential and commercial sectors because the electricity consumption levels in these two sectors have been found to be most sensitive to climatic variability and change [10]. However, our methodology can be easily extended to the industrial or transportation sectors. To develop our proposed prediction models, we selected the states of California (CA), Florida (FL), Illinois (IL), Louisiana (LA), New York (NY), Ohio (OH), Texas (TX), and Washington (WA). The selection of these states was based on the fact that: (a) they rank the highest in terms of the fraction of total energy consumption in the U.S., and (b) their diverse geographical distribution presents a unique case in offering a generalized understanding of the climate sensitivity of electricity demand [11,12].

We developed probabilistic composite predictive models of end-use demand based on a Bayesian non-parametric tree-ensembles algorithm, named Bayesian Additive Regression Trees (aka BART) [13,14]. This modeling approach was established as an effective methodology for characterizing the complex nexus between climate variability and electricity demand in previous literature [1–3,9]. We hypothesized the degree of the climate-sensitivity of load to be different for high-intensity end-users compared to the moderate-intensity end-users. To test the hypothesis, we developed two separate models, namely, high-intensity consumption model (HCM) (i.e., the end-use demand observations equal to or greater than the 3rd quartile) and moderate-intensity consumption model (MCM) (i.e., the observations below the 3rd quartile) for each of the residential and commercial sectors. Contrary to the previous studies of modeling each state separately [5,6], we developed a *single 'regional'* model to simultaneously characterize the climate-demand nexus across all the different states included in the analysis. Thus, each of the HCM/MCM models is referred to as a 'composite' model. We conjecture that the prevalence of separate, single-state models in the existing literatures [5,6] can be attributed to the commonly used linear model architectures. Developing composite models for multiple states using linear methodologies would likely lead to poor prediction capability. This is because the 'rigid' linear structure cannot capture the complex heterogeneities and non-linear variabilities in demographic composition, socio-economic patterns and topography across the states. We hypothesized that our proposed flexible, probabilistic predictive framework could in fact account for the complex nexus between climate variability and demand across the different regions. To test this hypothesis, we assessed the statistical performance of the composite models based on both (i) their goodness-of-fit, and (ii) their predictive accuracy (aka out-of-sample error estimation). We selected our model based on the generalizability of the developed models in addition to their capability in capturing the structure of the data. We also validated the models through leveraging a validation dataset that was not used in training and testing the models.

We have organized the article as follows. In Section 2, we start by presenting a brief overview of the literature on climate–demand nexus modeling and highlighting the research gaps in the existing body of knowledge. We discuss our data in Section 3. Section 4 summarize our methodologies, followed by Section 5 where we discuss our results and model validation. We conclude in Section 6 by summarizing our key findings and delineating future research directions.

2. Prior art of modeling climate–demand nexus

Over the past few decades, there has been a growing interest in analyzing the residential and commercial electricity demand trends as a function of different factors such as occupant behavior, equipment efficiency, climate change and variations, types of the households or buildings, etc. The studies vary widely in terms of the scale of the assessment, ranging from single and multiple households or buildings to single and multiple states and regional-level analyses.

Previous research has analyzed the energy demand at the level of individual commercial buildings under idealized conditions. Energy benchmarking models have also been developed for office buildings in all census regions of the U.S. [15]. Another study analyzed commercial energy consumption per unit floor area, considering 612 prototypical commercial buildings in a representative district incorporating stochastic occupant behavior, various zoning configurations, HVAC systems, and building construction characteristics [16]. The key predictors of the cooling demand intensity in commercial buildings across the U.S. were identified via leveraging statistical machine learning algorithms in order to improve building-level operational management policies [17]. Another research analysis simulated the energy consumption in the U.S. commercial buildings to estimate the cost of overcooling the commercial buildings [18]. Analyzing energy demands at household levels have also recently become more prevalent in the scientific literature. For example, one study investigated the relationship between energy demand in a representative residential building and the various factors such as heating and cooling degree days, total heating and cooling area, household size, price of electricity, housing type and age, neighborhood density, ownership and income [19]. Another study analyzed the cooling demand in residential buildings, considering the behavioral, physical and socio-economic parameters on cooling load. The objective of the study was to enable a more informed appraisal of interventions or incentives to improve energy efficiency [20]. In another article, researchers estimated the building sector's end-use demand intensity (Kilowatt-hour per square-meters of floor area, i.e., KWh/m^2), considering the different functions of the building and based on various building types such as residential single family, residential multi-family, office, store, education, health, warehouse and other commercial. While the above-mentioned studies paved the path to gaining an understanding of the energy consumption patterns at the individual commercial building or household level, these studies had a comparatively limited scope and spatial coverage. Thus, they did not account for regional variabilities and did not investigate the influence of the regionally diverse climatic variability on the electricity consumption.

Significant research efforts have been directed towards modeling the effects of climate variability and climate change on the regional energy demand in the residential and commercial sectors to help the electric utilities with more effective integrated adequacy planning [21–31]. Parametric generalized linear models have been widely used to establish the relationship between energy demand and climate and/or non-climatic predictors. Multiple linear regression (MLR) and time-series analysis have been leveraged to predict the seasonal electricity demands in the state of New South Wales in Australia under future climate change scenarios [32]. In a different study, influence of temperature on the energy demand in the residential and commercial sectors have been assessed leveraging a generalized linear model, while controlling for socio-economic factors [7]. However, in both the studies the authors did not consider other potentially relevant climate factors such as humidity, precipitation or wind speed [7,32]. MLR analysis is also used to analyze the relationship between the monthly electricity demand, and the meteorological variables and calendar effect for Italy to identify the influence of weather on aggregate national electricity demand; however, the study did not account for any socio-economic variabilities in the analysis [33]. MLR modeling approach is adopted to estimate the influence of various climatic and socio-economic factors on the monthly regional electricity demand in the Greek interconnected power systems [34]. Another research focused on developing a two-stage least square methodology to analyze the energy demand for all 50 states in the continental U.S. The author used the cross-sectional data collected for the year 1988 to estimate energy demand and price elasticities for the residential, commercial and industrial sectors [35]. Separate parametric linear models were developed to estimate the sensitivity of demand for electricity and natural gas to climate variability in each of the top eight energy intensive states viz., California, Louisiana, Texas, Florida, Washington, Illinois, Ohio and New York [6]. The separate, state-level linear regression-based models used the (heating and cooling) degree-day approach, and did not account for socio-economic (non-climatic) factors. A follow-up study based on the same methodology conducted a scenario-based sensitivity analysis to estimate the effect of climate change on the consumption of electricity for the residential and commercial sectors in each of the top eight energy intensive states in the U.S. [5]. Time series regression model was leveraged to analyze the relationship between temperature and energy demand for the state of Maryland [36].

Despite the advancements in understanding the climate sensitivity of the electricity demand in the residential or commercial sectors, there remain several knowledge gaps in the existing body of literature. Some of the limitations and knowledge gaps are summarized below:

(a) Majority of the household or building-level energy prediction models do not adequately account for the role of climate variability.
(b) Most of the regional studies used the conventional degree-days approach (i.e., the heating or cooling degree days) to establish the link between climate and electricity demand. However, previous research has demonstrated the inadequacy of the degree-days approach for capturing the climate sensitivity of electricity demand [3,4]. Moreover, many of the regional models are based on linear regression, developed using the highly correlated cooling and heating degree-days [37] which lead to biased inferencing. This is because the highly collinear variables might mask the effect of the others [3].
(c) Many of the existing approaches leverage linear, deterministic models to characterize the complex and non-linear climate-demand nexus. However, previous research demonstrated that generalized linear models often led to reduced predictive performance and generalizability due to their inadequacy in capturing the complex nonlinear climate-demand nexus [1–4].
(d) No previous study has assessed the non-uniform sensitivity of end-use demand to climate variability.

To address the gaps identified above, we propose a generalized, probabilistic composite model to characterize the stochastic and complex nexus between climate variability and electricity load in the residential and commercial sectors across the eight most energy intensive states. Based upon our previous research [1,3,4,9] we hypothesized that predictive models using a Bayesian, non-parametric, ensemble-of-trees algorithm could facilitate accurate estimation of the complex climate–end-use electricity demand nexus across multiple different geographical regions (i.e., eight different states in the U.S.). To test the hypothesis, we rigorously assessed the statistical performance of the developed models. In addition, we hypothesize that the climate variables that govern the sensitivity of electricity load differ for high-intensity demand versus moderate-intensity demand across the states. To test this hypothesis, we developed two separate 'composite' models, namely, high consumption model (HCM) and moderate consumption model (MCM) for each sector (refer to Section 1). Leveraging our proposed framework, the climate-demand nexus for multiple regions can be characterized *simultaneously* using a single model. The advantage of the proposed composite modeling approach over the cumbersome, individual state-level approach is that electricity market operators and policy analysts can implement swift and efficient sensitivity analysis for multiple regions simultaneously, instead of running separate models for each state. This is especially useful since state demarcations are not necessarily accurate boundaries of climate variability. Moreover, while we developed our models for the top-eight most energy intensive states in the U.S., any state/region could be used in our generalized modeling framework. We believe that the proposed composite modeling framework is poised to improve the operational management of regional electricity markets and enhance reliability levels by better anticipating climate-induced shifts in demand.

3. Data source, description and visualization

The performance of a statistical machine learning model depends on various factors including the volume and quality of the empirical data used during the model development. Hence, it is imperative to assemble a reasonably sized high-quality database consisting of both electricity consumption, climate and socio-economic data. The data obtained for the current analysis covers a 26-year period, starting from 01-Jan-1990 to 31-Dec-2015.

3.1. Input data preparation

3.1.1. Electricity data

Information pertaining to the state-level electricity consumption was retrieved from the database published by the United States Energy Information Administration (EIA) [38]. The monthly electricity sales (in Megawatt hours) and electricity price (in cents/kWh) data for the residential and commercial sectors in the above-mentioned eight states were extracted from the EIA-861M database [38].

For each state under consideration, the sectoral electricity sales data was normalized using the state-wide monthly population to obtain the per capita monthly electricity sales, denoted by $E(m, y)$ (in Kilowatt hours). We first obtained the state-level annual population data from the U.S. Census, and then estimated the state-wide monthly population through a linear interpolation between the years of interest. Subsequently, to isolate the influence of climatic factors on the electricity demand per capita—in other words, to remove the non-climatic factors' effect from the time series electricity demand (electricity sales) data—we trend-adjusted the per capita electricity sales data. The methodology used to trend-adjust the data is described in the following section [3,6]:

At first, we calculate the yearly average electricity sales consumption $\bar{E}(y)$ from the monthly data over the entire period of study.

$$\bar{E}(y) = \frac{\sum_{y=1}^{n_{years}} \sum_{m=1}^{12} E(m, y)}{n_{years}}$$

Then, the adjustment factor F_{adj} for each year was calculated from the following equation:

$$F_{adj} = \bar{E}(y)^{-1} \sum_{m=1}^{12} E(m, y)$$

Finally, each month's electricity data, $E(m, y)$, was trend-adjusted through a simple division by adjustment factor for that year as shown below:

$$E_{adj}(m, y) = E(m, y)/F_{adj}(y)$$

Figs. 1a and 2a show the raw time series of the per capita electricity sales (consumption) data aggregated across the selected eight states in the residential and commercial sectors, respectively. Figs. 1b and 2b, on the other hand, show the trend-adjusted time series of the per capita

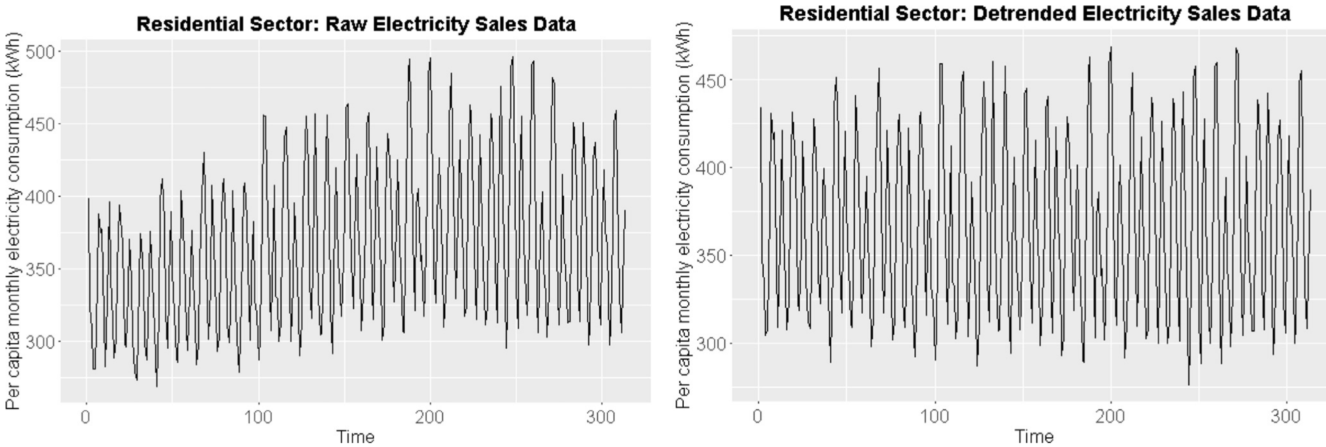

Fig. 1. Time-series distribution of per capita electricity sales in the residential sector: (a) raw data; (b) trend-adjusted data.

electricity sales data aggregated across the selected eight states in the residential and commercial sectors, respectively. The values on the x-axis range from 0 to 312 months, where 0 represents 01-Jan-1990 and 312 represents 31-Dec-2015. Each y-axis value indicates the per capita electricity sales for each of the eight states under consideration.

3.1.2. Climate and weather data

The periodicity of the electricity sales data dictated the periodicity of the climate data. The climate parameters were obtained from various sources as described in this section. These databases contain daily climate data from regional monitoring stations throughout the USA.

Monthly climate data was retrieved from the National Digital Forecast Database maintained by National Oceanic and Atmospheric Administration (NOAA) [39]. The monthly climate data was obtained from eleven different regional weather stations in each of the different states considered in our analysis; we aggregated the station-level data to compute the state-level climatic variable, leveraging the following formula:

$$X_{state}(m) = \frac{1}{N_s}\sum_{i=1}^{N_s} X_i$$

Here, N_s is the number of weather stations in a state and X represents the observed climate variable in the state during a month "m". In doing so, we assumed that every weather station reading is representative of the region in which it is situated. Finally, missing values in each climatic parameter are replaced via imputation. The climate data relevant to the current study is – "total precipitation (in mm)".

Daily weather data, aggregated across all the weather stations in a state, was obtained from National Climatic Data Center (NCDC) online database. We computed the *median* of the daily values to represent the monthly values of the following features—(a) Mean dew point temperature (in °C); (b) Mean wind speed (in m/s); and, (c) Maximum wind gust (in m/s).

Finally, the monthly-level climate and weather data were used in our analysis.

3.1.3. Socio-economic data

The data pertaining to economic indicators was obtained from the U.S. Department of Labor, Bureau of Labor Statistics. The dataset includes variables such as employment, unemployment, unemployment rate, gross state product (GSP), per capita income of the population and labor force [3,40]. However, we considered the features: (a) unemployment fraction, and (b) GSP of the states in our analysis because they were found to be important control variables in predicting the climate sensitive state-level electricity demand in previous studies [3,10]. In addition, since our analysis involved a multi-region time series analysis, these variables were found to better capture the socio-economic variabilities across the states during the years of the analysis. For example, unemployment fraction for the state of Illinois was as low as 4.1% in 1999 [February] and as high as 11.2% in 2009 [December]. On the other hand, Texas experienced the least variation over the years [4% in November-2000 to 8.4% in August-2009]. Inclusion of these control variables in our model was essential to isolate the effect of climate on end-use electricity demands.

The monthly gross state product of each state was converted to per capita income of that state by dividing it with appropriate state-wide population data for that month. Subsequently, the per capita income of

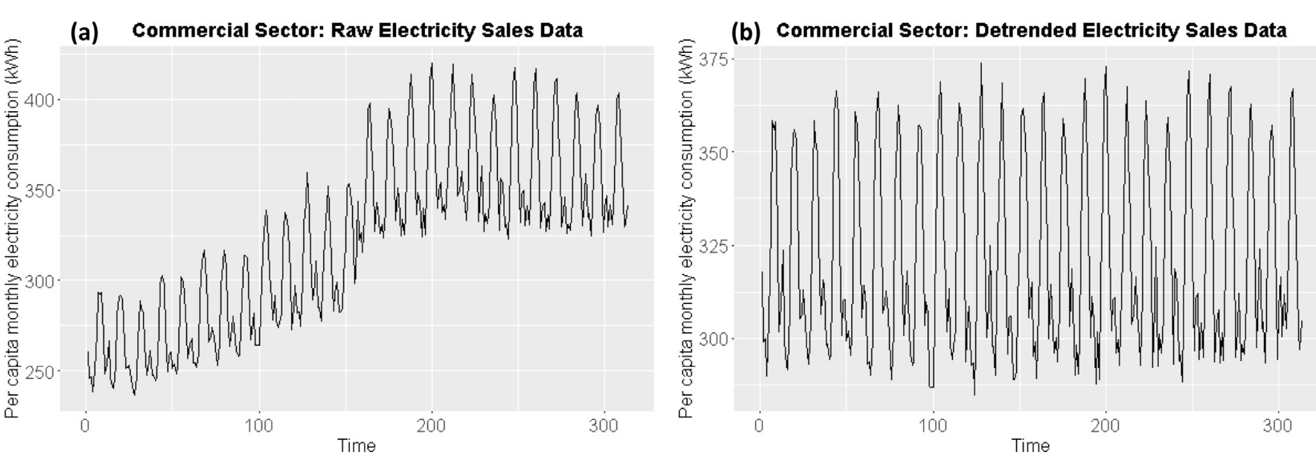

Fig. 2. Time-series distribution of per capita electricity sales in the commercial sector: (a) raw data; (b) trend-adjusted data.

the state and the electricity price was adjusted to account for the inflation over the entire period of study using the Consumer Price Index (specifically CPI Urban or CPI-U). The monthly CPI values were obtained from U.S Bureau of Labor Statistics (BLS) [41]. The standard baseline set by the BLS indicates that the CPI-U in the U.S. equals to 100, during the period of 1982–1984 [41]. Since our analysis was conducted at a monthly time-scale, we considered July 1983 (lying midway between 1982-January and 1984-December) as the baseline time-period for inflation adjustments. Thus, we adjusted our economic variables (per capita income of the state and the electricity price) using the Consumer Price Index (CPI) of 99.9 (i.e., the CPI for the month of July 1983) and the corresponding monthly CPI values during the period of our analysis from 1990 to 2015 [41]. The inflation adjustment factor was calculated as the ratio of the month's CPI to CPI in July 1983 (99.9). Based on that, the reported prices in all other months were adjusted by dividing each month's prices with this inflation adjustment factor. The benefit of following the standards set by BLS is that it facilitates direct comparison of the influence of GSP and electricity price on per capita electricity consumption in the future, irrespective of the time-period of analysis considered in any research study.

3.2. Description of final data-set used for model development

For each of the residential and commercial sectors, Table 1 summarizes the final dataset:

A portion of the final dataset, ranging from the year Jan 1990 till Dec 2011, was used to train and test our proposed statistical learning models, while the rest of the dataset (Jan 2012 till Dec 2015) was reserved for evaluating the predictive performance and validating our models.

3.3. Distribution and correlation of the input features and the response

The kernel distributions of the response variable, i.e., the per capita electricity sales data for the residential and commercial sectors in the eight states considered in our study are shown in Fig. 3. Kernel density estimation, refers to the nonparametric estimation of the probability density function of a random variable, which is the "per capita electricity sales" in the residential and commercial sector in this research. A kernel density function is used for representing the probability density function of a variable to avoid specific assumptions about the type of distribution of the data. From the individual states' kernel density plots, we observe that the distributions of end-use electricity demands are generally right-skewed (and bi-modal). Figs. 4 and 5 depict the kernel distributions of the per capita high-intensity and moderate-intensity electricity consumptions, respectively. The distribution of the high-intensity electricity demands in both the residential (Fig. 4a) and commercial (Fig. 4b) sectors are bi-modal and right-skewed, indicating a similarity in their distribution patterns. However, for the moderate-intensity demands, the pattern of residential end-use demand (Fig. 5a) is significantly different than that of the commercial sector (Fig. 5b); the former is bi-modal but not skewed, whereas the latter is tri-modal and left-skewed. In summary, it can be observed that the response variable, i.e., per capita electricity consumption (demand) aggregated across all the eight states, is not normally distributed in either of the residential or commercial sectors.

Fig. 6a and Fig. 6b show the seasonal variations of the high-intensity per capita monthly electricity demands aggregated across all the eight states in the residential and commercial sectors respectively. On the other hand, Fig. 7a and Fig. 7b show the seasonal variations of the moderate-intensity per capita monthly electricity demands in the residential and commercial sectors respectively, aggregated across all the eight states. It can be observed that the high-intensity electricity consumption significantly varies seasonally in both the residential (Fig. 6a) and commercial sectors (Fig. 6b), and it peaks during the summer months (July-August). On the other hand, the moderate-intensity demand in both the sectors do not show a significant variation across the different seasons (Fig. 7a, b).

A violin plot is a combination of a boxplot and a kernel density plot. In other words, in a violin plot a rotated kernel density plot is added on the two sides of the boxplot [42]. As observed from Figs. 8 and 9, violin plots provide an efficient visualization tool to compare the kernel distributions of the response variables—residential and commercial sector's electricity demands—on the same plot to help compare the two sectors' electricity demand. The violin plots in Fig. 8 show the characteristics of the residential and commercial sectors' per capita electricity consumption in the eight states, considered in this study. On the other hand, Fig. 9 shows the violin plots of the high-intensity and moderate-intensity per capita end-use electricity consumption in the residential and commercial sectors, aggregated across all the eight states. It can be observed from Figs. 8 and 9 that the residential electricity demand is much more heterogeneous and long-tailed compared to the commercial sector. For the combined eight states data, we also observe that the median values of high-intensity (Fig. 9a) and moderate-intensity (Fig. 9b) demands in the commercial sector are lower than that of the residential sector (Fig. 9). Moreover, in the commercial sector, the data points are clustered around median whereas in the residential sector the data points are scattered throughout the range, with a heavy concentration on the lower end of the range. Similar trends are also observed in the individual states' electricity demand distribution patterns (Fig. 8), except for the states of California and New York. In both these states, the commercial sector's median is higher, and the residential sector's data-points are clustered around the median.

Summary statistics of the higher and moderate intensity demands for the residential and commercial sectors are given in Table 2.

4. Methodology

We propose a composite predictive framework (depicted in Fig. 10) to evaluate the regional climate-electricity demand nexus. As discussed before, we collected data on state-level electricity consumption patterns, climate and weather information, and the relevant socio-economic information from various publicly available databases for all the eight states selected in our study. As depicted in Fig. 10, we performed various data transformations including: (a) trend-adjusting the electricity consumption data to allow for isolation of the effects of climate variability and removing the 'secular trends'; (b) spatiotemporal

Table 1
Summary of the final dataset.

Predictors	Response Variable	Data Sample Size
Electricity price Total precipitation in a month Mean dew point temperature Mean wind speed Maximum wind gust Per capita income of the state Unemployed population share	Electricity sales (consumption)	2495 observations pertaining to 312 months between 01/01/1990 and 31/12/2015, including all the 8 states considered in our analysis

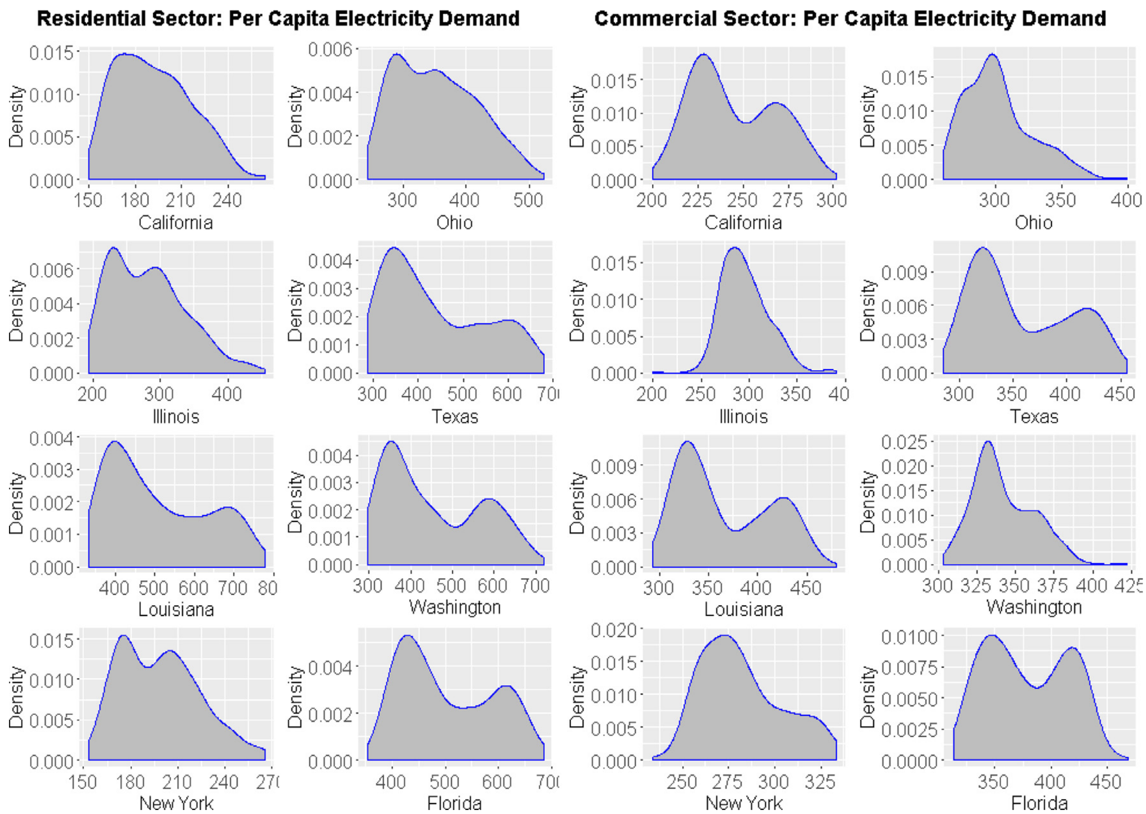

Fig. 3. Kernel distributions of per capita electricity demand (in KWh) in the individual states.

aggregation to aggregate daily data from all climate stations; and (c) adjusting the economic data based on the consumer price index (CPI) to remove the effect of inflation. As described earlier in the data section, the adjusted data for each of the states were aggregated using the year, month and the U.S. states as the common key variables to render the "Final Dataset".

We then leveraged supervised learning theory to develop sectoral energy-climate prediction models. Supervised learning involves estimating the unknown predictive function f that is able to generate predictions of the target variable of interest Y (in this case, sectoral per capita electricity demand) using a p dimensional vector of input variable X (i.e., climate and socio-economic variables in this study); where $Y = f(x) + e$. The objective of supervised learning is minimizing the loss function L which measures the deviation of observed (out-of-sample records) from the predicted values of Y. To conduct the proposed predictive modeling, the "Final Dataset" was split into the "Model Development Set" and "Model Validation Set". The *model development set*, which was used to train and test the prediction model, included all the observations from Jan 1990-Dec 2011, i.e., all the observations in the first 22 years. The *model validation set*, on the other hand, was used to evaluate and validate our models in terms of their future predictive performance. Each of the model development, including training and test sets', and model validation sets' observations were categorized into the two levels: (a) "High-intensity Consumption" (all observations with electricity consumption above the 3rd quartile) and (b) "Moderate-intensity Consumption" (all observations with electricity consumption below the 3rd quartile). The high and moderate consumption categories of training data were trained leveraging the Bayesian additive regression trees (BART) algorithm [13,14] for both the residential and commercial sectors. The composite predictive framework facilitated the identification of the key predictors of the high- and moderate- electricity consumption in the residential and

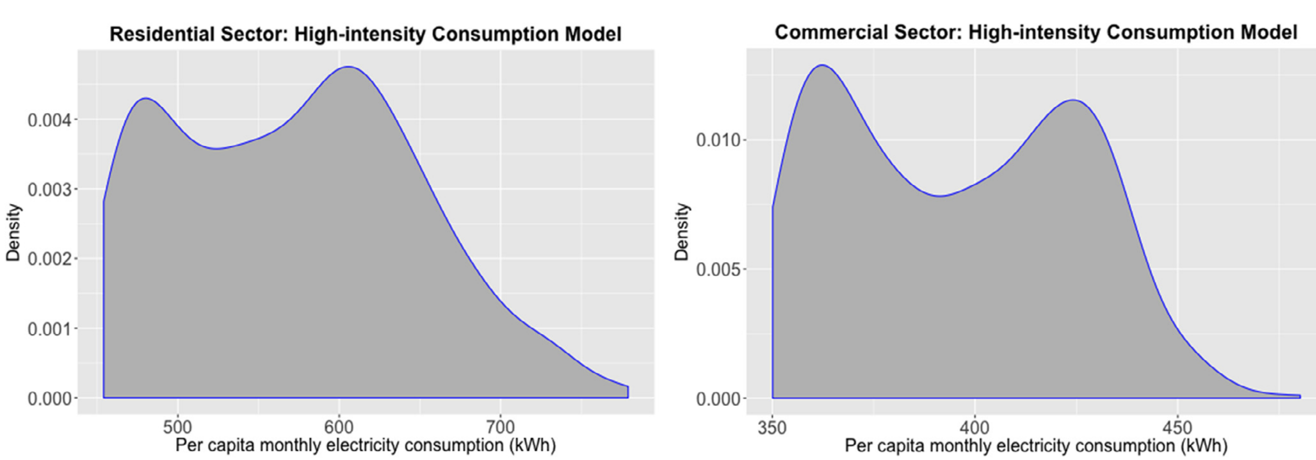

Fig. 4. Kernel density plots of the high-intensity end-use consumption in (a) residential and (b) commercial sectors.

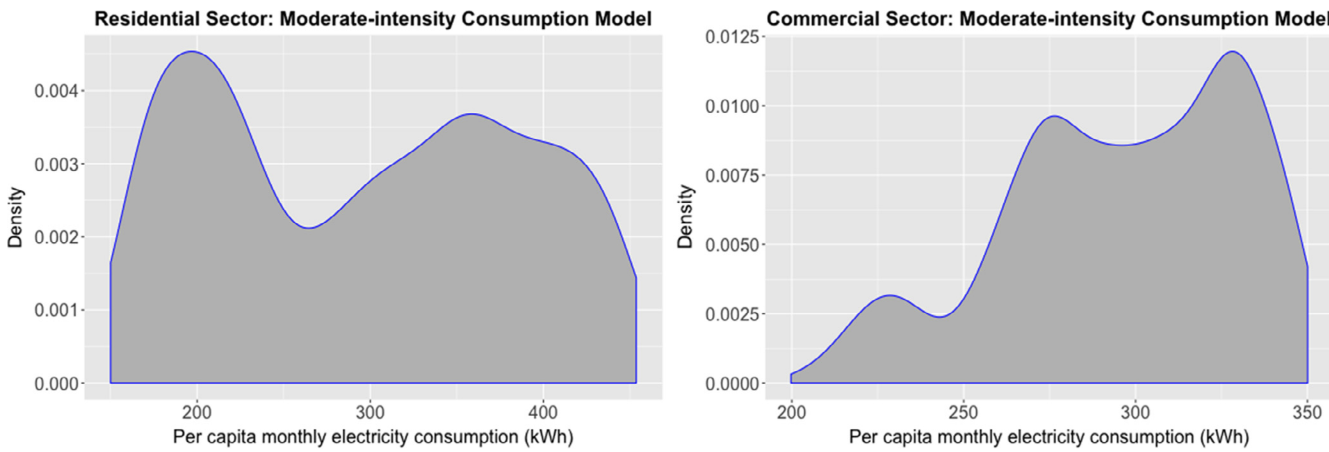

Fig. 5. Kernel density plots of the moderate-intensity end-use consumption in (a) residential and (b) commercial sectors.

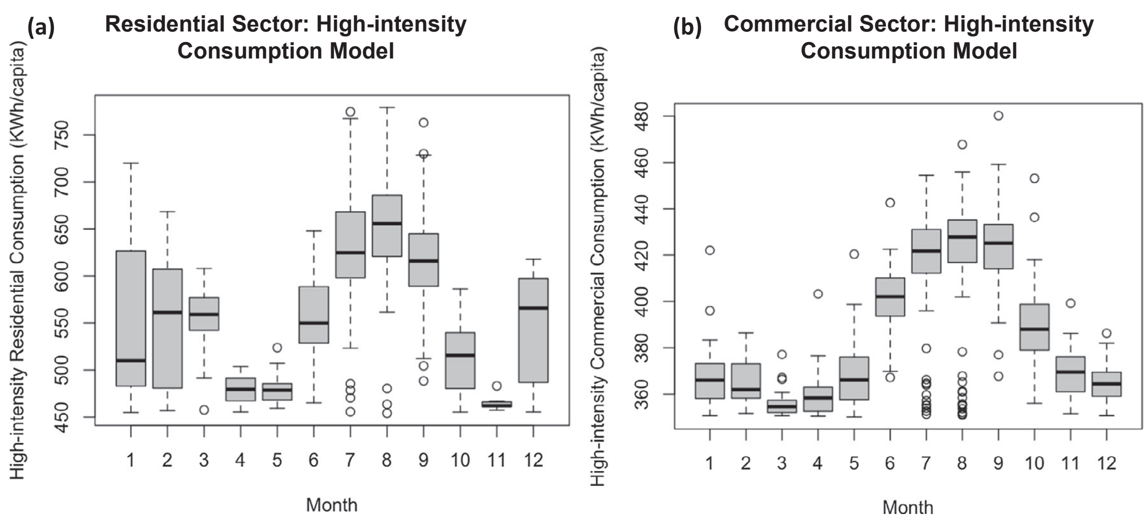

Fig. 6. Seasonal variation of high-intensity electricity consumption in the (a) residential and (b) commercial sectors.

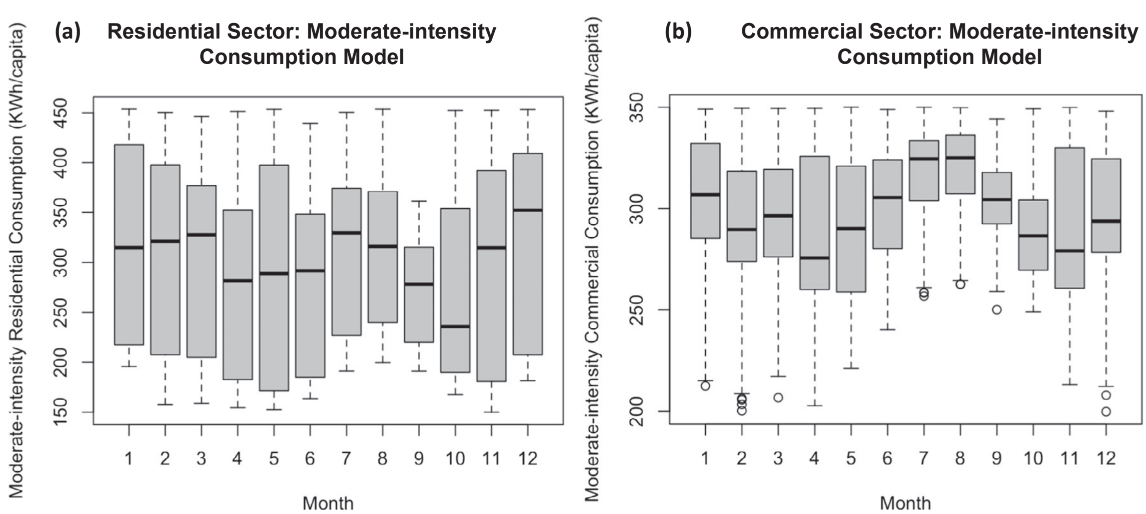

Fig. 7. Seasonal variation of moderate-intensity electricity consumption in the (a) residential and (b) commercial sectors.

commercial sectors for the top energy intensive states considered in this study.

4.1. Bayesian additive regression trees (BART)

Bayesian additive regression trees (BART) is a nonparametric, Bayesian tree-ensemble model. Mathematically, it can be represented as a sum-of-trees model, where the response function Y is approximated by aggregating the estimates from m 'shallow' decision trees as shown below [13,14,43]:

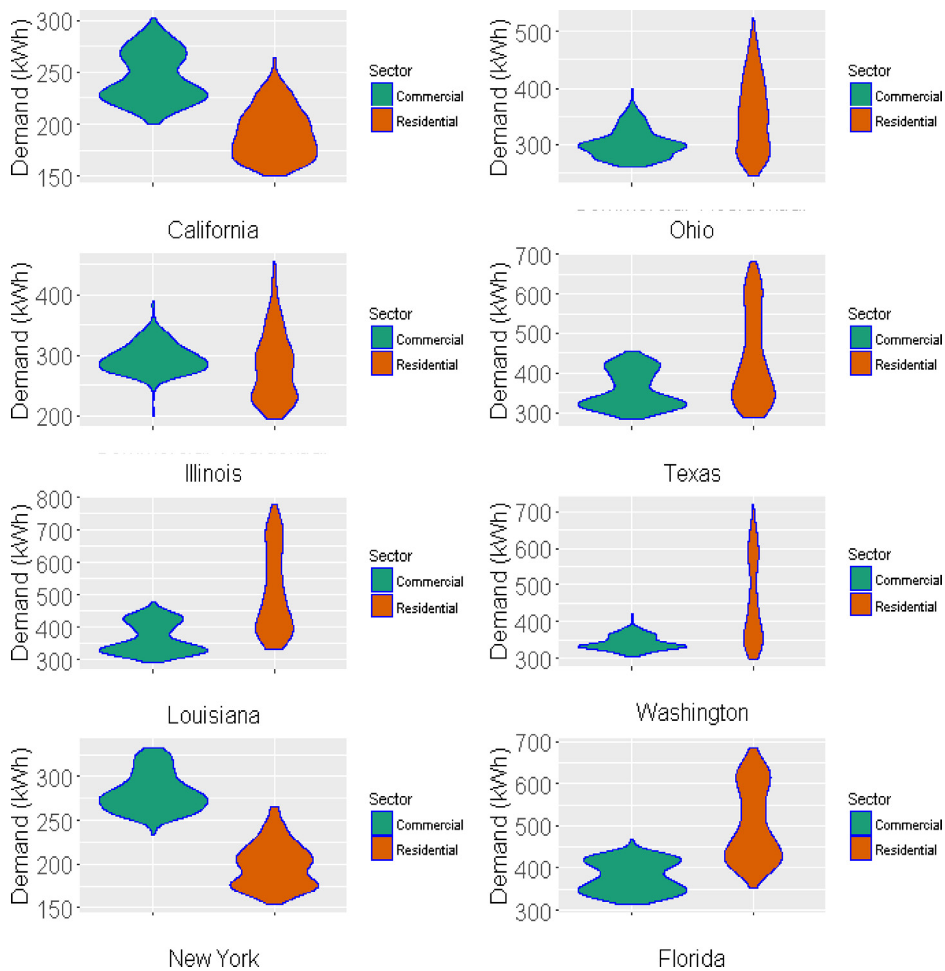

Fig. 8. Characteristics of individual states' per capita electricity demand in the commercial and residential sectors.

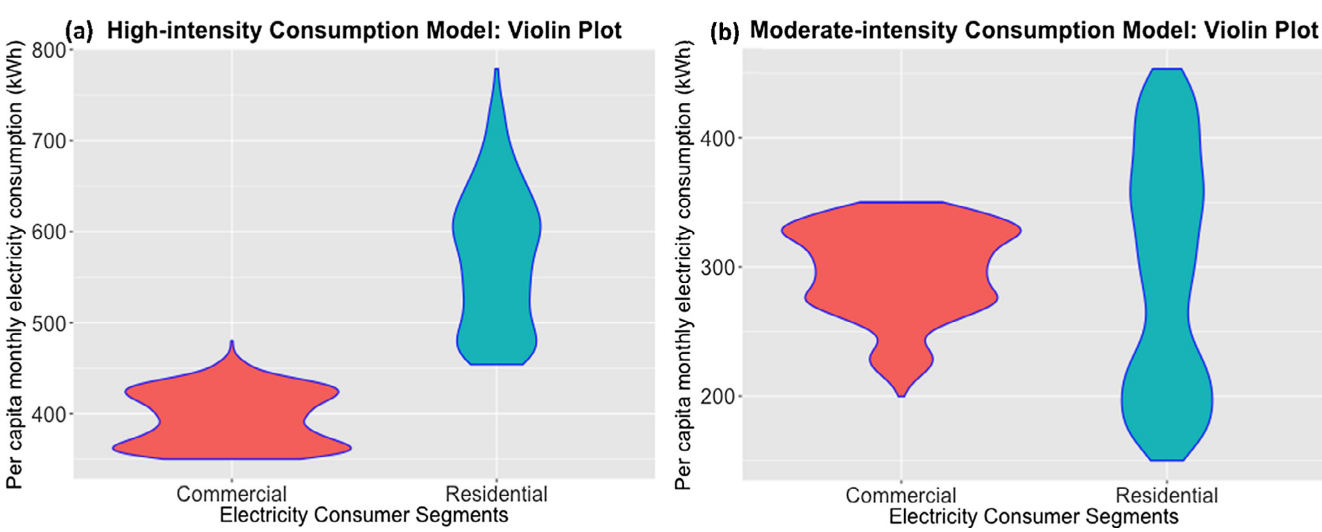

Fig. 9. Characteristics of combined eight states' (a) higher intensity and (b) moderate intensity consumption patterns in the commercial and residential sectors.

$$Y = \left(\sum_{j=1}^{m} g(\boldsymbol{X};\, T_j,\, M_j)\right) + \epsilon, \qquad \epsilon \sim N(0,\, \sigma^2)$$

In the above equation, $g(X;\, T,\, M)$ is the function which assigns the parameters M of the tree T to the predictor matrix X across all the m trees. The additive stochastic noise ϵ—referred to as irreducible error—reflects the dependence of the response variable on quantities other than the input features that are neither observed, nor measured. Regularization priors are used on both the tree structures and the conditional expectations at each terminal node. The regularization priors are leveraged to control model complexity and restrict the overwhelming influence of the large tree components. Regularization priors eliminate an individual tree's effect being unduly influential on the sum-of-trees model [13,14]. Combining the prior distributions with

Table 2
Descriptive statistics (i.e., mean, median, minimum (Min), maximum (Max), interquartile range (IQR) and standard deviation (Std. Dev)) of high- and moderate-intensity electricity consumption in the residential and commercial sectors during Jan 1990–Jan 2016.

Demand levels	Sectors	Per capita electricity consumption (in KWh) during Jan 1990–Jan 2016					
		Mean	Median	Min	Max	IQR	Std. dev
High	Residential	572.7	575.2	454.0	779.0	122.1	75.8
	Commercial	395.7	396.0	350.1	480.2	55.5	29.8
Moderate	Residential	296.5	301.4	150.0	453.6	163.4	89.9
	Commercial	296.3	300.1	199.7	350.1	53.0	34.8

tree-model likelihoods, yields a posterior distribution of the tree models. The sum-of-trees model is fitted using a Bayesian back-fitting MCMC algorithm [31].

4.1.1. Model inference

To facilitate model inferencing for the proposed non-parametric predictive model, we generated variable importance ranking and partial dependence plots (PDPs). The variable importance ranking is computed based on 'variable inclusion proportion', which indicates the fraction of times a given predictor was used in growing a regression tree. Unlike the parametric models, where inferences can be made based on their regression coefficients, partial dependence plots (PDPs) are used to make inferences in non-parametric models. In general, PDPs help in understanding the individual effects of the predictor variables x_j in a *ceteris paribus* condition (i.e., controlling for all the other predictors). Mathematically, the estimated PDP is given as [11]:

$$\widehat{f_j}(x_j) = 1\Big/n\sum_{i=1}^{n}\widehat{f_j}(x_j, x_{-j,i})$$

Here, $\widehat{f}$ denotes the statistical model; n denotes the number of observations in the training dataset; x_{-j} denotes all the variables except x_j. The estimated PDP of the predictor x_j provides the average value of the function $\widehat{f}$ when x_j varies over its marginal distribution keeping x_{-j} fixed.

4.1.2. Bias-variance tradeoff for assessing the model performance

The generalization performance of a predictive model depends on its capability to make good predictions on an independent test sample. Balancing the bias-variance trade-off is the key to minimized generalization error [44,45]. Cross validation techniques (e.g., k-fold, leave-one-out, etc.), bootstrap and validation set approaches are some examples of resampling techniques widely used for balancing bias and variance. In this paper, we leveraged k-fold cross validation technique in addition to the validation set approach to train, test and validate our models.

Cross validation is one of the most widely used resampling methods (as mentioned before) in balancing bias and variance. The method of k-fold cross validation is generally used to estimate predictive accuracy. k-fold cross-validation involves randomly dividing the data into k equally-sized subsets. In each iteration, the model is fitted to the subsets except the kth held-out sample, and the predictive accuracy is calculated based on the models' performance on the kth held-out subset. Training error is the average error over the training sample and is given as: $\bar{\mathrm{err}} = \frac{1}{N}\sum_i^n L(y_i, \hat{f}(x_i))$ [45]; the test error/prediction error is given as $\mathrm{Err}_\tau = E[L(y_i, \hat{f}(x_i))|\tau]$ [45]. Here, L is the loss function and τ refers to the specific training corresponding to which the test error is evaluated. The model with minimum test error is then assessed against the validation set to quantify its predictive performance. In this paper, we used a randomized holdout technique to estimate the models' predictive performance. More specifically, we implemented a 30-fold random holdout validation tests on the training data (Jan 1990-Dec 2011), where in each iteration, 20% of the data was randomly held out ("Test Set") and the model was trained with the remaining 80% data ("Training Set"), and then tested using the held-out sample, i.e., the *test set* (refer to Fig. 10). Since the cross validation is based on randomized sub-sectioning the data into 80–20% portions, "30 repetitions" is considered as a conservative measure to ensure that all the observations

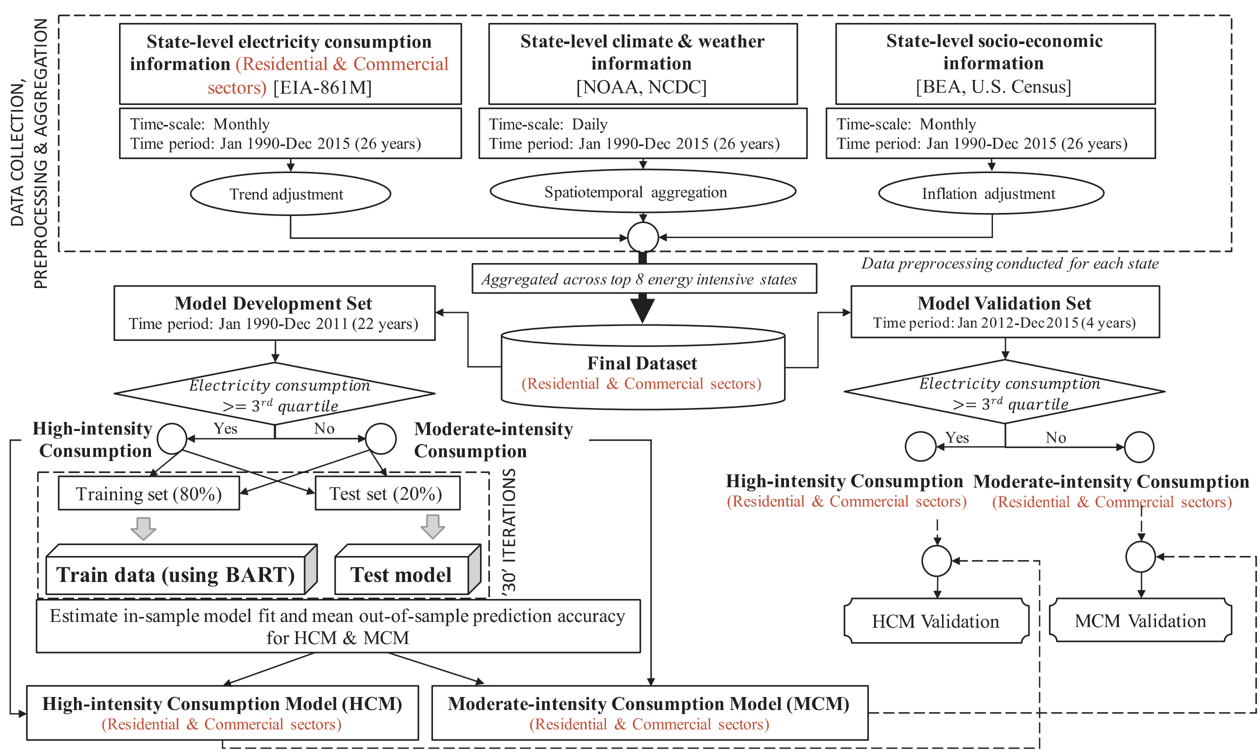

Fig. 10. The schematic of the proposed composite predictive modeling framework.

have been used at least once [45,46].

On the other hand, the *validation set approach* is a type of method that estimates a model error rate by holding out a subset of the data from the fitting process (creating a validation dataset). Once we developed our models and selected them based on the model's performance both in terms of in-sample model fit and out-of-sample prediction accuracy (leveraging the *model development set*), we used the *model validation set* (Jan 2012–Dec 2015) to validate the predictive performance of our models (refer to Fig. 10). Using the predictor set of the validation data, we predicted the future per capita electricity consumptions (KWh) in the residential and commercial sectors. Finally, we compared the predicted values (given by our models) and the actual observations during the period of Jan 2012–Dec 2015 for validation.

5. Results

In this section, we describe the performance of the prediction models of the per capita electricity demand in the residential and commercial sectors developed for the top eight energy intensive states of California, Florida, Illinois, Louisiana, New York, Ohio, Texas, and Washington. As discussed before, we developed two types of models—HCM (high-intensity consumption model) and MCM (moderate-intensity consumption model)—for the residential and commercial sectors. For each model, we assessed the in-sample goodness-of-fit and out-of-sample predictive accuracy and plotted the percentage coverage of the observed values within the credible and prediction intervals. The credible interval (aka Bayesian confidence interval) characterizes the posterior distribution of the predicted values [14], while the prediction interval characterizes the confidence levels related to the (future) observations that are yet to be observed.

5.1. Model performance

The performance of the developed prediction models was assessed based upon both in-sample and out-of-sample root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). The error rates reported in Tables 3 and 4 indicate the cross-validation errors. To benchmark the performance of our predictive models, we also provided information about the 'null' or 'mean-only' models. Comparison with the 'null model' reveals the extent to which a predictive model contributes to explaining the variance of the response, beyond the historical mean values [1–4]. Tables 3 and 4 reveal the significant predictive performance of both the high consumption model (HCM) and the moderate consumption model (MCM) in the residential and commercial sectors compared to the "mean-only" model.

From Table 3, it can be observed that the magnitude of the HCM's mean absolute error (MAE) related to the in-sample model fit, improves by 51.26% and 61.23% in the residential and commercial sectors respectively, compared to the mean-only model. On the other hand, the out-of-sample prediction errors (in-terms of MAE) of the HCM models are improved by 40.48% and 50.15% respectively in the residential and commercial sectors, compared to the mean-only model. The lower predictive performance (i.e., higher error rate) of the HCM model for the residential sector relative to the commercial sector can be attributed to the higher level of variation in the residential sector (Fig. 9). This is consistent with previous research that identified much higher levels of heterogeneity in the residential sector compared to the commercial sector [1,47].

Table 3
Statistical performance related to the high consumption models (HCM) in the residential and commercial sectors.

Sector	Model	R^2	In-sample model fit		Out-of-sample prediction accuracy	
			MAE	RMSE	MAE	RMSE
Residential	BART	0.71	32.19	41.50	38.94	50.32
	mean-only	-NA-	66.04	77.36	65.42	77.57
	% Improvement	-NA-	51.26	46.35	40.48	35.13
Commercial	BART	0.80	10.24	13.47	13.12	17.52
	mean-only	-NA-	26.41	29.82	26.32	29.72
	% Improvement	-NA-	61.23	54.83	50.15	41.05

Table 4
Statistical performance related to the moderate consumption models (MCM) in the residential and commercial sectors.

Sector	Model	R^2	In-sample model fit		Out-of-sample prediction accuracy	
			MAE	RMSE	MAE	RMSE
Residential	BART	0.95	16.20	20.83	19.93	25.95
	mean-only	-NA-	79.31	89.23	79.47	89.38
	% Improvement	-NA-	79.57	76.66	74.92	70.97
Commercial	BART	0.92	7.63	9.70	9.65	12.19
	mean-only	-NA-	28.96	34.70	28.87	34.61
	% Improvement	-NA-	73.65	72.05	66.57	64.78

Similarly, Table 4 shows the magnitude of the MCM's performance in the residential and commercial sectors. In the case of in-sample fit, the mean absolute error (MAE) of the developed models improves by 79.57% and 73.65%—in the residential and commercial sectors respectively—over the mean-only model. On the other hand, the out-of-sample prediction errors (in terms of mean absolute error) of the MCM models are improved by 74.92% and 66.57% respectively in the residential and commercial sectors. In contrast to the HCMs' performance in the residential and commercial sector, the MCM models perform differently in the two sectors. The lower percentage improvement (both in terms of goodness of fit and out-of-sample prediction accuracy) in the MCMs' performance over the mean-only model in the commercial sector indicates that moderate level of the commercial electricity demand has higher variations and associated uncertainties compared to the residential sector's electricity demand. Thus, moderate-levels of residential electricity demand can be better predicted compared to that of the commercial electricity demand. We reason that since moderate-intensity commercial demand is less sensitive to climate compared to the residential demand (refer to Fig. 7).

5.2. Model diagnostics

Figs. 11 and 12 depict the normal probability plots of the in-sample residuals for the HCM and MCM models respectively. Figs. 11a, b and 12b reveal that the residuals of the residential and commercial HCM models, as well as the residuals of the commercial MCM model follow a normal distribution with an expected value of zero. However, the normality assumption is slightly violated at the upper tails of the residential MCM (Fig. 12a). This minor violation of the assumption suggests that the electricity demand of the residential consumers with moderate demand intensity may depend on other (non-climatic) factors—such as human behavior and spatiotemporal heterogeneities—that were not accounted for in the presented model.

The model performance of the HCM and MCM in terms of the in-sample fits are given in Figs. 13–16—using credible and prediction intervals. In case of the HCM, we observe that the 95% credible intervals provide 63.71% and 70.7% coverage for all the observations in the residential (Fig. 13a) and commercial (Fig. 13b) sectors, respectively. This indicates that the commercial HCM has a better in-sample fit compared to that of the residential sector. From Fig. 14, we observe a similar trend for the commercial MCM, where 95% credible intervals provide 63.8% (Fig. 14a) and 67.91% (Fig. 14b) coverage for all the observations in the residential and commercial sectors respectively.

Fig. 15 shows that the 95% prediction interval of the HCM offers a

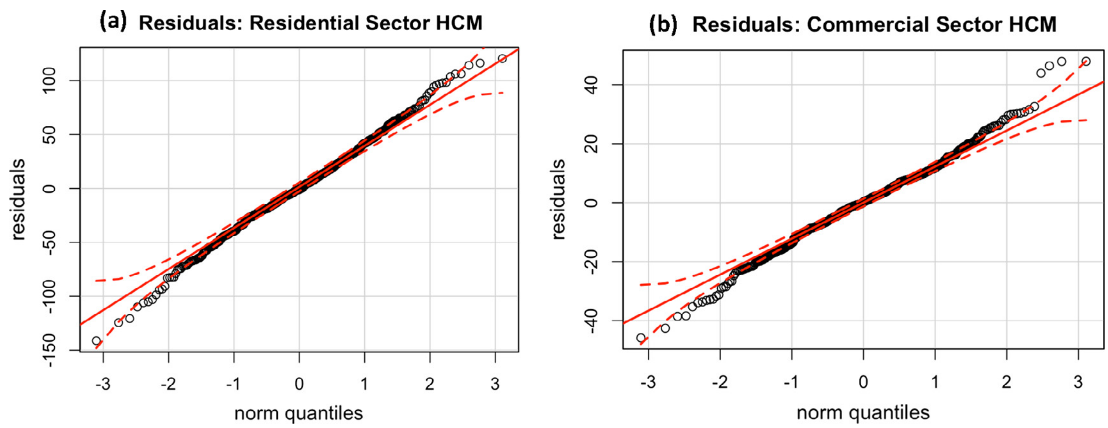

Fig. 11. Residual plots for the high-intensity consumption model (HCM): (a) residential and (b) commercial sectors.

97.16% (Fig. 15a) and 98.3% (Fig. 15b) coverage for the residential and commercial electricity demand observations respectively.

On the other hand, in case of the MCM (refer to Fig. 16), the 95% prediction interval provides a 97.54% (Fig. 16a) and 98.55% (Fig. 16b) coverage for the residential and commercial electricity demand observations respectively.

5.3. Model validation

As discussed before, we leveraged the developed predictive model (trained and tested using the data from Jan 1990–Dec 2011) to predict the per capita electricity consumption in the residential and commercial sectors for the period of Jan 2012–Dec 2015. We conducted this validation technique for both the high-intensity and moderate-intensity consumption models in both the sectors. We observed that the Pearson correlation coefficients between the predicted and the observed (actual) values of the high-intensity per capita electricity demand in the residential sector is $\rho = 0.72$ (Fig. 17a) and that in the commercial sector is $\rho = 0.88$ (Fig. 17b). This indicates that our proposed predictive HCM better captures the complex climate-electricity relationship in the commercial sector compared to the residential sector for the high intensity consumption levels (refer to Section 5.1, Table 3).

Figs. 18 and 19 show the comparison between the seasonal trends in the predicted and the actual values of the high-intensity per capita electricity demand in the residential and commercial sectors respectively, during the years 2012–2015. The predicted values closely follow the trend of the actual values, which reveals the high predictive power of the developed models in making predictions of the future end-use demands under climate variability. It is observed that out of the eight states considered in this study, only five states—Florida, Louisiana, Ohio, Texas and Washington—are recorded for high-intensity per capita electricity consumption in the residential sector (Fig. 18). However, in case of the commercial sector, we observe that, in addition to the above mentioned five states, the state of Illinois experiences high-intensity per capita electricity demand. It is noteworthy that neither the residential nor the commercial sector in the states of California and New York experience high-intensity per capita electricity demand during the period of 2012–2015. This can be attributed to the higher investments in energy conservation and efficiency in the states of California and New York [48,49].

In case of the moderate intensity demand models (residential MCM and commercial MCM), we observe that the Pearson correlation coefficients between the predicted and the observed values of the per capita electricity demand in the residential sector is $\rho = 0.96$ (Fig. 20a) and that in the commercial sector is $\rho = 0.94$ (Fig. 20b). This indicates that our proposed MCM model adequately captures the complex climate-demand relationship and can project the demand into the future. In other words, we can infer that moderate-intensity demand can be better predicted than the high-intensity electricity demand. This is attributable to the high-intensity demand being a function of unobserved, non-climatic heterogeneities such as variations in occupant behavior and building characteristics, among other factors [47]. In addition, Fig. 20a and Fig. 20b also indicate that the residential sector's moderate intensity demand can be better predicted than that of the commercial sector (also refer to Section 5.1, Table 4 discussed before).

Figs. 21 and 22 show how the predicted values closely follow the trend of the actual observations for moderate-intensity consumption in the residential and commercial sector. Unlike the high-intensity consumption trends, we observe that all the eight states—i.e., California, Florida, Illinois, Louisiana, New York, Ohio, Texas, and

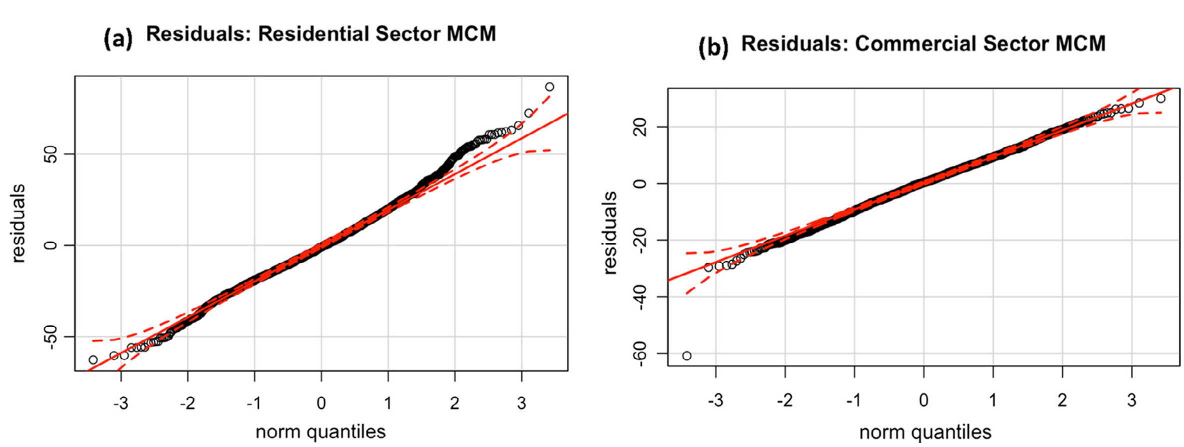

Fig. 12. Residual plots for the moderate consumption model (MCM): (a) residential and (b) commercial sectors.

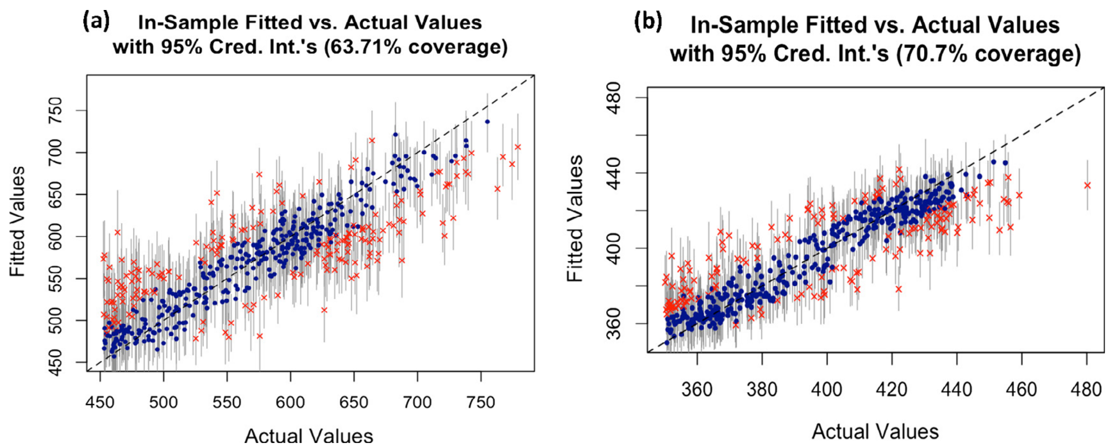

Fig. 13. In-Sample fit of the HCM models vs. actual values using 95% credible intervals: (a) residential and (b) commercial sectors.

Washington—are present in the moderate-intensity per capita electricity demand dataset during the period of 2012–2015 (validation period). It is important to note here that the states of California and New York are consistently at the bottom level of the seasonal residential and commercial sectors' consumption curves during the years 2012–2015. As discussed before, this lower intensity of per-capita electricity demand is mostly attributed to the higher energy efficiency standards and implementation of several sponsored programs for energy conservation and efficiency in the states of California and New York [48–50]. We also observe that the moderate-intensity consumption has a higher variation in the non-summer months; while during the warmer summer months all the represented states cluster more tightly close to higher values of demand.

In summary, the significant predictive performance of the HCM and MCM models for the residential and commercial sectors—as depicted in Figs. 17–22—confirms our initial hypothesis that the Bayesian tree-ensemble algorithm can capture the complex and non-linear climate–-demand nexus across all regions and adequately predict the demands into the future. In this research, although we leveraged the Bayesian predictive model for medium-term projections, the developed model can be used to make long-term projections under IPCC climate scenarios.

5.4. Model inference

The variable importance plots for the HCM in the residential and commercial sectors are given in Fig. 23a and Fig. 23b respectively, while that for the MCM are provided in Fig. 24a and Fig. 24b. The variable rankings in the variable importance plots are estimated based on the inclusion proportion of the predictors in the ensemble-of-trees model. In other words, the inclusion proportion for any given predictor represents the proportion of times that a split using the predictor appears in the ensemble tree among all splitting variables [13,14]. The rank of the covariates depicts the degree of influence on the response variable.

It can be observed from the HCM's variable importance plots (refer to Fig. 23) that the mean dew point temperature has the greatest influence on the per capita electricity demand in both the residential and commercial sectors; followed by the monthly mean wind speed, peak wind gust in a month, and the total precipitation in a month. We also observe that the per capita income of the state is a more significant predictor of the commercial sector's per capita electricity demand compared to that of the residential sector. This makes intuitive sense because, besides climatic factors, commercial electricity consumption is dependent on many non-climatic factors such as the commercial building age and type, lease incentive terms and human capital [47].

Fig. 24a and Fig. 24b once again identify the *mean dew point temperature* as the most important predictor of the moderate per capita electricity consumption in both the residential and commercial sectors. Unlike the residential and commercial HCM models, non-climatic factors appear to be playing a bigger role in predicting the moderate-intensity residential and commercial demand levels. For instance, *electricity price* appears to be the second most important predictor of moderate per capita electricity demand in both the residential and commercial sectors. We also observed that in the residential sector, the per capita electricity demand is comparatively more influenced by another non-climatic predictor, the *per capita income of the state*, than the commercial sector. Moreover, the *monthly mean wind speed* is a more influential variable in predicting commercial electricity consumption compared to the residential sector. As initially hypothesized, except for

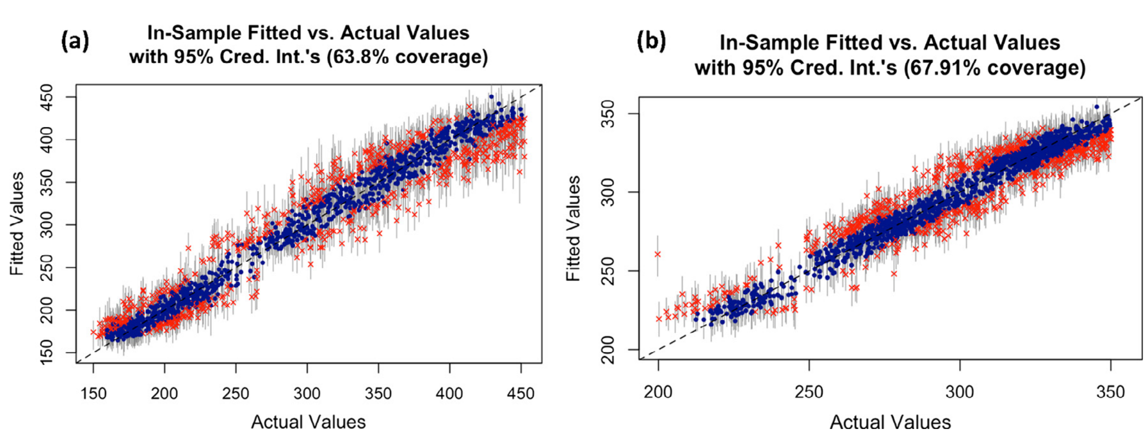

Fig. 14. In-Sample fit of the MCM models vs. actual values using 95% credible intervals: (a) residential and (b) commercial sectors.

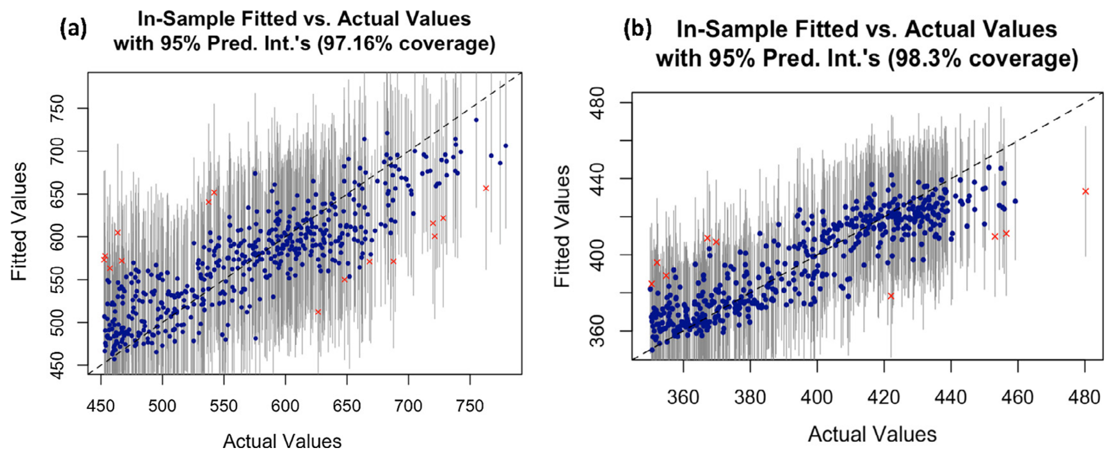

Fig. 15. In-Sample fit of the HCM models vs. actual values using 95% prediction intervals: (a) residential and (b) commercial sectors.

the *mean dew point temperature*, the influence of climate variables on end-use demand differs across the high-intensity and moderate-intensity consumption levels. The moderate-intensity consumption levels appear to be more sensitive to non-climatic factors compared to the high consumption levels, indicating the high-intensity consumption to be more seasonally variable than the moderate-intensity demand (refer to Figs. 6 and 7).

It is noteworthy that although we included the state-level *unemployment fraction* as one of the predictors in our analysis, it did not appear in the list of the top six important predictors (refer to Figs. 23 and 24). This indicates that the influence of *unemployment fraction* on per capita electricity consumption is suppressed by other economic predictors' influence, such as the *electricity price* and *per capita monthly income of the state*. Since our analysis was restricted to the top six important predictors, we have not plotted the marginal effect of *unemployment fraction* on the electricity consumption in this paper.

To understand the marginal influence of the important predictors, we plotted the partial dependencies between electricity demand and the most important predictors (for details on partial dependence plots refer to Section 4.1.1). We considered the top five important predictors—*mean dew point temperature*, *monthly mean wind speed*, *total precipitation* in a month, *electricity price* and *per capita income of the state*. The PD plots are plotted in black and 95% credible intervals plotted in blue; the points on both the PD plot and the 95% credible intervals are plotted at specific quantiles (i.e., 5-, 10-, …, 95-percentile values of the predictor). Lines plotted between the points approximate the PDP by linear interpolation [14].

5.4.1. Influence of mean dew point temperature

From the plots of the high- and moderate-intensity electricity consumption in the residential and commercial sectors (Figs. 25 and 26), it can be observed that there is a distinct non-linear relationship between the per capita electricity consumption and the mean dew point temperature. Fig. 25 shows that in both sectors, there is a steady increase in the per capita electricity consumption with increasing dew point temperature, especially at the higher levels of electricity consumption. Further, note that over the mean dew point temperature range of 0–25 °C, the average residential electricity demand increases by around 200 KWh per month, while the increase in the commercial sector over the same temperature range is around 80 kWh per month. Thus, the residential electricity consumption is more sensitive to the ambient temperature compared to the commercial sector, on average. This is intuitive because residential electricity consumption is more spatio-temporally heterogeneous and dependent upon the occupant behavior compared to that of the commercial sector [3].

However, the dependency between the moderate-intensity per capita electricity consumption and mean dew point temperature is distinctively different compared to that of the high-intensity consumption. Fig. 26 reveals that the monthly per capita consumption first decreases until a threshold value and then steadily increases. In the residential sector, the average threshold value (across all the states considered in this study) is found to be around 8–10 °C (equivalent to a surface temperature of 19–21 °C, with a 50% relative humidity) while that for the commercial sector is found to be around 3–4°C (equivalent to a surface temperature of 13–14 °C, with a 50% relative humidity). The threshold temperature refers to the comfortable temperature level when

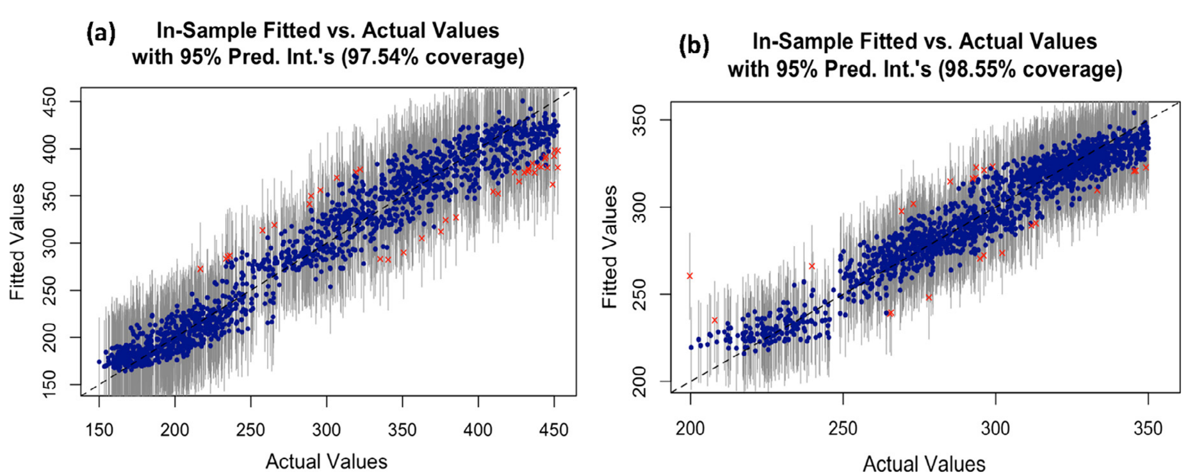

Fig. 16. In-Sample fit of the MCM models vs. actual values using 95% prediction intervals: (a) residential and (b) commercial sectors.

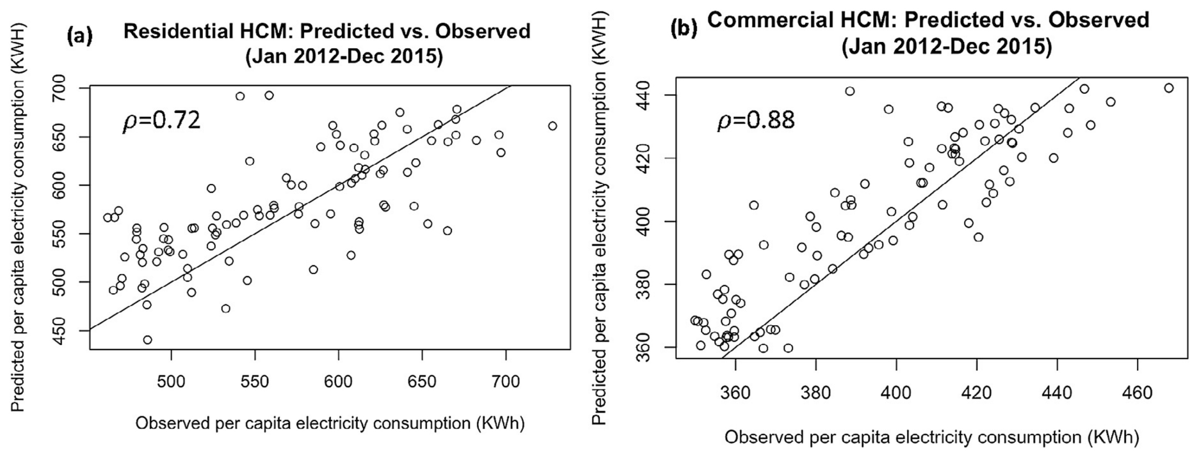

Fig. 17. Predicted vs. actual observations for the high-intensity consumption in the (a) residential and (b) commercial sectors during Jan 2012–Dec 2015.

the amount of electricity needed for space conditioning is the least. The shape of the curve in the residential sector—steep drop in electricity demand until the threshold temperature, and then steep increase—is consistent with our previous research that established the residential consumption as more heterogeneous and spatiotemporally varied compared to the commercial sector [3]. The drop in the electricity demand during colder months (dew point temperature ranging around −7 °C to 7 °C) can be attributed to higher heating requirement that is mostly supplied by natural gas, in most of the states. Thus, drop in electricity demand is mostly due to natural gas substitution in the colder months.

5.4.2. Influence of wind speed

Figs. 27 and 28 depict the marginal influence of wind speeds on the per capita electricity demands in the residential and commercial sectors. In the residential sector, we observed that with the increase in wind speeds from 2 m/s to 4m/s, the per capita electricity demand sharply decreased by about 60 KWh per month on average and did not show a significant variation on average following that (Fig. 27a). However, the larger confidence bounds (shaded grey area) around the higher wind speeds (i.e., wind speed ≥4 m/s) indicates that the sensitivity of demand to higher wind speeds is more variable across the different states. Although a similar pattern is observed in the relationship between the per capita electricity consumption and the monthly mean wind speed in the commercial sector, the magnitude of the decrease in demand is much less than that of the residential sector (Fig. 27b). In the commercial sector, the per capita demand on average decrease from around 410 KWh to 380 KWH, i.e., by 30 KWh, while that in the residential sector decrease from around 620 KWh to 560 KWh, i.e., by 60 KWh.

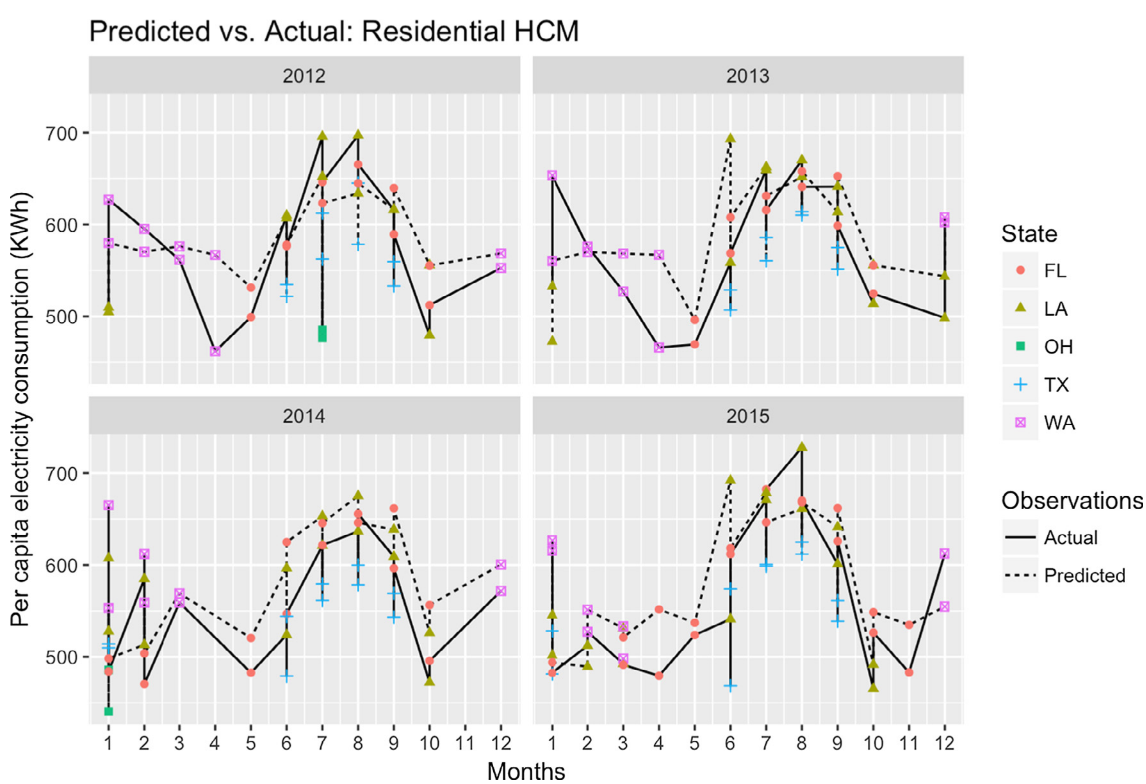

Fig. 18. Trends of actual observations vs. predicted values for the high-intensity consumption in the residential sector during Jan 2012–Dec 2015 (the vertical lines indicate the interpolation of the state-level per capita electricity consumptions, when more than one state is observed in a month). States experiencing high intensity consumption in the residential sector include: FL (Florida), LA (Louisiana), OH (Ohio), TX (Texas), and WA (Washington).

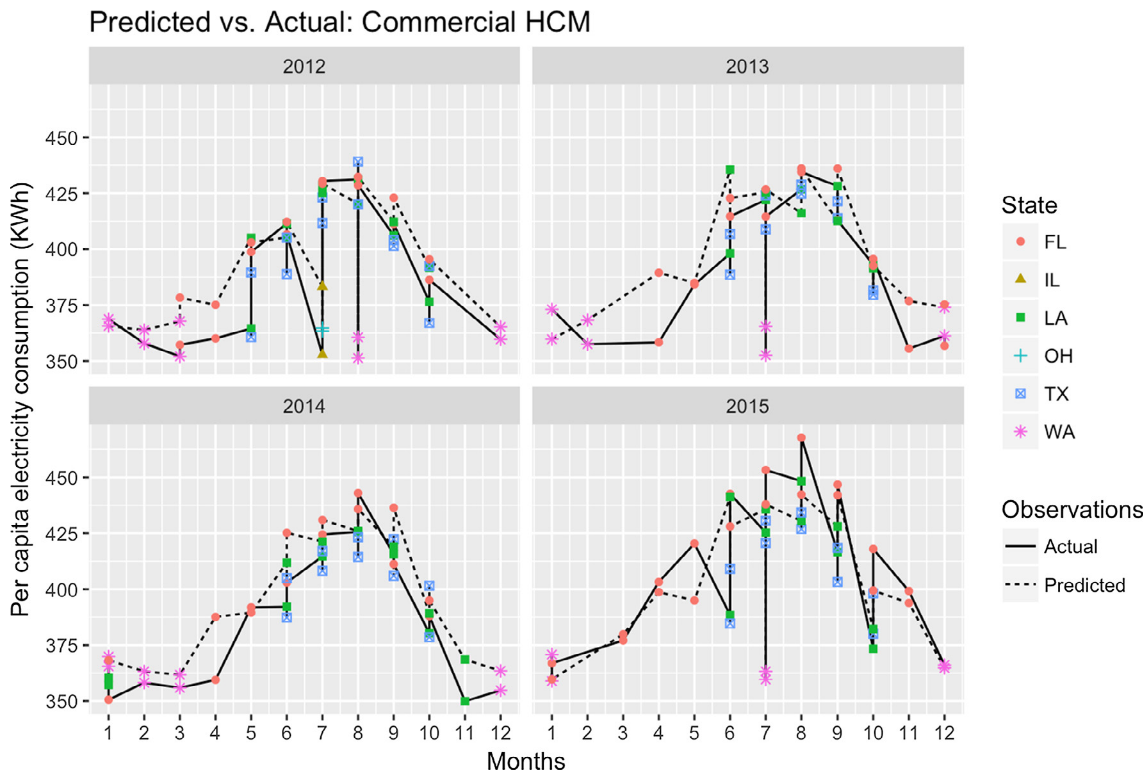

Fig. 19. Trends of actual observations vs. predicted values for the high-intensity consumption in the commercial sector during Jan 2012–Dec 2015 (the vertical lines the indicate interpolation of the state-level per capita electricity consumptions, when more than one state is observed in a month). States experiencing high intensity consumption in the commercial sector include: FL (Florida), IL (Illinois), LA (Louisiana), OH (Ohio), TX (Texas), and WA (Washington).

The influence of wind speed on the moderate-intensity demand follows a decreasing trend, similar to that of the high-intensity consumption, in both the residential (Fig. 28a) and commercial (Fig. 28b) sectors. However, in this case, the patterns of dependence (mean values as well as the uncertainty bounds) of the sectoral end-use electricity consumption to increasing wind speeds are quite different than that of the high-intensity electricity demand. Unlike the high-intensity sectoral demand, the moderate-intensity demand is much less variable with respect to increasing wind speeds. Over a range of 2–6m/s wind speed, the mean per capita moderate-intensity electricity demand decreases by about 15 KWh and 7 KWh in the residential and commercial sectors respectively (as is observed in Fig. 28).

In summary, a decreasing trend in the per capita high- and moderate- intensity of demand is observed with increasing wind speeds in both the residential and commercial sectors. This is intuitive as increasing wind speeds blow away the heat from the residential and commercial buildings, leading to a decreasing need for spatial cooling during warmer months.

5.4.3. Influence of total monthly precipitation

Figs. 29 and 30 show the variations of the residential and commercial sectors' per capita electricity demand as it relates to the total monthly precipitation. As observed from Fig. 29, with increasing levels of total precipitation, the per capita high-intensity electricity consumption decreases in both sectors. As the total monthly precipitation increases over a range of 10–240 mm, the median electricity demand decreases by about 30 KWh in the residential sector (Fig. 29a), whereas in the commercial sector it decreases by approximately 10 KWh (Fig. 29b). We hypothesized that since precipitation has a moderating effect on temperatures, with increased precipitation levels, the need for

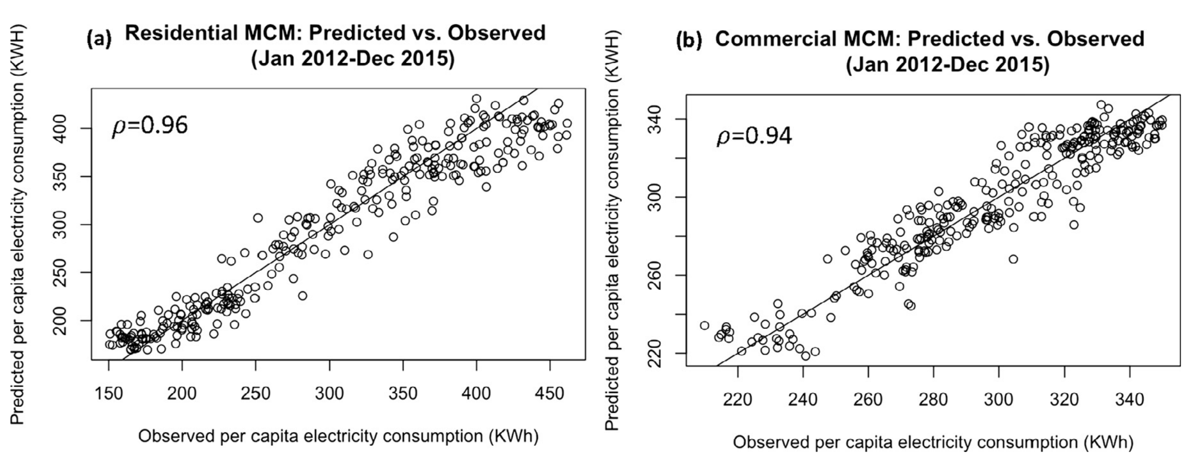

Fig. 20. Predicted vs. actual observations for the moderate-intensity consumption in the (a) residential and (b) commercial sectors during Jan 2012–Dec 2015.

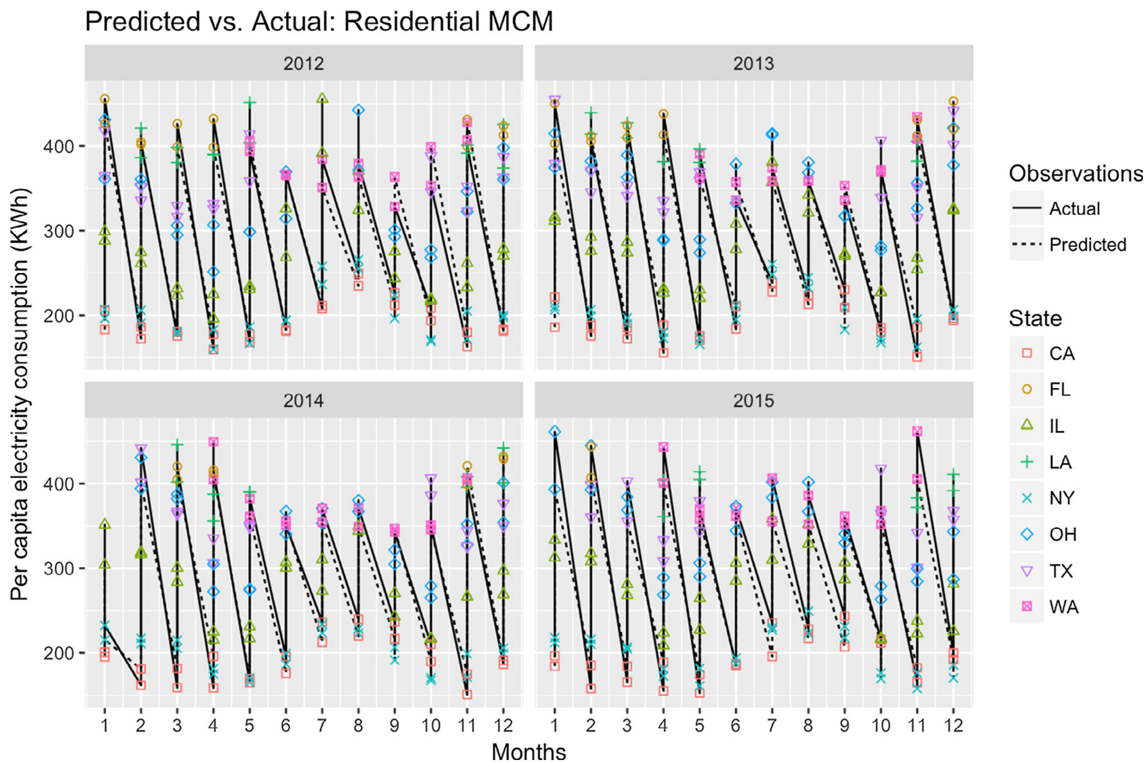

Fig. 21. Trends of actual observations vs. predicted values for the moderate-intensity consumption in the residential sector during Jan 2012–Dec 2015 (the vertical lines indicate interpolation of the state-level per capita electricity consumptions, when more than one state is observed in a month). States experiencing moderate intensity consumption in the residential sector include: CA (California), FL (Florida), IL (Illinois), LA (Louisiana), NY (New York), OH (Ohio), TX (Texas), and WA (Washington).

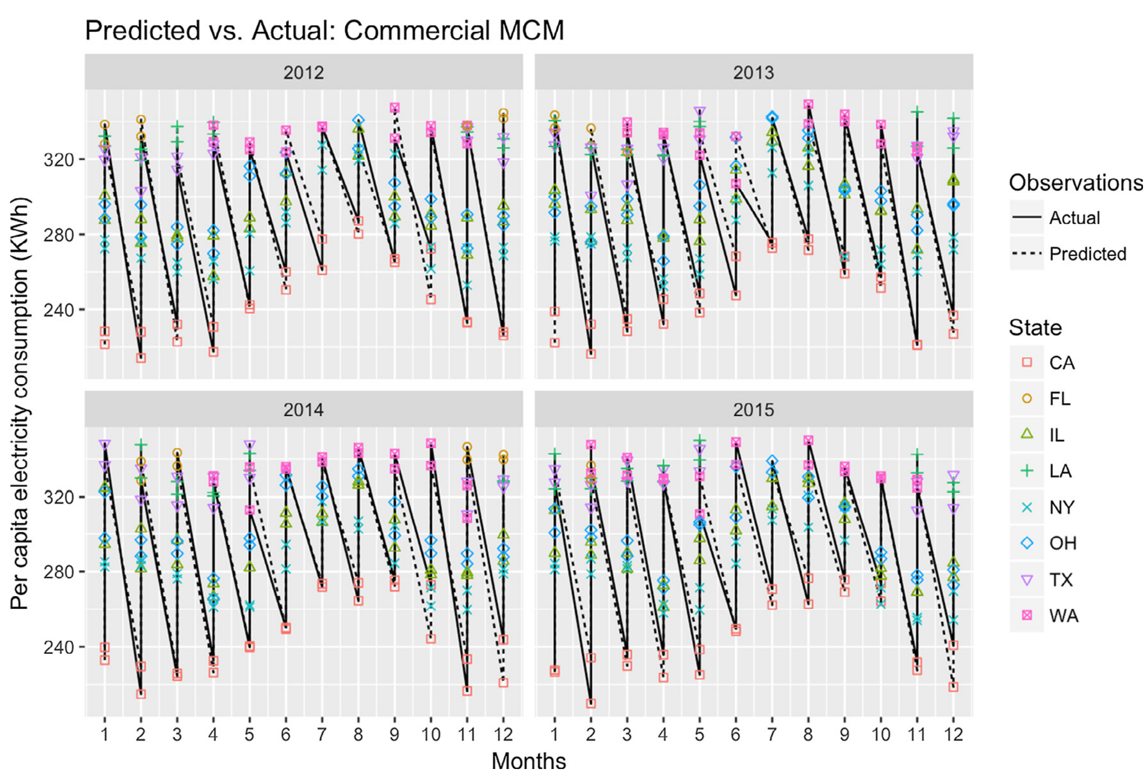

Fig. 22. Trends of actual observations vs. predicted values for the moderate-intensity consumption in the commercial sector during Jan 2012–Dec 2015 (the vertical lines indicate the interpolation of the state-level per capita electricity consumptions, when more than one state is observed in a month). States experiencing moderate intensity consumption in the commercial sector include: CA (California), FL (Florida), IL (Illinois), LA (Louisiana), NY (New York), OH (Ohio), TX (Texas), and WA (Washington).

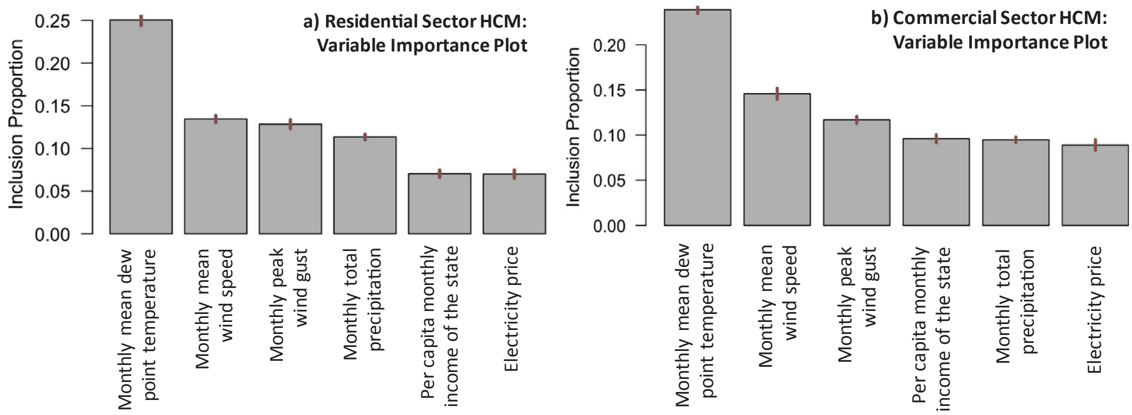

Fig. 23. Predictor importance ranking for the HCM: (a) residential and (b) commercial sectors.

spatial conditioning decreased leading to a reduced consumption of electricity.

From Fig. 30a, we observe that the variation in the moderate-level of per capita monthly electricity consumption in the residential sector is irregular and does not show any specific trends, unlike the high consumption model. We hypothesize that the irregular nature of the moderate-intensity residential consumption is due to the following reasons: (a) variation in geographical locations of the states, and (b) the difference in consumption patterns in these states across the different seasons. For example, states such as Louisiana, Florida, and Texas are characterized by hot and humid summer months and temperate winter months. Higher levels of precipitation in the winter months are associated with a cooling effect that might lead to infrequent increases in the electricity demand for heating requirement on average (indicated by "spikes" in Fig. 30a). On the other hand, higher precipitation levels in the warm and humid months bring down the temperature to a comfortable threshold, leading to a decrease in the electricity demand. In the commercial sector, the moderate electricity consumption does not show much variations with respect to precipitation levels, suggesting the residential sector to be more sensitive to seasonal climate variability. Thus, we can conclude that both the residential and commercial sectors' end-use electricity demands are insensitive to the different levels in precipitation.

5.4.4. *Influence of electricity price*

Fig. 31 reveal that the high-intensity electricity consumption in the residential (Fig. 31a) and commercial (Fig. 31b) sectors are price-insensitive, as variations in price do not change the consumption levels significantly. On the other hand, in the case of moderate-intensity electricity consumption (Fig. 32), we observe that both the residential and commercial sectors are sensitive to the variations in electricity price. In the residential sector (Fig. 32a), we observe that the mean per capita moderate-intensity electricity demand decreases from 320 KWh to 280 KWh, when the electricity price increases from 3.5 cents/KWh to 9 cents/KWh on average. Similarly, in the commercial sector (Fig. 32b), the mean per capita electricity demand decreases from around 310 KWh to 280 KWh, when the electricity price increases from 3 cents/KWh to 8 cents/KWh on average.

5.4.5. *Influence of per capita income of the state*

Figs. 33 and 34 show the variation of per capita high- and moderate-intensity electricity consumption respectively in the residential and commercial sectors. We observe a decreasing trend in the per capita high-intensity electricity demand in both the residential (refer to Fig. 33a) and commercial (refer to Fig. 33b) sectors, with the increasing per capita income of the states on average. We hypothesize that the higher purchasing power of a given state—as reflected by the higher state income—could suggest higher investments in more energy efficient electrical appliances, and thus the decreasing trend in demand.

From Fig. 34, we observe a lower sensitivity of the moderate-intensity consumption to the monthly state income in both the residential (Fig. 34a) and commercial (Fig. 34b) sectors, relative to the high-intensity consumption. In the commercial sector (Fig. 34b), we observe the following trend—with increase in per capita income, the per capita electricity consumption decreases in the beginning, until a threshold is reached and then it is followed by a slight increasing trend. However, the mean per capita electricity demand ranges between 295 KWh and 300 KWh, over the entire range of per capita income from 16,000 USD to 26,000 USD. This variation in per capita electricity demand being small, we can conclude that the per capita moderate-intensity

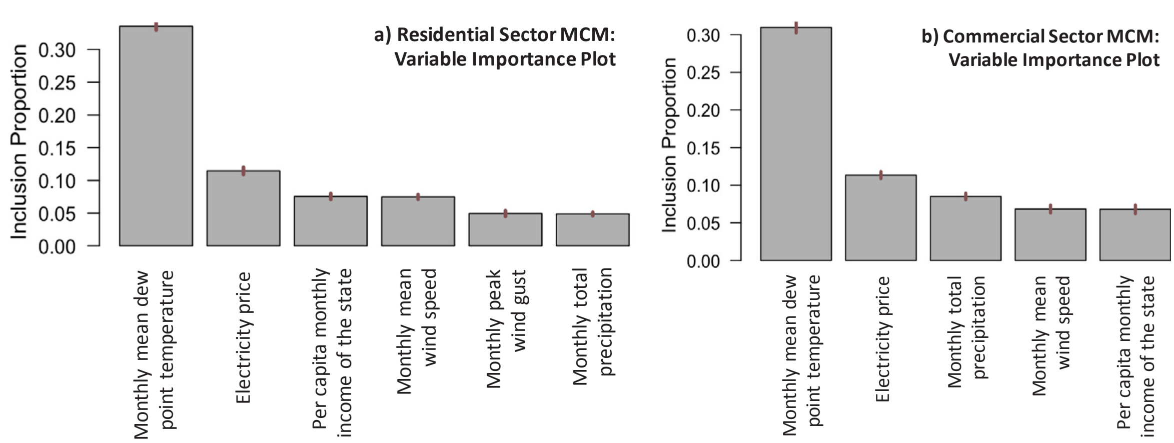

Fig. 24. Predictor importance ranking for the MCM: (a) residential and (b) commercial sectors.

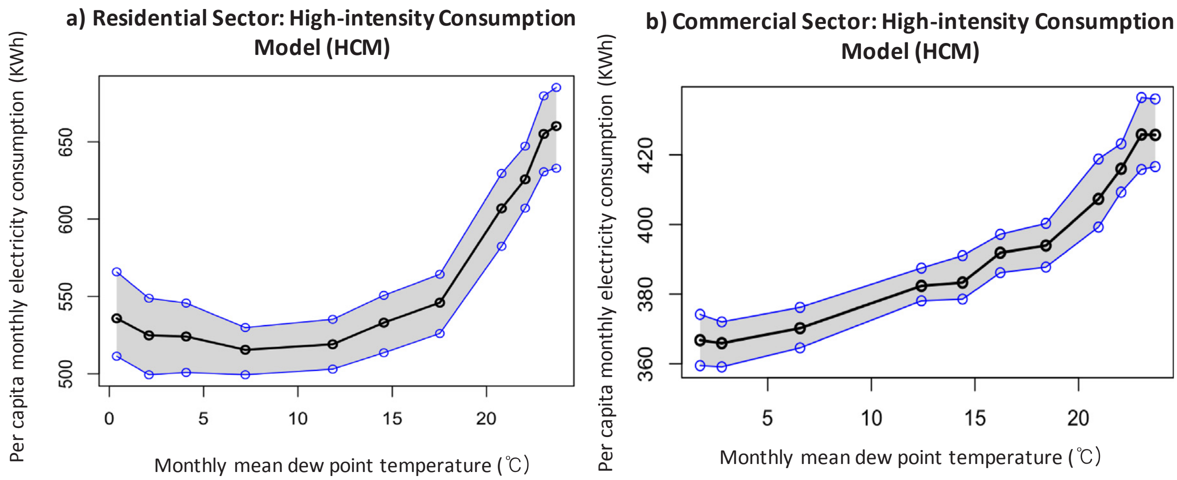

Fig. 25. Influence of mean dew point temperature on high levels of per capita electricity consumption in (a) residential and (b) commercial sectors.

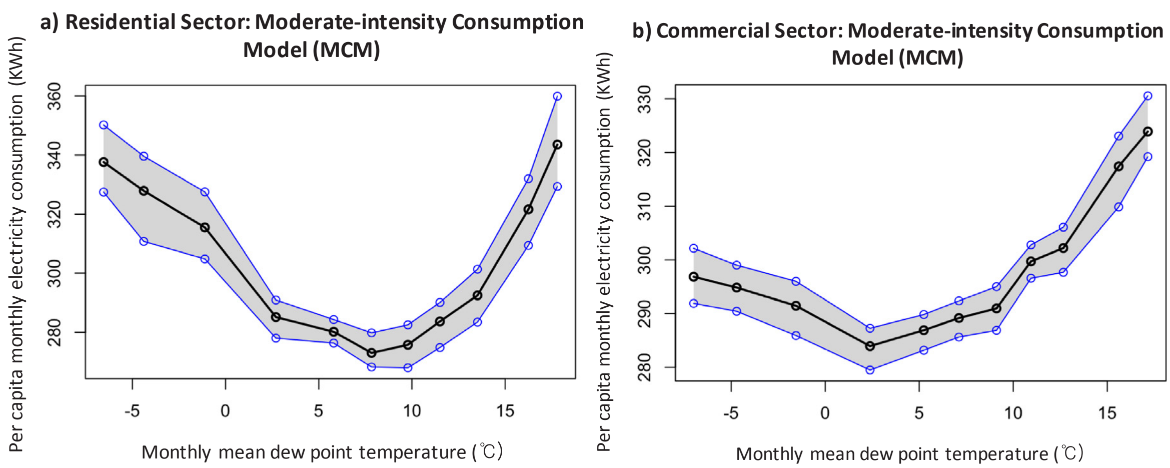

Fig. 26. Influence of mean dew point temperature on moderate levels of per capita electricity consumption in (a) residential and (b) commercial sectors.

electricity demand is relatively insensitive to per capita income variation, as compared to the high-intensity electricity demand.

6. Discussion and conclusion

Analyzing and quantifying the relationship between electricity demand and climate variability is of utmost importance. Mischaracterizing the climate-demand nexus could lead to incorrect estimation of capacity margins, leading to over- or under-investments in capacity expansion plans, which could manifest as inefficient planning or insufficient supplies. In this paper, we presented a methodology for developing rigorously validated prediction models of "electricity

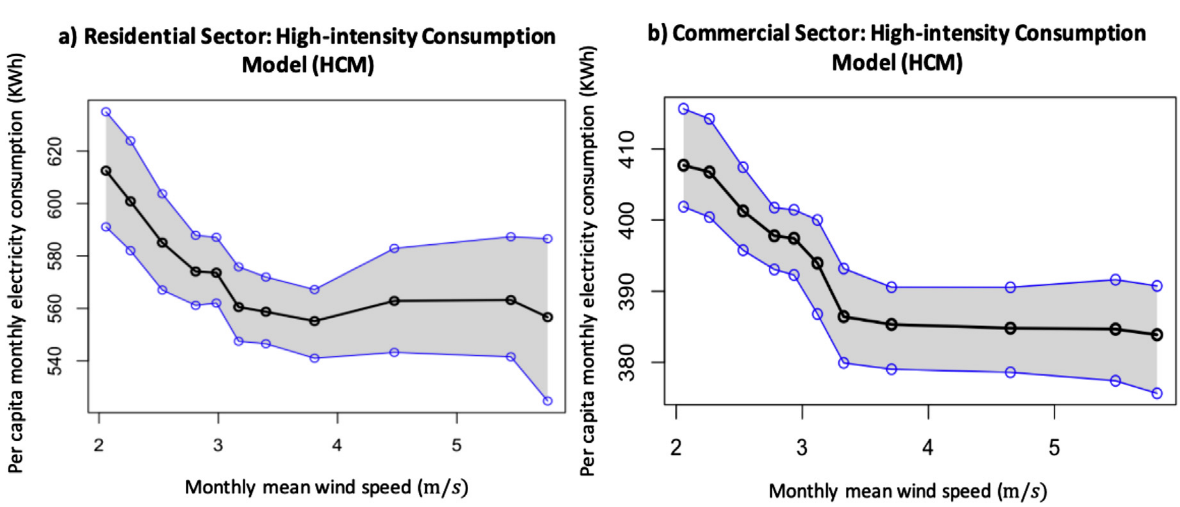

Fig. 27. Influence of wind speed on high-intensity levels of per capita electricity consumption in the (a) residential and (b) commercial sectors.

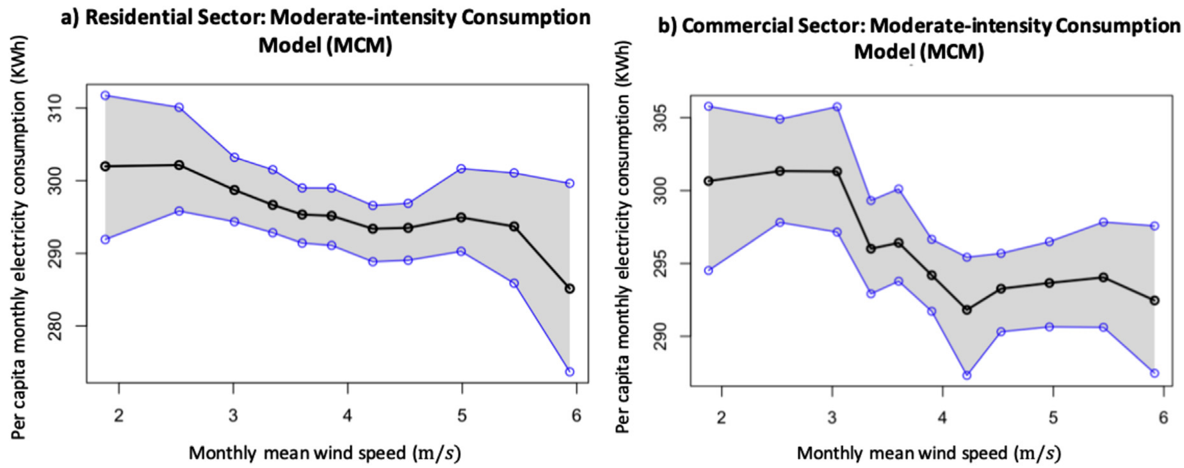

Fig. 28. Influence of wind speed on moderate-intensity levels of per capita electricity consumption in the (a) residential and (b) commercial sectors.

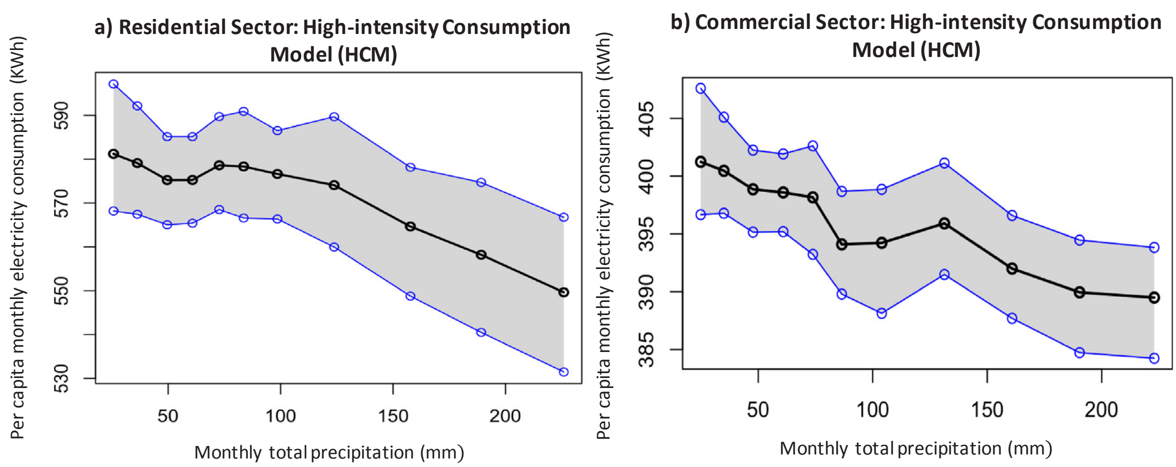

Fig. 29. Influence of total monthly precipitation on high-intensity levels of per capita electricity consumption in the (a) residential and (b) commercial sectors.

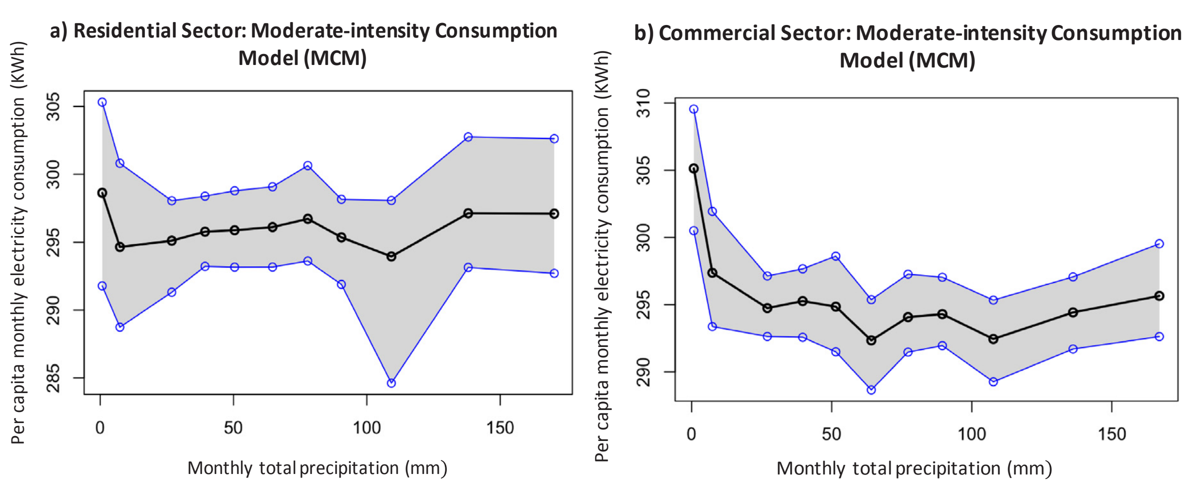

Fig. 30. Influence of total monthly precipitation on moderate-intensity levels of per capita electricity consumption in the (a) residential and (b) commercial sectors.

demand—climate nexus" for high- and moderate-intensity electricity consumption in the residential and commercial sectors. We considered the top eight energy intensive states in the U.S. viz., California, Florida, Illinois, Louisiana, New York, Ohio, Texas, and Washington. Unlike previous approaches of developing separate models for different geographical locations [5,6], we presented a novel composite Bayesian predictive framework to simultaneously characterize the demand–climate nexus in multiple states in the U.S. While we developed models for the eight most energy intensive states, our proposed methodology can be easily extended to other states/regions that might be of interest.

Unlike the widely used generalized linear models, our proposed models were based on a flexible, non-parametric Bayesian tree-ensembles approach that was effective in capturing the non-linear and complex relationship between electricity demand and climate variability. We considered mean dew point temperature instead of the conventionally used heating and cooling degree day variables (HTDD and

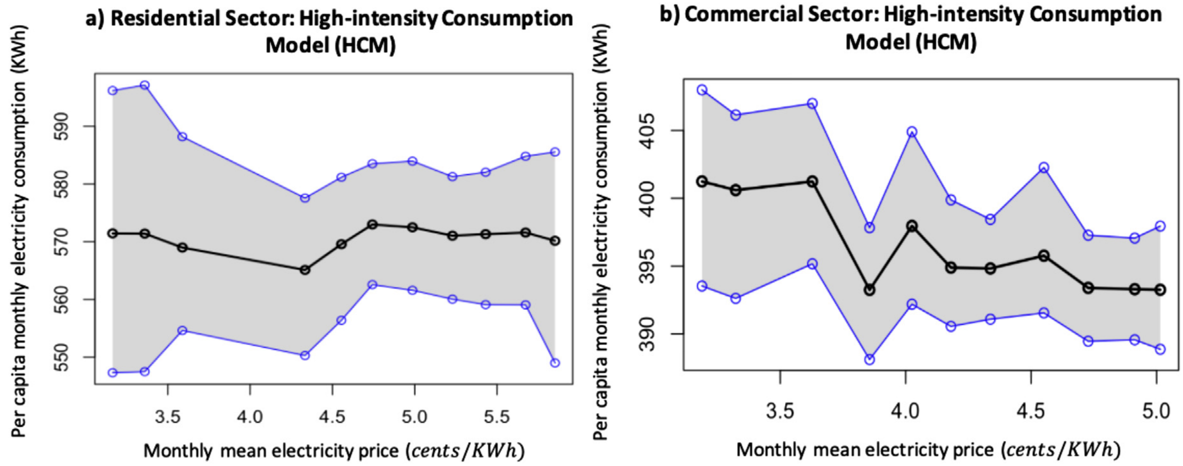

Fig. 31. Influence of electricity price on high-intensity levels of per capita electricity consumption in the (a) residential and (b) commercial sectors.

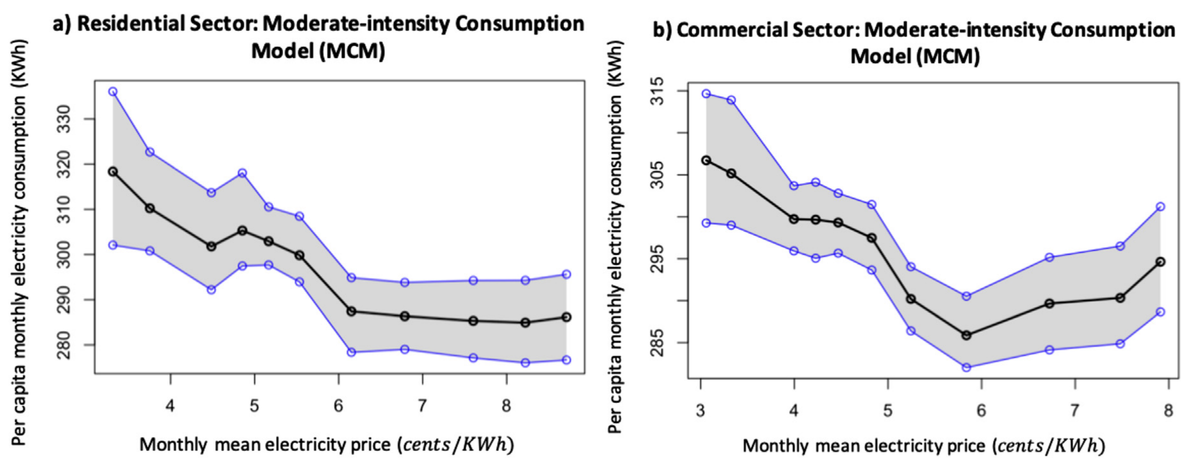

Fig. 32. Influence of electricity price on moderate-intensity levels of per capita electricity consumption in the (a) residential and (b) commercial sectors.

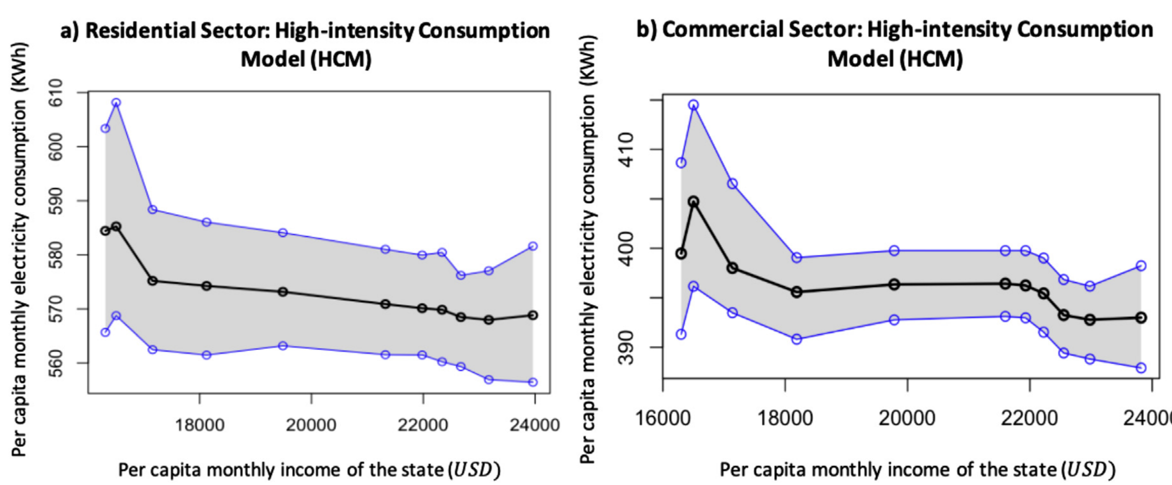

Fig. 33. Influence of per capita income of the state on high-intensity levels of per capita electricity consumption in the (a) residential and (b) commercial sectors.

CLDD). This is because, our previous research established mean dew point temperature as a more effective predictor of the climate-sensitive portion of electricity demand compared to HTDD or CLDD [3]. This study helped establish the mean dew point temperature as the most important predictor of both residential and commercial electricity demand irrespective of the intensity of the consumption levels. Mean dew point temperature was found to have a monotonically direct relationship with the high levels of electricity consumption in both sectors. However, for moderate consumption levels, the mean dew point temperature–per capita electricity demand partial dependency plots have a (non-symmetrical) parabolic shape; where the demand first decreases until a threshold (about 8–10 °C in residential sector or 3–4°C in the commercial sector) followed by a uniformly increasing trend.

We identified the higher-intensity consumption levels of electricity to be more sensitive to climate variability compared to the moderate consumption levels in both the residential and commercial sectors. In

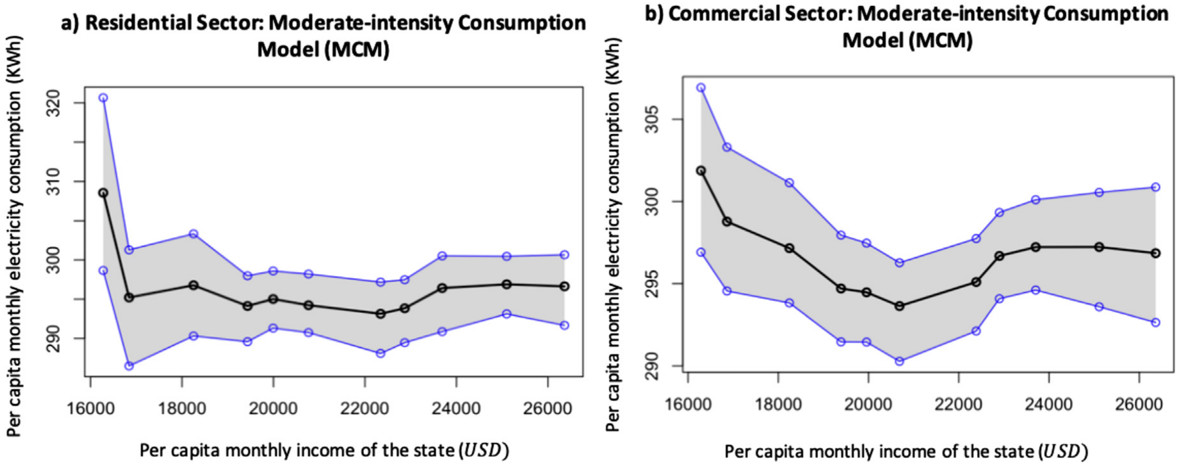

Fig. 34. Influence of per capita income of the state on moderate-intensity levels of per capita electricity consumption in the (a) residential and (b) commercial sectors.

addition to the mean dew point temperature, we found that wind speeds, wind gusts, and precipitation levels also influence the per capita electricity demand. We found that with increasing wind speeds, both the high- and moderate-intensity electricity demands show a decreasing trend in the residential and commercial sectors. We observed monthly precipitation levels to have a significant influence on the high-intensity electricity demands in both the residential and commercial demands; however, moderate-intensity electricity demand was found to be comparatively less sensitive to precipitation levels.

Electricity price was found to be the second most important predictor of the moderate-intensity electricity consumption, after the mean dew point temperature, in both the residential and commercial sectors. We observed that the high-intensity electricity consumption is relatively insensitive to the electricity price, whereas the moderate-intensity consumption has a decreasing trend with the increasing electricity price, in both the residential and commercial sectors. The per capita income of the states was also found to be an important predictor for the sectoral electricity demands; although the effect was not significant, we found that the electricity demand in both sectors shows a slight decreasing trend with the increasing per capita income of the states.

The major contribution of this paper is to illustrate the asymmetry in the climate sensitivity of electricity load. More specifically, our results demonstrated that the high-intensity end-use demands are more climate-sensitive, compared to the moderate-intensity end-use demands. Besides analyzing the asymmetry in climate sensitivity of load, our proposed composite Bayesian predictive model framework helps to identify the key climate predictors that significantly influence the high- and moderate- intensity demands. This framework can be also integrated with the General Circulation Model (GCM) climate change scenarios to project the future electricity demands in both the residential and commercial sectors. Moreover, although our predictive models are developed for projecting medium-term electricity demands, they can also be used for long-term demand projections under the IPCC climate scenarios. Our results also show that electricity sectors in the top eight energy-intensive states are vulnerable to extreme climate change and variations during the high demand periods (warmer summer months and intermediate seasons). If adequate capacity margins or resource allocations are not planned for in advance considering such climate effects, the risk of system inadequacy during the high-demand periods will increase [10], leading to frequent brownouts and blackouts. Our proposed framework will aid the utility stakeholders in informed decision making for capacity margins planning and resource allocations/reallocations, enhancing the resilience of the electricity sector in face of climate change and variations.

Acknowledgments

The authors would like to thank Mr. Jackson B. Bennett, and acknowledge the NSF grants #1728209 & #1826161.

References

[1] Nateghi R, Mukherjee S. A multi-paradigm framework to assess the impacts of climate change on end-use energy demand. PLoS ONE 2017;12:e0188033. https://doi.org/10.1371/journal.pone.0188033.
[2] Mukherjee S, Nateghi R. Estimating climate-demand nexus to support long-term adequacy planning in the energy sector. 2017 IEEE power energy soc. gen. meet. IEEE Xplore; 2018. p. 1–5.
[3] Mukherjee S, Nateghi R. Climate sensitivity of end-use electricity consumption in the built environment: an application to the state of Florida, United States. Energy 2017;128:688–700. https://doi.org/10.1016/j.energy.2017.04.034.
[4] Mukherjee S, Nateghi R, Hastak M. A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S. Reliab Eng Syst Saf 2018;175:283–305. https://doi.org/10.1016/j.ress.2018.03.015.
[5] Sailor DJ. Relating residential and commercial sector electricity loads to climate—evaluating state level sensitivities and vulnerabilities. Energy 2001;26:645–57. https://doi.org/10.1016/S0360-5442(01)00023-8.
[6] Sailor DJ, Muñoz JR. Sensitivity of electricity and natural gas consumption to climate in the U.S.A.—methodology and results for eight states. Energy 1997;22:987–98. https://doi.org/10.1016/S0360-5442(97)00034-0.
[7] Amato AD, Ruth M, Kirshen P, Horwitz J. Regional energy demand responses to climate change: methodology and application to the commonwealth of Massachusetts. Clim Change 2005;71:175–201. https://doi.org/10.1007/s10584-005-5931-2.
[8] Moral-Carcedo J, Pérez-García J. Temperature effects on firms' electricity demand: an analysis of sectorial differences in Spain. Appl Energy 2015;142:407–25. https://doi.org/10.1016/j.apenergy.2014.12.064.
[9] Raymond L, Gotham D, McClain W, Mukherjee S, Nateghi R, Preckel P, et al. Projected climate change impacts on Indiana's energy demand and supply. Clim Change 2018. [in press].
[10] Mukherjee S, Nateghi R. A data-driven approach to assessing supply inadequacy risks due to climate-induced shifts in electricity demand. Risk Anal 2018. https://doi.org/10.1111/risa.13192.
[11] Climate DOE. Change and the U.S. energy sector: regional vulnerabilities and resilience. Solutions 2015.
[12] DOE. Effects of climate change on Federal Hydropower; 2013.
[13] Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. Ann Appl Stat 2010;4(1):266–98. https://doi.org/10.1214/09-AOAS285.
[14] Kapelner A, Bleich J. bartmachine: Machine learning with bayesian additive regression trees. arXiv: Prepr arXiv13122171 2013.
[15] Yalcintas M, Aytun Ozturk U. An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database. Int J Energy Res 2007;31:412–21.
[16] Yamaguchi Y, Shimoda Y, Mizuno M. Proposal of a modeling approach considering urban form for evaluation of city level energy management. Energy Build 2007;39:580–92. https://doi.org/10.1016/j.enbuild.2006.09.011.
[17] Lokhandwala M, Nateghi R. Leveraging advanced predictive analytics to assess commercial cooling load in the U.S. Sustain Prod Consum 2018;14:66–81. https://doi.org/10.1016/j.spc.2018.01.001.
[18] Derrible S, Reeder M. The cost of over-cooling commercial buildings in the United States. Energy Build 2015;108:304–6.
[19] Kaza N. Understanding the spectrum of residential energy consumption: a quantile

regression approach. Energy Policy 2010;38:6574–85.
[20] Yun GY, Steemers K. Behavioural, physical and socio-economic factors in household cooling energy consumption. Appl Energy 2011;88:2191–200.
[21] Christenson M, Manz H, Gyalistras D. Climate warming impact on degree-days and building energy demand in Switzerland. Energy Convers Manage 2006;47:671–86. https://doi.org/10.1016/j.enconman.2005.06.009.
[22] Webber H, Zhao G, Wolf J, Britz W, de Vries W, Gaiser T, et al. Climate change impacts on European crop yields: do we need to consider nitrogen limitation? Eur J Agron 2015;71:123–34. https://doi.org/10.1016/j.eja.2015.09.002.
[23] Isaac M, Van Vuuren DP. Modeling global residential sector energy demand for heating and air conditioning in the context of climate change. Energy policy 2009;37(2):507–21. https://doi.org/10.1016/j.enpol.2008.09.051.
[24] Scott MJ, Wrench LE, Hadley DL. Effects of climate change on commercial building energy demand. Energy Sources 1994;16:317–32. https://doi.org/10.1080/00908319408909081.
[25] Dirks JA, Gorrissen WJ, Hathaway JH, Skorski DC, Scott MJ, Pulsipher TC, et al. Impacts of climate change on energy consumption and peak demand in buildings: a detailed regional approach. Energy 2015;79:20–32. https://doi.org/10.1016/j.energy.2014.08.081.
[26] Crowley C, Joutz FL. Weather effects on electricity loads: modeling and forecasting; 2005.
[27] Bugnion V, Sogomonian A, Swindle G. Modelling weather-sensitive electrical loads. Energy Risk 2007:84–9.
[28] Nahid-Al-Masood Ahsan Q. A Methodology for Identification of Weather Sensitive Component of Electrical Load Using Empirical Mode Decomposition Technique. Energy Power Eng 2013;5:293–300.
[29] Sullivan P, Colman J, Kalendra E. Predicting the response of electricity load to climate change; 2015.
[30] Franco G, Sanstad AH. Climate change and electricity demand in California. Clim Change 2008;87:139–51. https://doi.org/10.1007/s10584-007-9364-y.
[31] Perera ATD, Coccolo S, Scartezzini J-L, Mauree D. Quantifying the impact of urban climate by extending the boundaries of urban energy system modeling. Appl Energy 2018;222:847–60. https://doi.org/10.1016/J.APENERGY.2018.04.004.
[32] Ahmed T, Muttaqi KM, Agalgaonkar AP. Climate change impacts on electricity demand in the State of New South Wales, Australia. Appl Energy 2012;98:376–83. https://doi.org/10.1016/j.apenergy.2012.03.059.
[33] Apadula F, Bassini A, Elli A, Scapin S. Relationships between meteorological variables and monthly electricity demand. Appl Energy 2012;98:346–56. https://doi.org/10.1016/j.apenergy.2012.03.053.
[34] Mirasgedis S, Sarafidis Y, Georgopoulou E, Kotroni V, Lagouvardos K, Lalas DP. Modeling framework for estimating impacts of climate change on electricity demand at regional level: case of Greece. Energy Convers Manage 2007;48:1737–50. https://doi.org/10.1016/j.enconman.2006.10.022.
[35] Badri MA. Analysis of demand for electricity in the United States. Energy 1992;17:725–33. https://doi.org/10.1016/0360-5442(92)90080-J.
[36] Ruth M, Lin A-C. Regional energy demand and adaptations to climate change: methodology and application to the state of Maryland, USA. Energy Policy 2006;34:2820–33. https://doi.org/10.1016/j.enpol.2005.04.016.
[37] Singh S, Kennedy C. Estimating future energy use and CO_2 emissions of the world's cities. Environ Pollut 2015;203:271–8. https://doi.org/10.1016/j.envpol.2015.03.039.
[38] EIA. Form EIA-861M (formerly EIA-826) detailed data. Dec 6, 2017 2017: Sales and revenue aggregated, 1990 till present. <https://www.eia.gov/electricity/data/eia861m/index.html> [accessed June 27, 2016].
[39] Mukherjee S, Nateghi R. Climate, weather, socio-economic and electricity usage data for the residential and commercial sectors in FL, U.S. Data Br 2017;13:192–5. https://doi.org/10.1016/j.dib.2017.05.031.
[40] USDL. Bureau of Labor Statistics: Data Tools. United States Dep Labor 2016. <http://www.bls.gov/data/> [accessed June 7, 2016].
[41] Bureau of Labor Statistics. Table 24. Historical Consumer Price Index for All Urban Consumers (CPI-U): U.S. city average, all items. CPI Detail Rep Data December 2015; 2015.
[42] Hintze JL, Nelson RD. Violin plots: a box plot-density trace synergism. Am Stat 1998;52:181–4.
[43] Hastie T, Tibshirani R, Friedman JH. Neural network. Elem. stat. learn. data mining, inference predict. Springer; 2008. p. 389–415.
[44] James G, Witten D, Hastie T, Tibshirani R. Bias-variance trade-off for k-fold cross-validation. An introd. to stat. learn. – with appl. R. Springer; 2013. p. 24.
[45] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning. second ed. Springer; 2008.
[46] Hoard M, Homer J, Manley W, Furbee P, Haque A, Helmkamp J. Systems modeling in support of evidence-based disaster planning for rural areas. Int J Hyg Environ Health 2005;208:117–25. https://doi.org/10.1016/j.ijheh.2005.01.011.
[47] Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. Renew Sustain Energy Rev 2009;13:1819–35.
[48] Energy Commission: State of California. California's Energy Efficiency Standards Have Saved Billions. Efficiency; 2018. <http://www.energy.ca.gov/efficiency/savings.html> [accessed April 9, 2018].
[49] Energy Commission: State of California. California Energy Commission – Tracking Progress; 2017.
[50] Office of Energy Efficiency and Renewable Energy. Energy Incentive Programs, New York. Fed Energy Manag Progr; 2015. <https://www.energy.gov/eere/femp/energy-incentive-programs-new-york> [accessed April 9, 2018].