Applied Energy 262 (2020) 114419

Contents lists available at ScienceDirect

Applied Energy

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

Evaluating the climate sensitivity of coupled electricity-natural gas demand using a multivariate framework

Renee Obringer[a], Sayanti Mukherjee[b,*], Roshanak Nateghi[a,c]

[a] Environmental and Ecological Engineering, Purdue University, 500 Central Drive, West Lafayette, IN 47907, USA
[b] Department of Industrial and Systems Engineering, University at Buffalo - The State University of New York, Buffalo, NY 14260, USA
[c] School of Industrial Engineering, Purdue University, 315 N. Grant Street, West Lafayette, IN 47907, USA

HIGHLIGHTS

- Multi-sector assessment of climate-sensitive portion of NY's energy demand.
- Multivariate framework more accurately predicts energy demand than univariate model.
- Season-to-date heating/cooling degree-days are critical across seasons and sectors.
- Dew point is crucial for predicting demand in intermediate and winter months.
- Essential for utilities that need to predict electricity-natural gas coupled demand.

ARTICLE INFO

Keywords:
Coupled electricity-natural gas demand
Climate-energy nexus
Season-specific climate sensitivity
Multi-sector approach
Multivariate analysis
Predictive modeling

ABSTRACT

Projected climate change will significantly influence the shape of the end-use energy demand profiles for space conditioning—leading to a likely increase in cooling needs and a subsequent decrease in heating needs. This shift will put pressure on existing infrastructure and utility companies to meet a demand that was not accounted for in the initial design of the systems. Furthermore, the traditional linear models typically used to predict energy demand focus on isolating either the electricity or natural gas demand, even though the two demands are highly interconnected. This practice often leads to less accurate predictions for both demand profiles. Here, we propose a multivariate, multi-sector (i.e., residential, commercial, industrial) framework to model the climate sensitivity of the coupled electricity and natural gas demand simultaneously, leveraging advanced statistical learning algorithms. Our results indicate that the season-to-date heating and cooling degree-days, as well as the dew point temperature are the key predictors for both the electricity and natural gas demand. We also found that the energy sector is most sensitive to climate during the autumn and spring (intermediate) seasons, followed by the summer and winter seasons. Moreover, the proposed model outperforms a similar univariate model in terms of predictive accuracy, indicating the importance of accounting for the interdependence within the energy sectors. By providing accurate predictions of the electricity and natural gas demand, the proposed framework can help infrastructure planners and operators make informed decisions towards ensuring balanced energy delivery and minimizing supply inadequacy risks under future climate variability and change.

1. Introduction

Projected climate change—characterized by hot and humid summers, warmer and milder winters, shifts in precipitation patterns and more frequent extreme weather events such as droughts and heatwaves—will significantly influence the shape of the end-use energy demand profiles for space conditioning. It is projected that under a 1.8° F increase in mean temperature, the demand for cooling is expected to increase by 5–20%, whereas that for heating is likely to decrease by about 3–15% across the United States [1]. Such shifts in cooling and heating needs will significantly influence the shape of the demand curves for electricity end-use (mostly used for space cooling) [2–9] and natural gas end-use (mostly used for space heating) [10,11] respectively. Moreover, as the climate continues to warm and extreme weather events intensify, cities will need to work to improve the resilience of their energy infrastructure to cope with the changing

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

https://doi.org/10.1016/j.apenergy.2019.114419
Received 20 September 2019; Received in revised form 15 December 2019; Accepted 17 December 2019
Available online 16 January 2020

conditions in addition to the changing demand profiles [12]. Furthermore, several non-climatic factors such as population shifts, socioeconomic changes, and technological advancements, influence the patterns of electricity and natural gas demands [13], adding additional complexity to demand forecasting research. Although a number of studies have focused on establishing the end-use demand curves, accurate demand forecasting still remains a challenging task. This is mostly due to inability of the traditional models (e.g., generalized linear models) to capture the nonlinearities and interdependencies within the energy demand sectors, as well as the exclusion of some significant weather factors, such as dew point temperature, which was found to have significant influence on the energy demand patterns in the previous studies [2–5]. Additionally, owing to the growing utilization of natural gas-fired power plants, the dependence of the electricity sector on natural gas fuel is increasing [14]. This growing interdependence between the two energy sectors combined with the lack of accurate forecasts has created efficiency and reliability challenges in both the sectors [14].

Given the complex interdependence between the natural gas and electricity sectors, we propose a data-driven, integrated *multivariate* predictive modeling framework to harness the covariance structure between the energy sectors while evaluating the climate sensitivity of the demand nexus. Electricity and natural gas consumption is highly correlated, making the proposed model a good candidate for the current study. In order to assess the added benefits in leveraging a multivariate framework, we also implemented a univariate methodology based on the same tree-boosting algorithm applied in the multivariate framework. Our proposed framework uses a multi-sector approach, meaning that the model accounts for the interdependence in energy consumption across the various sectors—residential, commercial, industrial and electric power plant (EPP). Sector-level studies are important for modeling energy demand because the patterns and types of energy usage vary significantly across the different sectors, which has a significant impact on the climate-sensitivity of energy demand [15]. For example, the residential sector's energy demand is more sensitive to climate and has more spatiotemporal heterogeneity compared to the commercial and industrial sectors because it is more influenced by consumer behavior [5,8].

Additionally, state-level studies have been found to be effective to implement policy analysis [5,11]. The reasons being, both state-level [5] and region-specific [11] studies can (a) account for the geographical variations in climate change impacts; (b) minimize the effects of unobserved heterogeneities that may arise from regional/state-level policy, characteristics of the population and industries, etc; and, (c) help in analyzing sectoral demand as the data is mostly collected at state-/region- levels. Thus, we conducted a state-level analysis in this study. We established our framework using the state of New York as a case study because:

(i) it is the largest northeastern state and the fourth most populous state in the nation with the third-largest economy [16];
(ii) it is the fifth-largest natural gas consumer in the U.S. as of 2017 [16];
(iii) in the state, the majority of the natural gas consumption is attributed to electric power generation and space heating during the colder months [16];
(iv) as of 2017, two-fifths of the state's net electricity generation came from natural gas. In fact, more than half of NY's generating capacity is at natural gas-fired power plants while more than one-third of the capacity is at the units with dual-fuel capability that can use either natural gas or fuel oil [16].

Although, in this paper, we present the results specific to the state of New York (used as a case study in this research), our proposed multivariate, multi-sector framework is generalizable, such that similar studies can be conducted to investigate the climate sensitivity of coupled electricity-natural gas demand in the other states and geographical regions, conditioned on the availability and accessibility of the relevant data and information. Thus, our work contributes to the discipline of risk-informed decision making, particularly as it applies to the capacity planning and climate-informed demand projections by the utility managers.

The structure of this article is as follows: Section 2 presents a brief overview of the existing literature highlighting the current knowledge and research gaps. Data collection, data preprocessing and aggregation are presented in Section 3. Sections 4 and 5 discuss our methodology and results respectively. Section 6 summarizes our key findings and the relevant policy implications of this research.

2. Background information in climate impacts on energy demand

The relationship between climate and energy demand has been well established in previous studies. However, most of the existing body of literature is dedicated to isolated assessments of climate impacts on specific types of end-use energy demand (i.e., electricity or natural gas) or aggregated energy demand. It is likely that these studies have missed valuable information by not considering the complex, interdependent nature of electricity and natural gas demands. On the electricity side, for example, Sailor [17] modeled the climate change impact on the residential and commercial sectors' electricity consumption by leveraging a linear regression model using a degree-day approach, while not accounting for socioeconomic and other non-climatic factors. Similarly, Sullivan et al. [18] assessed the impact of climate change on long term electric loads across 300 transmission zones and 16 seasonal and diurnal time periods in the U.S., leveraging simple linear regression. However, these models were based on simple linear regression, which is often an inaccurate representation of real systems. To improve this aspect of electricity demand forecasting, Mirasgedis et al. [19] implemented a multiple linear regression model to evaluate the climate sensitivity of electricity demand within the Greek interconnected power system, while accounting for the influence of non-climatic socioeconomic factors such as population, monthly-regional gross domestic product (GDP), etc. Likewise, Lam [20] investigated the relationship between the residential electricity demand, the climate, and the economic factors for Hong Kong using a multiple linear regression model. The authors concluded that household income, household size, electricity price and cooling degree-days were the key factors influencing both the seasonal and annual residential electricity demand [20]. These studies were able to improve upon previous work by implementing more advanced linear regression models. However, recent work has shown that the relationship between climate and electricity demand is non-linear, thus necessitating the move from linear regression to more complex methodologies. Nahid-Al-Masood and Ahsan [21], for example, proposed to identify the climate sensitive portion of the electrical load based on Empirical Mode Decomposition (EMD). The authors identified temperature and humidity to be the key climate factors influencing the climate sensitive portion of the electricity demand [21]. Mukherjee and Nateghi [4,5] evaluated the climate sensitivity of residential and commercial electricity demands and concluded that the complex non-linear climate-electricity demand relationship is best captured by advanced learning models such as Random Forests or Bayesian Additive Regression Trees. Leveraging these advanced learning models, Mukherjee et al. [2] investigated the climate sensitivity of the electricity demand for the top eight energy-intensive states in the U.S., and found that the climate sensitivity of electricity demand is asymmetric; with high-intensity end-use demand being more sensitive to climate variability compared to moderate-intensity end-use demand. The authors identified dew point temperature as the key climate predictor for both the high- and moderate-intensity demand [2].

In contrast to the electricity sector, only a limited number of studies have focused on assessing the impact of climate change on the natural gas end-use demand. Amato et al. [11] estimated the temperature-sensitivity of both electricity and natural gas demand for the Commonwealth of Massachusetts using a two-step modeling and estimation procedure. The study controlled for the influence of non-climatic socioeconomic factors, but failed to consider the potential key climate features such as humidity, precipitation and wind speed or the impact of sector interdependence. Ruth and Lin [22] evaluated the climate change impacts on natural gas, electricity and heating oil end-use demand in the residential and commercial sectors. The authors modeled the energy demand as a function of degree-days (heating and cooling), energy prices, daylight hours, and trend variables, leveraging the fixed effects regression model. Similarly, Warren and LeDuc [15] implemented a linear regression model to relate the natural gas demand in the residential and commercial sectors to gas prices and heating degree-days in a nine-region model for the U.S. The authors recommended that a complete climate sensitivity assessment of natural gas demand necessitates joint consideration of all the users to ensure a more holistic assessment [15]. Finally, Sailor and Muñoz [10] analyzed the climate sensitivity of natural gas end-use demand along with that of the electricity demand for the top eight energy-intensive states, viz., California, Louisiana, Texas, Florida, Washington, Illinois, Ohio and New York leveraging a linear regression model.

Several other studies have focused on analyzing the climate impacts on aggregate energy demand. Aggregate energy demand refers to the total energy demand including both electricity and natural gas. Mukherjee and Nateghi [6] evaluated the climate sensitivity of the sectoral energy demand for the state of Indiana using Bayesian Additive Regression Trees (BART). The authors found that the maximum sustained wind speed, dew point temperature and snowfall were the most important predictors of the residential and commercial energy demand. In another study, Nateghi and Mukherjee [3] developed a generalized predictive model to project the climate-sensitive portion of the aggregate sectoral energy demand until the year 2100, under both the business-as-usual emissions scenario (RCP8.5) and a scenario based on reduced emissions consistent with a 2°C increase in global mean temperature (RCP4.5). Additionally, Raymond et al. [7] estimated the changes in the climate-induced portion of the aggregate energy demand for the state of Indiana under both the aforementioned climate change scenarios, with a focus on potential policy implementations.

The literature review above helps identify the important gaps in the current body of knowledge. First, to the best of our knowledge, none of the studies focused on understanding the climate sensitivity of the energy demand considering the coupled nature of the energy market. Moreover, the spatiotemporal variations in energy demand across the U.S. over various seasons necessitates studying the climate impacts on the heating and cooling energy demand separately across the different seasons in a particular region [22]. Analyzing the climate sensitivity of aggregate energy demand is likely to lose much of the important information related to season-, sector- or region-induced variations in the energy demand. Studies based on aggregate demand often result in forecasting negligible changes in annual energy demand, mostly because the changes in cooling and heating demand offset one another [22]. Second, the models used to evaluate the climate sensitivity of natural gas often assume a linear relationship between climate and the end-use demand. However, in the light of the fact that previous studies established the relationship between electricity demand and climate to be highly non-linear [4,5], it is likely that generalized linear models do not adequately capture the complex and potentially non-linear relationships between the climate and natural gas consumption. Third, most of the studies modeled natural gas demand as a function of degree-days or surface air temperatures. We hypothesize that these are inadequate measures for capturing the trends in heating and cooling as they do not account for relative humidity. Under climate change scenarios, such models will potentially underestimate the climate sensitivity of the demand, as both temperature and humidity are projected to increase [23].

In this study, we propose to address the above-mentioned gaps by developing a data-driven multivariate predictive model to account for the complex interdependence between the electricity and natural gas demand across the various sectors (residential, commercial, industrial and electric power plants). Using the state of New York as a test case, we analyze the climate sensitivity of the coupled energy demand nexus separately for three different seasons—summer, winter, and intermediate (i.e., autumn and spring)—to consider and compare the seasonal variabilities in the demand. In addition to the multivariate model, we also develop a univariate model to evaluate the importance of considering the coupled nature of the energy market while predicting energy demand.

3. Data collection, preprocessing and aggregation

In this section, we present a brief description of the input and response variables obtained from various sources, data preprocessing and trend-adjustment, and, finally the seasonal separation and variable screening processes.

3.1. Input data types and description

Different types of variables—monthly electricity demand, monthly natural gas demand, monthly climate and daily weather data, and socioeconomic information—were obtained from multiple sources (described below).

(i) **Electricity demand**: Monthly electricity consumption (sales) data for the state of New York was obtained from the U.S. Energy Information Administration (EIA) Form EIA-861 M (formerly EIA-826) [24]. The EIA collects data at an aggregated level for each energy sector, i.e., residential, commercial, industrial, etc. Thus, this dataset does not include information on the user profiles (e.g., number of homes, square footage, etc.), etc.; rather it presents a holistic view of the total electricity use across the state of New York. Data was obtained for the three sectors—residential, commercial, and industrial—during the period of January 2001-December 2017.

(ii) **Natural gas demand**: Monthly natural gas consumption by end-use types was extracted from the U.S. Energy Information Administration (EIA) [25]. Similar to the electricity demand, the natural gas demand was available aggregated at the state-level, and therefore there was no data on user profile details. We extracted the data for the four different sectors—residential, commercial, industrial and electric power plants (EPP)—during the period of January 2001-December 2017.

(iii) **Climate and weather data**: Monthly climate data for the state of New York was obtained from National Oceanic and Atmospheric Administration (NOAA), starting from January 2001-December 2017. Variables include monthly and season-to-date heating and cooling degree days, snow depth, number of days above 90° F (32° C) or below 0° F (-18° C), etc. [26]. Daily weather data was requested from NOAA's National Centers for Environmental Information (NCEI) for the period 01-January-2001 through 31-December-2017. The weather data, recorded at multiple weather stations (95 different stations spread across the state of NY), included information on the following weather variables: dew point temperature, daily average and maximum air temperatures, daily mean and maximum wind speeds, maximum wind gust and visibility [26].

(iv) **Socioeconomic data**: Economic data for the state of New York was extracted from the U.S. Department of Labor, Bureau of Labor Statistics' database [27]. We considered unemployment rate and gross state product (GSP) in our analysis as these were found to be the important socioeconomic features that need to be controlled for to evaluate the climate-sensitivity of energy demands [2,5].

3.2. Response variable trend-adjustment

Following the data collection step, the response variables (electricity and natural gas consumption in the residential, commercial, industrial, and electricity power plant sectors) were detrended. Trend adjustment was performed, following the method presented by Sailor and Muñoz [10], to limit the impact of non-climatic phenomena including technological advancements and cultural shifts that also affect the demand structure. Since the goal of this study was to evaluate the climate sensitivity of the interdependent electricity and natural gas demand, this trend adjustment process was especially important. More information on the trend adjustment process, including equations, can be found in the Supplementary Materials.

3.3. Input variables preprocessing

Daily weather variables obtained from 95 various weather stations were spatiotemporally aggregated to render monthly average values for the state of NY over the years of the analysis. The monthly climate variables obtained from NOAA were also spatially aggregated to produce the mean values for the state of NY for the period of the analysis. The preprocessed climate and weather data were then merged with the socioeconomic data to generate the predictor variable dataset.

3.4. Seasonal separation

As discussed before, we implemented our analysis separately for the three different seasons—summer, winter, and intermediate—to account for the seasonal variability in the electricity and natural gas consumption. All the data (response and predictor variables) was divided into seasons based on the standard seasonal changes for the state of NY. Specifically, spring is from March-May, summer is from June-August, autumn is from September-November, and winter is from December-February [28]. For modeling purposes, the spring and autumn months were combined to form a single 'intermediate period'. This seasonal separation was performed to account for differences in the usage patterns of electricity and natural gas—electricity use increases in the summer owing to increased space cooling, while natural gas use increases in the winter due to higher demand of space heating. Going forward, the trend-adjusted, seasonally-separated datasets were used throughout the modeling and analysis process.

3.5. Input variable screening

A total of 57 input features related to weather, climate, and socioeconomic variables were gathered to be utilized within our framework. While the predictive performance of our proposed modeling framework (described in Section 4) is not directly impacted by multicollinearity, the presence of high-dimensional, correlated predictors could lead to a 'masking effect' of certain variables or model overfitting [29–32]. In other words, the model will still perform well in terms of predictive accuracy in the face of multicollinearity, but there may be indirect effects, such as the masking or suppression of uncorrelated predictor variables in favor of the correlated ones. For example, if there are five temperature related predictors that are highly correlated and one relative humidity predictor that is not, the effect of relative humidity on the response variable may not be considered important within the algorithm due to the overshadowing caused by the other correlated temperature variables. However, if only one or two temperature variables are included in the model, there might be a stronger signal from the relative humidity variable, which was previously unseen. With this in mind, we decided to remove the highly correlated variables ($\rho > 0.8$) to limit the potential for overfitting or the masking of important, non-correlated variables. The correlation plots of all the initial variables included in the study are located in the Supplementary Material (Figs. S1–S3). After the initial variable screening, the variables considered during the summer season were: dew point temperature, number of days with a temperature above 90 °F (32 °C), number of cooling degree-days, unemployment rate, and state GDP. For winter, the final variables included were: number of heating degree days (HDD) in the season, number of days below 0 °F (-18 °C), dew point temperature, snow depth, unemployment rate, and state GDP. Finally, the intermediate seasons' variables included number of days with a temperature above 90 °F (32 °C), the number of heating degree days (HDD) in the season, number of days below 0 °F (-18 °C), dew point temperature, snow depth, unemployment rate, and state GDP.

4. Methodology

The methodology used in this study is based on a state-of-the-art multivariate tree boosting algorithm [33], which has recently begun to grow in popularity within the area of infrastructure analysis [34–36]. In these studies, the algorithm was used to evaluate resilience of electric distribution networks to hurricanes for the first time. The results indicated that using a multi-outcome algorithm to measure resilience led to analyzing different areas of vulnerability throughout the network as well as different proactive measures being selected as optimal [35]. Additionally, the algorithm has been used to study the water-electricity nexus, where it is established that it is highly beneficial to use a multi-outcome model in terms of the predictive accuracy of the models [34]. Finally, the model was recently used to assess the resilience of infrastructure systems to tsunamis in Japan. In this study, using the multivariate algorithm resulted in identifying different metrics of resilience being important as well as the models led to an increase in the overall predictive accuracy [36]. Overall, the multivariate algorithm has had its success when used in the studies of interconnected systems. However, it has not yet been used to model the interdependent electricity and natural gas demand, and climate nexus. This novel application will allow utility managers and policy makers to gain a deeper understanding of the relationship between electricity use, natural gas use, and climate across a variety of consumer sectors. In the following sections, we will first describe the modeling framework before delving into details of the algorithm.

4.1. Modeling framework

There were three main steps to the modeling framework: (1) data collection, aggregation, and preprocessing; (2) model training and testing; and (3) statistical inferencing. These steps are depicted in

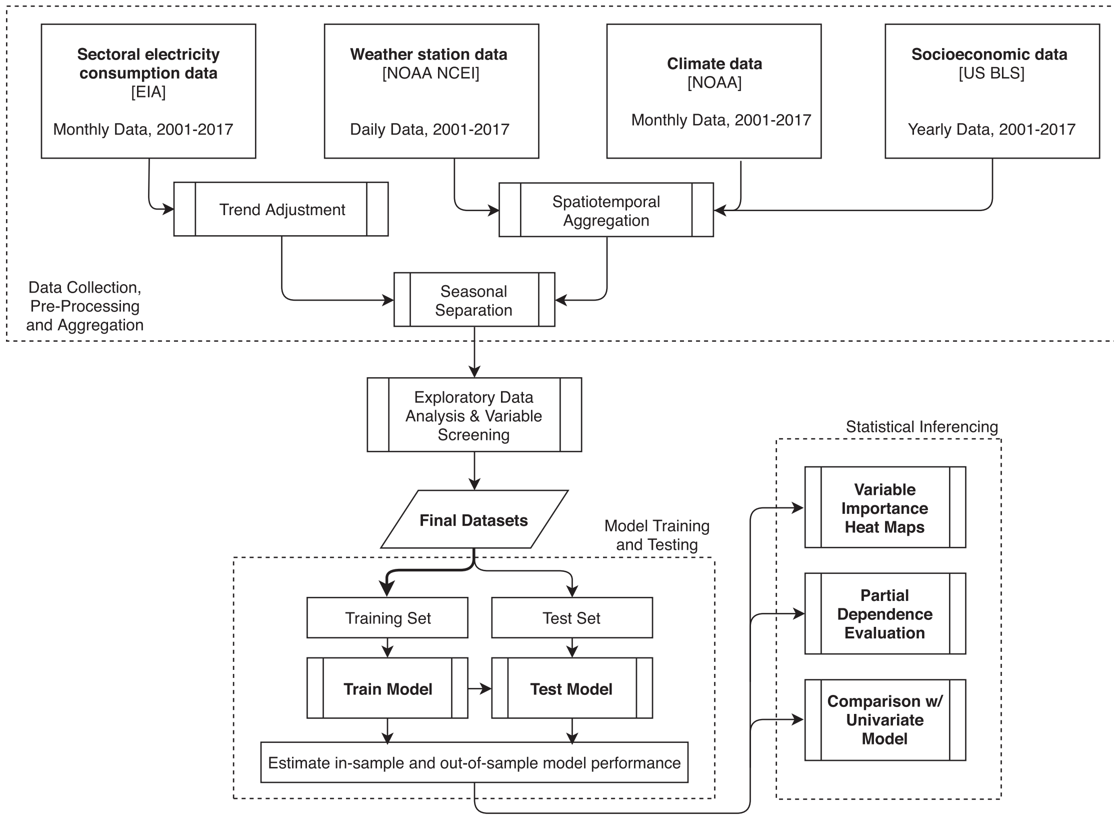

Fig. 1. Schematic depicting the proposed multi-variate, multi-sector modeling framework.

Fig. 1. The first step—data collection, preprocessing and aggregation—was described in Section 3.2. In the second step—model training and testing—the algorithm was applied within a cross-validation loop (discussed in Section 4.1.1). Finally, the third step—statistical inferencing (discussed further in Section 4.1.2)—was performed using the results from the final model run. This included three main aspects: (1) a comparison between the multivariate, univariate, and null model runs; (2) an evaluation of the variable importance; and (3) an analysis of the partial dependency between the response variable and the key predictors. Additional details on these steps can be found below.

4.1.1. Predictive modeling

We leverage predictive modeling (aka supervised learning) to characterize the climate sensitivity of the coupled electricity-natural gas demand nexus. Supervised learning is a mapping of inputs to outputs, i.e., $f: \mathscr{X} \rightarrow \mathscr{Y}$. The goal is to estimate the prediction function such that the lost function (L) of interest—some measure of the distance (l) between the true values and the predicted values—is minimized, i.e.: $L(f) \triangleq \mathbb{E}_{x,y\sim P}[l(f(x), y)]$ [37].

Among the wide library of supervised predictive algorithms, tree-based methods are one of the most popular non-parametric learning techniques [37], particularly in modeling energy demand [2–7]. Tree-based models are competitive—in terms of predictive accuracy—with many of the state-of-the art statistical machine learning algorithms [38–40], and lend themselves more easily to interpretation and inferencing compared to other "black box" algorithms, such as deep learning and support vector machines [37,41]. In this paper, we used an advanced predictive learner—based on tree-ensembles—that takes advantage of the covariance structure of multiple response variables to accurately estimate the complex and non-linear dependencies between the target variables and the independent variables. Specifically, the predictive model of the coupled electricity and natural gas demands was developed based on a multivariate extension of the gradient boosted regression trees (described under the Algorithm Specification section). We also developed a univariate model for the individual response variables, using gradient tree boosting, to evaluate the effectiveness of the multivariate methodology in capturing the dependencies in the coupled electricity-natural gas demand. For testing the out-of-sample predictive performance of the trained models, we implemented the leave-one-out cross-validation (LOOCV). Cross-validation, in general, balances the bias and variance in the model, resulting in more robust predictions [37]. LOOCV is a style of cross-validation that is often done with smaller datasets, such as the one considered in this study.

Algorithm Specification: Gradient boosted regression trees is an ensemble-of-trees method that takes advantage of the boosting meta-algorithm to improve model predictive accuracy [42]. The boosting meta-algorithm works by sequentially fitting decision-tree classifiers-- where in each iteration more weight is given to the better classifiers than the misclassified points in order to reduce the overall loss function and enhance the predictive accuracy. Boosting is represented mathematically in the equation below.

$$\mathscr{F}(x) = \sum_{m=1}^{M} \omega_m \mathscr{T}_m(x) \tag{1}$$

Here $\mathscr{F}(x)$ is the final ensemble model, M is the total number of iterations to be completed, ω_m is the weight of each prediction (step size) that controls how quickly the model fits to the observed data, and $\mathscr{T}_m$ is the tree models fitted to the input variable x at iteration m.

Multivariate tree boosting extends gradient boosted regression trees to a multivariate (i.e., multi-response) case. Thus, the multivariate algorithm allows for *simultaneous* prediction of multiple response variables [33]. Specifically, the algorithm iteratively fits trees by minimizing the squared loss for each target variable and maximizing the covariance discrepancy in the multidimensional response variable at each gradient step. Thus, at each iteration, a prediction is made for each response variable such that the loss function is minimized and the covariance discrepancy between the current and previous prediction is maximized. This allows each subsequent estimation to be incrementally more accurate than the previous step, while ensuring the independent variables that account for the largest fraction of the covariance in the nexus of the response variables are selected. The steps of the algorithm are summarized below:

Algorithm 1. Multivariate Ensemble Tree Boosting Algorithm [33]

```
for m in 1, …, M steps (regression trees) do
   for r in 1, …, R quantitative response variables (e.g., electricity and natural gas demands) do
      train tree m^(r) to residuals, and estimate the covariance discrepancy D_{m,r}
   end for
   Select the response y^(r) corresponding to the regression tree that yielded the maximum D_{m,r}
   Update residuals by subtracting the predictions of the tree fitted to y^(r), multiplied by step-size.
end for
```

4.1.2. Statistical inferencing

Statistical inferencing consisted of three main steps. The first step involved models' performance comparison. In the literature, much of the energy demand predictions are performed using univariate approaches that do not account for the interdependent nature of the two demand structures. As discussed before, we evaluated the differences in performances between the multivariate model presented above and a univariate model. The latter uses a similar algorithm, namely gradient tree boosting [43], which is the basis for the multivariate tree boosting algorithm used here [33]. Additionally, we compared the model results to the null model. In the second step, we analyzed the relative importance of the independent variables on the electricity and natural gas demands across the residential, commercial, industrial and EPP sectors. Finally, in the third step, we analyzed the partial dependence of the predictors on the response variables within the multivariate model. This allowed us to discern which predictor variables were most important for accurately predicting the energy demands as well as the nature of those relationships. The results from this analysis are detailed in the following section.

5. Results and discussion

In this section we will discuss the results from the model predictions as well as the statistical inferencing techniques (steps) outlined above.

5.1. Model comparison

To compare the multivariate, univariate, and null models, we assessed the models' performance using a variety of measures, including the normalized root-mean-square error (NRMSE), shown in Eq. 2. The NRMSE is especially beneficial when comparing multiple models' outputs having different units, as it puts all the measures of error on the same scale.

$$NRMSE = \frac{\sqrt{\frac{\sum(\hat{x}-x)^2}{n}}}{x_{max}-x_{min}} \tag{2}$$

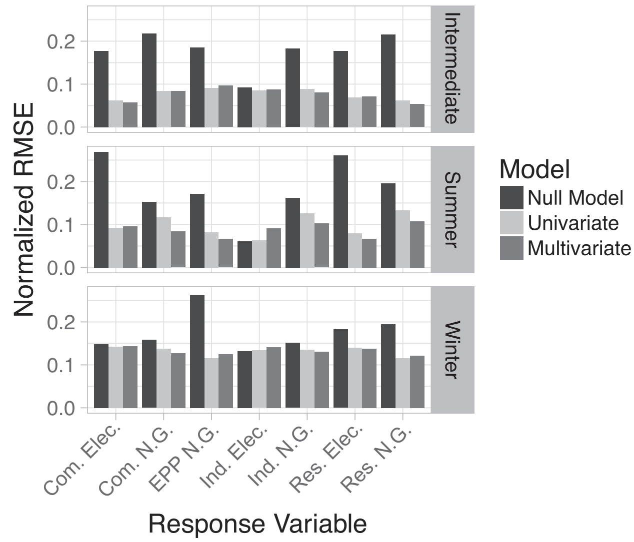

Fig. 2. Normalized root mean square error (NRMSE) for the multivariate, univariate, and null models. NRMSE is the RMSE normalized by the data range at the time of calculation, such that each response variable is on the same scale.

The NRMSE analysis results for each of the response variables are presented in the form of a bar plot in Fig. 2. In this plot, the response variables are presented along the x-axis while the three different seasons are depicted in separate panels, as labeled on the right-hand y-axis. From this plot, we can observe that the null model (i.e., the black bars) generally performs worse than either of the predictive modeling algorithms, which is expected as the null model represents the prediction based solely on the mean value of the observed response variable. However, in some cases, namely the industrial electricity sector, the null model performs better than both the multivariate and univariate models. This indicates a lack of climate sensitivity in the industrial electricity demand, as the mean value is a better predictor than the range of climate variables included in this study. This result is in line with the findings from previous studies, which have demonstrated that the industrial sector is the least climate sensitive followed by commercial sector, while the residential sector is most sensitive to climate [4,5]. This is likely due to the nature of end-uses [4]. Further exploration is needed to determine the optimal variables for predicting the industrial electricity demand.

In addition to comparing the null models to the predictive modeling algorithms, Fig. 2 demonstrates the improved performance of the multivariate model over the univariate model in the various energy sectors considered in our research. The multivariate model (i.e., the dark gray bars) has a lower NRMSE than the univariate model (i.e., the light gray bars), especially in the summer period. Since the main difference between the multivariate and univariate models is the consideration of response variable interdependence, the results indicate that the interdependence plays a key role in accurately predicting the electricity and natural gas demands across the various sectors during the summer months. A notable exception is the industrial electricity demand, in which the univariate model performs significantly better than the multivariate model in the summer period. This suggests that there is little interdependence between industrial electricity and the other sectors. In the winter and intermediate months, there is more similarity between the univariate and multivariate model performance, however, the multivariate model shows marginal improvement in the various sectors. That being said, even marginal improvement can lead to significant impacts, when one considers the size of the energy sector. For example, the predictions for residential natural gas use in the intermediate months is slightly better in the multivariate model than the univariate model (i.e., a NRMSE of 0.053 compared to 0.062; see Table 1), which is approximately a 15% difference in predictive accuracy (i.e., out-of-sample NRMSE). In terms of the residential natural gas use in the state of New York, this could mean an underestimation of 4574 million cubic feet (MMcf; 129 million m^3) in natural gas demand in an average month (i.e., 15% of the mean residential natural gas consumption in the intermediate months, i.e., 15% of 30,496 MMcf which equals to 4574 MMcf), and up to 11,384 MMcf (322 million m^3) in a peak month. If one considers the other natural gas sectors, there would be no difference in the commercial sector, a surplus of 1515 MMcf (42 million m^3) in the electric power plant sector, and a shortage of 523 MMcf (14 million m^3) in the industrial sector, under average conditions. In other words, if the natural gas plant managers relied on the univariate model for their projections during the intermediate seasons, they could experience a shortage of 2536 MMcf (71 million m^3) in average conditions. In peak months, this shortage could be up to 8261 MMcf (233 million m^3).

Interestingly, the EPP natural gas use in the winter and intermediate months is better modeled by the univariate model. This suggests that during the colder months, electric power plant (EPP) natural gas use has lower interdependence with the rest of the energy sectors than during the warmer months. That being said, overall, the multivariate model shows improvement over both the null model and the univariate model. Additional measures of out-of-sample (i.e., test data) model performance are shown in Table 1. Here, the benefits of the multivariate model are further demonstrated. In particular, Table 1 presents the NRMSE values that are plotted in Fig. 2, as well as the percent of improvement over the null model. The improvement data shows the relative increase (or decrease) in predictive accuracy that each model has shown compared to the null, or mean-only, model. In general, both models show improvement over the null model, but that the multivariate model has slightly better improvement. See Supplemental Table S1 for the in-sample (i.e., training data) model performance.

Additionally, Table 1 shows the out-of-sample R^2 values. The R^2 values, which indicate how much variance in the original response data is explained by the model (i.e., goodness-of-fit), are generally closer to one in the multivariate model compared to the univariate model. This indicates that the multivariate model is able to fit the original data better than the univariate model. In Table 1, for example, one can see the ability of the multivariate model to accurately predict the energy consumption across the sectors in the intermediate months. In fact, with the exception of industrial electricity, which has been shown to be less sensitive to climate variables [4,5], all of the sectors are fairly well-predicted by the multivariate model (e.g., all the R^2 values for the test set are greater than 0.70, with the exception of industrial electricity use). This trend continues during the summer months. Here, the multivariate model accurately predicts the demand data in each sector, except the industrial electricity sector, although the R^2 values are slightly lower than those for the intermediate months, especially in natural gas use (see Table 1). These R^2 values represent the fit between the multivariate model predictions and the observed values in the test set, meaning higher values corresponding to a better fit of the model.

Table 1
The out-of-sample (i.e., testing data) model performance of each sector in the multivariate and univariate models in each seasonal period. The improvement over null refers to the percent improvement in error compared to the null (i.e., mean-only) model. Note that a negative number indicates a decrease in model accuracy when compared to the null model. See Supplemental Table S1 for the in-sample (i.e., training data) model performance.

Sector	Multivariate Model			Univariate Model		
	R^2	NRMSE	Improvement over null	R^2	NRMSE	Improvement over null
			Intermediate months			
Commercial electricity	0.92	0.057	68%	0.88	0.062	66%
Commercial natural gas	0.86	0.084	64%	0.83	0.084	64%
Electric power plant	0.75	0.096	52%	0.76	0.091	54%
Industrial electricity	0.40	0.088	4%	0.37	0.085	7%
Industrial natural gas	0.84	0.081	57%	0.79	0.088	54%
Residential electricity	0.84	0.071	60%	0.79	0.069	61%
Residential natural gas	0.87	0.053	75%	0.86	0.062	71%
			Summer months			
Commercial electricity	0.86	0.095	66%	0.82	0.092	64%
Commercial natural gas	0.68	0.084	57%	0.38	0.117	28%
Electric power plant	0.78	0.066	63%	0.70	0.082	53%
Industrial electricity	0.37	0.091	−56%	0.08	0.063	−6%
Industrial natural gas	0.64	0.102	42%	0.46	0.126	25%
Residential electricity	0.89	0.067	74%	0.87	0.079	70%
Residential natural gas	0.69	0.108	56%	0.47	0.133	42%
			Winter months			
Commercial electricity	0.47	0.144	4%	0.27	0.143	2%
Commercial natural gas	0.62	0.127	32%	0.46	0.137	14%
Electric power plant	0.78	0.125	61%	0.78	0.116	59%
Industrial electricity	0.40	0.141	−14%	0.20	0.134	−3%
Industrial natural gas	0.55	0.131	12%	0.31	0.136	14%
Residential electricity	0.61	0.138	23%	0.43	0.140	22%
Residential natural gas	0.70	0.121	49%	0.61	0.116	50%

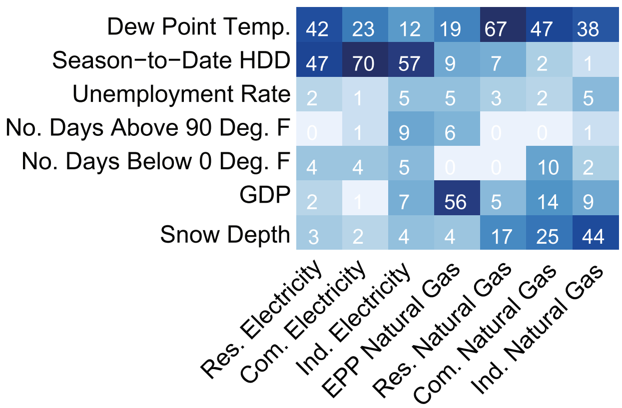

Fig. 3. Relative influence of the various predictors on the response variables for the intermediate months.

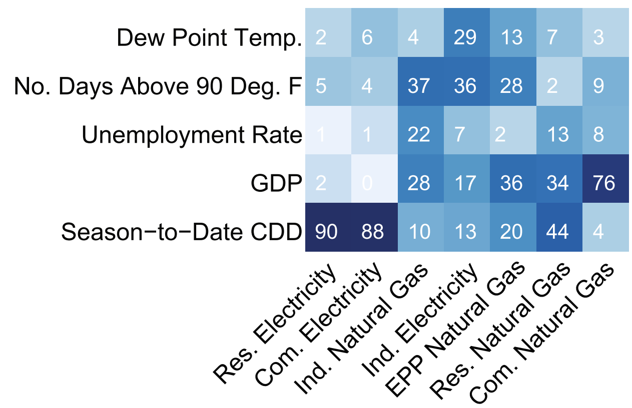

Fig. 5. Relative influence of the various predictors on the response variables for the summer months.

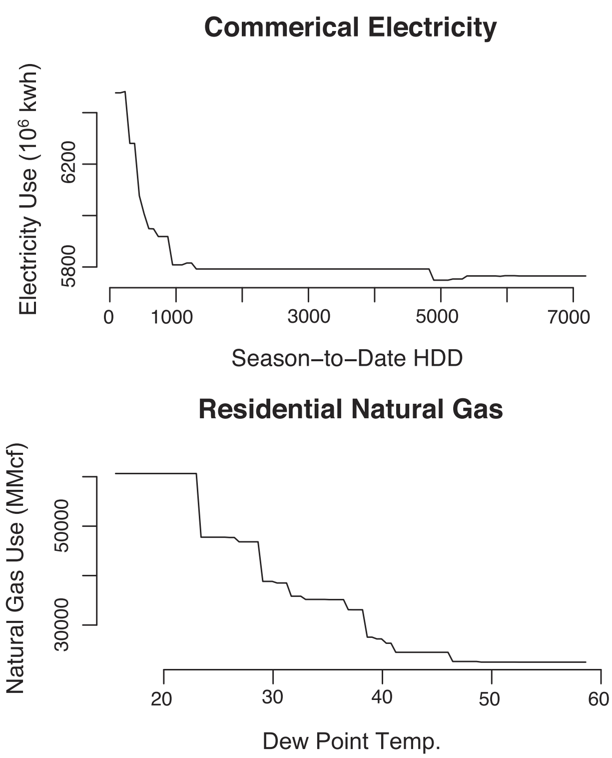

Fig. 4. A selection of the partial dependence plots for the intermediate months: (a) Influence of season-to-date heating degree days (HDD) on electricity demand in the commercial sector; and (b) influence of dew point temperature on natural gas demand in the residential sector.

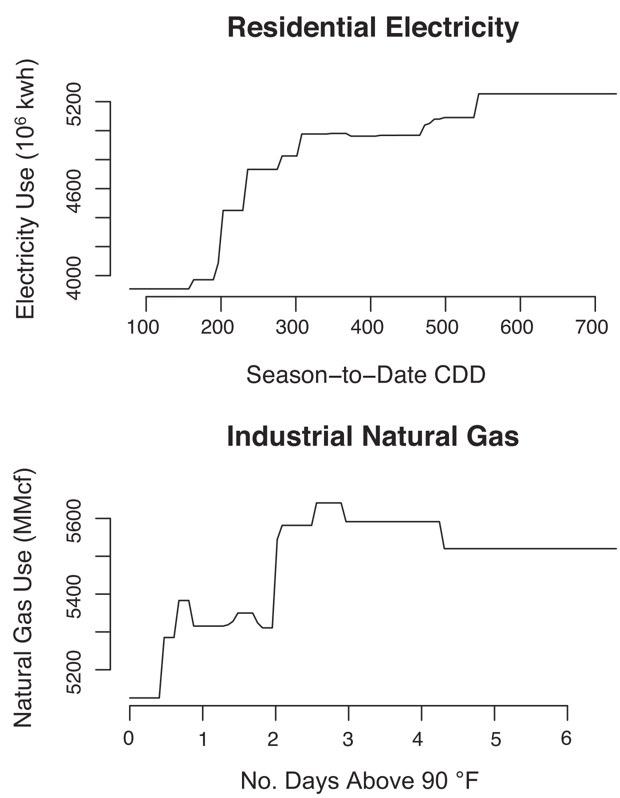

Fig. 6. A selection of the partial dependence plots for the summer months: (a) Influence of season-to-date cooling degree days (CDD) on electricity demand in the residential sector; (b) Influence of number of days above 90 °F on natural gas demand in the industrial sector.

Interestingly, although much less amount of natural gas is used during the summer, at least in the residential and commercial sectors, the multivariate model still accurately predicts the climate-sensitive portion of the demand (e.g., R^2 values greater than 0.60). Finally, in the winter months, the models still fit the data relatively well, but not to the level of accuracy seen in the intermediate and summer months. This suggests that the energy demand during the winter months is less climate sensitive as compared to the summer and intermediate seasons. Visual representations of the model fit can be found in the Supplementary Material (Figs. S4–S6).

5.2. Statistical inferencing: Analyzing climate sensitivity

One of the benefits of tree-based predictive modeling techniques is the ability to interpret the model and make inferences based on the results. One such way to make inferences is to look at the relative influence of the predictors on the response variables. The relative influence can be thought of as the importance of a given predictor to the predictive accuracy of the model. In other words, more influential (or important) predictors are responsible for a larger share of the overall improvement in predictive accuracy. Figs. 3, 5, and 7 depict the relative influence of each predictor variable on each response variable, separated by season. These plots are clustered using hierarchical clustering, such that the predictors that affect the response variables similarly are grouped together. The darker squares represent higher relative influence over the predictive accuracy of a given response variable (the numbers being the actual relative influence score, shown as a fraction out of 100 for each sector). The relative influence plots demonstrate the

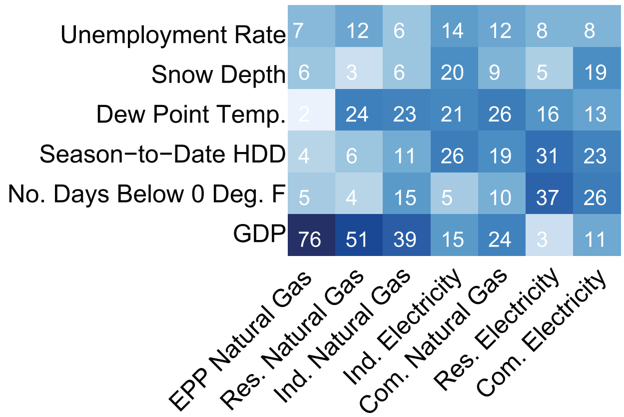

Fig. 7. Relative influence of the various predictors on the response variables for the winter months.

importance of each predictor, but to determine the nature of the relationship (e.g., if the value of the predictor increases, what effect does that have on the response), one needs to assess the partial dependence.

Figs. 4, 6, and 8 show a selection of the partial dependence plots for each season. The predictors plotted are the two most important predictors over all the sectors (i.e., the highest total relative influence score, summed across the sectors), while the response variables were selected based on the highest relative influence score for the chosen predictor variables. For example, in Fig. 6, dew point temperature has the highest total relative influence score (i.e., 42 + 23 + 12 + 19 + 67 + 47 + 38 = 248) and in particular, residential natural gas is influenced the most by dew point temperature (i.e., a relative influence score of 67). Therefore, we chose to show the partial dependence plot of dew point temperature and residential natural gas

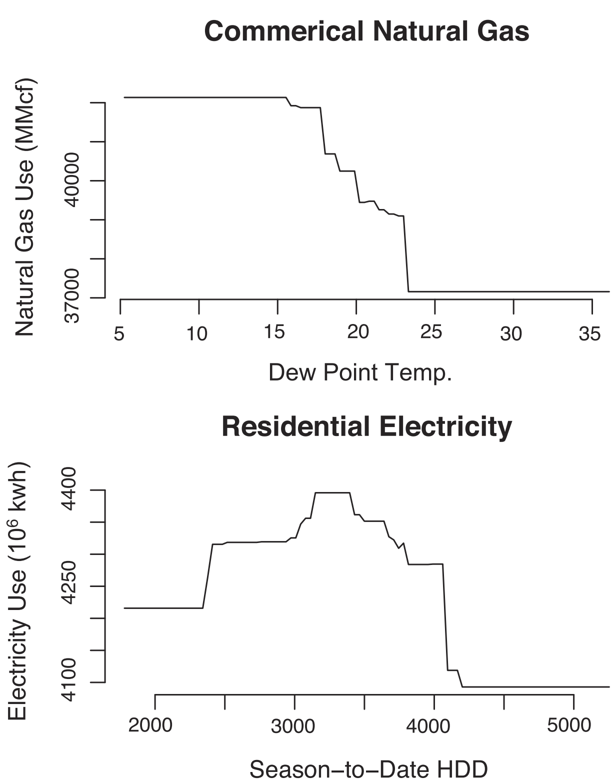

Fig. 8. A selection of the partial dependence plots for the winter months: (a) Influence of season-to-date dew point temperature on natural gas demand in the commercial sector; and (b) influence of season-to-date heating degree days (HDD) on electricity demand in the residential sector.

use. Additional partial dependence plots for the climate variables and sectors not included in this text can be found in Supplementary Figs. S7–S13. In the partial dependence plots, all but the one predictor is held constant, such that the relationship between the response variable and the given predictor can be assessed. It is worth noting that the GDP is highly influential across all the sectors and all the periods, likely signaling the effect of the energy sector on the economy as a whole. That being said, since the purpose of this study was to evaluate the climate sensitivity, GDP was not considered in the partial dependence plot selection.

5.2.1. Intermediate months

During the intermediate months, which included both spring (March–May) and autumn (September–November), dew point temperature and season-to-date heating degree-days are found to be highly influential across all the sectors (see Fig. 3). The importance of dew point temperature echoes with the results of previous studies, which suggested that dew point temperature is a better predictor than heating or cooling degree-days in all the seasons [2,4,5]. Dew point temperature, which is the temperature at which the air becomes saturated with water, is often used in demand studies because it accounts for the moisture in the air, which has the potential to make the temperature appear different to people. This means that as the dew point increases, for example, and people *feel* hotter, they will generally increase their air conditioning use, thus leading to higher electricity consumption, even if the actual temperature is lower.

It is noteworthy that the season-to-date heating degree-days was found to be important, especially for the electricity demand. It is likely that as the season progresses and the number of heating degree-days increases or remains the same (i.e., cooling degree-days are increasing), people are opting to maintain heating or cooling levels out of habit, rather than adjusting daily. In this sense, the season-to-date heating degree-days captures a longer trend in the demand structure than the monthly heating degree-days, hence the importance when predicting seasonal energy use. This is further demonstrated by the partial dependence plots in Fig. 4, which shows as season-to-date heating degree-days increase (i.e., an increasing number of cold days), the electricity use decreases—likely due to less space-cooling and more space-heating (via natural gas). Similarly, as the dew point temperature increases, there is a drop in natural gas use, which is primarily used for space-heating during the winter months.

5.2.2. Summer months

In the summer months (i.e., June–August), the season-to-date cooling degree-days is the most important variable for residential electricity and natural gas end-use, as well as commercial electricity demand, while the number of days above 90 °F (32 °C) is important for the remaining sectors (see Fig. 5). As the number of cooling degree-days increases, indicating that it is getting hotter and people are increasing their use of air conditioning, the electricity use is increasing (see Fig. 6). It is interesting to note that the number of days above 90 °F (32 °C) is the most influential variable on industrial natural gas use (see Fig. 5), and that, as the number of days increases, the amount of natural gas used also increases, up to a point, and then remains constant (see Fig. 6). Since industrial natural gas is used in a variety of manufacturing processes in addition to space-heating, these results could be capturing a trend of increasing productivity during the warmer months.

5.2.3. Winter months

Finally, in the winter months (December–February), dew point temperature and the season-to-date heating degree-days were found to be the most important climate predictors (see Fig. 4). Since the most intensive energy end-use in the winter months is space-heating, it is logical that colder temperatures (and more heating degree-days) lead to increased energy use, regardless of the source (i.e., natural gas or electricity). According to Fig. 8, residential electricity use increases as the number of heating degree-days increases, before dropping, as the HDD count continues rise. This likely points to increased space-heating—whether due to an electric furnace or electric space heaters. Likewise, as the dew point increases, the commercial natural gas use drops, signifying a rise in temperature and reduction in space-heating.

Overall, the relative influence plots show us which variables are important and the partial dependence plots show us how the relationship works—both allow us to make inferences about the interconnected energy demand structure. These results have implications on policy and decision-making at the utility scale, which may not include interdependent modeling of electricity and natural gas demand. For example, predicting the demand tends to be the most difficult during the intermediate months, due to the fluctuations in temperature, which lead to rapid switches from electricity to natural gas use, and vice versa. Our results show that dew point temperature is an important predictor for both electricity and natural gas demand, so if utilities work together, they could potentially improve their projections based on the dew point alone. Although, to get a more accurate prediction, it is recommended that utility managers take all the climate variables into account, which play various roles in predicting consumption within the different energy sectors.

6. Conclusion and policy implications

The purpose of this study was to evaluate the climate sensitivity of the coupled electricity-natural gas demand nexus. Using the multivariate tree boosting algorithm, we were able to demonstrate the importance of considering the interdependencies between electricity and natural gas use in the state of New York, particularly during the warmer summer months. When compared to a similar univariate algorithm, the multivariate model performed better in terms of normalized root-mean-square error, indicating the value of considering multiple sectors at once, even if one is only interested in residential electricity, for example. Additionally, we identified the important variables for each season; we found that dew point temperature was fairly important across all seasons and sectors, especially in predicting the climate-induced demand during the intermediate and winter months and within the natural gas sectors. This aligns with the previous studies on climate sensitivity of energy demand, which established the importance of dew point temperature as a critical climate factor influencing end-use electricity demand [2,4,5]. Our results also indicate the importance of season-to-date heating and cooling degree-days. Previous studies have focused on monthly degree-days, but it is likely that considering the total number of degree-days that have occurred in a season could lead to better predictive accuracy of the energy demand-climate models. This is likely due to the fact that season-to-date measures take into account the trends within a season that may lead to consumer decisions on setting thermostats. Additionally, this framework focused on seasonal models, as the trends in electricity and natural gas demand are intricately linked to the seasonal changes in the surrounding environment. For example, natural gas demand usually picks up in the late autumn and continues through the early spring, when people are switching to space-heating. In this sense, a model that doesn't separate by seasons could be missing major trends within the demand profiles.

The modeling framework and results presented here can be used by utility managers and decision-makers to improve their ability to predict electricity and natural gas use. The framework would be especially helpful in the intermediate months, when the daily fluctuations make it difficult to obtain accurate predictions. For example, in the intermediate months, dew point temperature and season-to-date heating degree-days were found to be important climate predictors. Utility managers that are interested in predicting the demand during the intermediate season, can use these two variables to obtain a close approximation of the demand structure. Moreover, the utility managers in this situation would also see improvement in their predictions if they consider the interdependencies between electricity and natural gas demand. Improving the predictive capabilities of electricity and natural gas demand, especially as the demand structure relates to the climatic conditions, will become increasingly important as the climate becomes more variable and populations continue to rise in future. Although this study focused on the state of New York, the modeling framework is generic enough that can be applied in any number of regions or states around the world. Moreover, there is potential for extending the framework to include other utilities, such as water.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Renee Obringer: Formal analysis, Methodology, Writing - original draft, Writing - review & editing. **Sayanti Mukherjee:** Conceptualization, Methodology, Supervision, Writing - original draft, Writing - review & editing. **Roshanak Nateghi:** Methodology, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgments

The authors would like to acknowledge support from the Purdue University Center for the Environment as well as NSF grants #1826161 & #1832688.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.apenergy.2019.114419.

References

[1] EPA. Climate Impacts on Energy; 2016 < https://19january2017snapshot.epa.gov/climate-impacts/climate-impacts-energy_.html > .
[2] Mukherjee S, Vineeth C, Nateghi R. Evaluating regional climate-electricity demand nexus: a composite bayesian predictive framework. Appl Energy 2019;235:1561–82.
[3] Nateghi R, Mukherjee S. A multi-paradigm framework to assess the impacts of climate change on end-use energy demand. PloS One 2017;12(11). e0188033.
[4] Mukherjee S, Nateghi R. A data-driven approach to assessing supply inadequacy risks due to climate-induced shifts in electricity demand. Risk Anal 2019;39(3):673–94.
[5] 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.
[6] Mukhopadhyay S, Nateghi R. Estimating climate—demand nexus to support long-term adequacy planning in the energy sector. In: 2017 IEEE power & energy society general meeting. IEEE; 2017. p. 1–5.
[7] Raymond L, Gotham D, McClain W, Mukherjee S, Nateghi R, Preckel PV, et al. Projected climate change impacts on indiana's energy demand and supply. Clim Change 2018:1–15.
[8] Lokhandwala M, Nateghi R. Leveraging advanced predictive analytics to assess commercial cooling load in the US. Sustain Product Consump 2018;14:66–81.
[9] Parkpoom S, Harrison GP. Analyzing the impact of climate change on future electricity demand in Thailand. IEEE Trans Power Syst 2008;23(3):1441–8. https://doi.org/10.1109/TPWRS.2008.922254.
[10] Sailor DJ, Muñoz JR. Sensitivity of electricity and natural gas consumption to climate in the USA—methodology and results for eight states. Energy 1997;22(10):987–98.
[11] 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(1–2):175–201.
[12] Arab A, Khodaei A, Han Z, Khator SK. Proactive recovery of electric power assets for resiliency enhancement. IEEE Access 2015;3:99–109. https://doi.org/10.1109/ACCESS.2015.2404215.
[13] Ching-Lai Hor SJ, Watson S Majithia. Analyzing the impact of weather variables on monthly electricity demand. IEEE Trans Power Syst 2005;20(4):2078–85. https://doi.org/10.1109/TPWRS.2005.857397.
[14] Ji Z, Huang X. Day-ahead schedule and equilibrium for the coupled electricity and natural gas markets. IEEE Access 2018;6:27530–40.
[15] Warren HE, LeDuc SK. Impact of climate on energy sector in economic analysis. J

Appl Meteorol 1981;20(12):1431–9.
[16] EIA. New York quick facts; 2018 < https://www.eia.gov/state/?sid=NY > .
[17] Sailor DJ. Relating residential and commercial sector electricity loads to climate—evaluating state level sensitivities and vulnerabilities. Energy 2001;26(7):645–57.
[18] Sullivan P, Colman J, Kalendra E. Predicting the response of electricity load to climate change, Tech. rep., National Renewable Energy Lab. (NREL). Golden (CO, United States); 2015.
[19] Mirasgedis S, Sarafidis Y, Georgopoulou E, Kotroni V, Lagouvardos K, Lalas D. Modeling framework for estimating impacts of climate change on electricity demand at regional level: case of Greece. Energy Convers Manage 2007;48(5):1737–50.
[20] Lam JC. Climatic and economic influences on residential electricity consumption. Energy Convers Manage 1998;39(7):623–9.
[21] Ahsan Q, et al. A methodology for identification of weather sensitive component of electrical load using empirical mode decomposition technique. Energy Power Eng 2013;5(04):293.
[22] Ruth M, Lin A-C. Regional energy demand and adaptations to climate change: methodology and application to the state of maryland, usa. Energy Pol 2006;34(17):2820–33.
[23] IPCC. Climate change 2013: the physical science basis. Tech. rep.; 2013.
[24] EIA. Form EIA-861M (formerly EIA-826) detailed data; 2019 < https://www.eia.gov/electricity/data/eia861m/ > .
[25] EIA, Natural gas consumption by end use; 2019 < https://www.eia.gov/dnav/ng/ng_cons_sum_dcu_SNY_m.htm > .
[26] NOAA. Global summary of the day (GSOD); 2019 < https://www.ncdc.noaa.gov/data-access/quick-links#gsod > .
[27] USDL. Bureau of labor statistics: data tools; 2018 < http://www.bls.gov/data/ > .
[28] NYC Visitor Guide. Weather facts New York; 2019 < https://www.nyc.com/visitor_guide/weather_facts.75835/ > .
[29] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning vol. 112. Springer; 2013. Ch. 2, pp. 15–58.
[30] Wheelwright S, Makridakis S, Hyndman RJ. Forecasting: methods and applications. John Wiley & Sons; 1998. Ch. 6, pp. 239–310.
[31] Vaughan TS, Berry KE. Using monte carlo techniques to demonstrate the meaning and implications of multicollinearity. J Statist Educ 2005;13(1).
[32] Shmueli G, et al. To explain or to predict? Statist Sci 2010;25(3):289–310.
[33] Miller PJ, Lubke GH, McArtor DB, Bergeman C. Finding structure in data using multivariate tree boosting. Psychol Meth 2016;21(4):583.
[34] Obringer R, Kumar R, Nateghi R. Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States. Appl Energy 252. doi:https://doi.org/10.1016/j.apenergy.2019.113466.
[35] Nateghi R. Multi-dimensional infrastructure resilience modeling: an application to hurricane-prone electric power distribution systems. IEEE Access 2018;6:13478–89.
[36] Obringer R, Nateghi R. Multivariate modeling for sustainable and resilient infrastructure systems and communities. In: Romeijn H, Schaefer A, Thomas R, editors. Proceedings of the 2019 IISE annual conference; 2019. arXiv:1905.05803.
[37] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction, springer series in statistics; 2009.
[38] Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms, in. Proceedings of the 23rd international conference on machine learning, ACM. 2006. p. 161–8.
[39] Nateghi R, Guikema SD, Quiring SM. Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes. Risk Anal: Int J 2011;31(12):1897–906.
[40] Obringer R, Nateghi R. Predicting urban reservoir levels using statistical learning techniques. Scient Rep 8 (1), doi:https://doi.org/10.1038/s41598-018-23509-w.
[41] Nateghi R. Modeling hurricane activity in the Atlantic Basin and reliability of power distribution systems impacted by hurricanes in the US. The Johns Hopkins University; 2012.
[42] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning, vol. 1. Springer series in statistics. New York; 2001.
[43] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Statist 2001:1189–232.