THE GAME CHANGER: ANALYTICAL METHODS FOR ENERGY DEMAND PREDICTION UNDER CLIMATE CHANGE

by

Debora Maia Silva

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Division of Environmental and Ecological Engineering

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Roshanak Nateghi, Chair

School of Industrial Engineering

Dr. Hua Cai

School of Industrial Engineering

Environmental and Ecological Engineering

Dr. Rohini Kumar

Helmholtz Centre for Environmental Research

Dr. Tom Logan

Civil & Natural Resources Engineering, University of Canterbury

Dr. Shweta Singh

Agricultural and Biological Engineering

Environmental and Ecological Engineering

Approved by:

Dr. John A. Howarter

ACKNOWLEDGMENTS

I saw something interesting today on social media (I know, it shocked me too). It was a video of a woman, probably in her thirties, in an environment that was clearly an underdeveloped country, showing impressive skills in freestyle soccer. She wore flipflops and a long skirt. The video caption:

"Talent is evenly distributed.

Opportunity isn't."

As a soccer fan, I thought the video was amazing. But as I had to sit down and write the acknowledgement section of this thesis, I also realized that I am able to be in this position today because of the opportunities I had; and that I was only gifted those opportunities because certain people, for whatever reason and at precise times, decided to believe in me and trust me, even when they had little to no evidence to guide that decision.

So here I want to acknowledge the people in my life who gave me the many opportunities that brought me here.

First and foremost, I would like to thank God for my life and making it possible to meet the friends who helped me in this adventure. Secondly, I would like to thank my family, who, even unwillingly, let a 17-year-old pursue college three states away that back then was just a dream – but it was the first step of this journey. And they stood by my side at every step of it. Geraldo, Keyla, Davi (and now Gabi and Levi!): *thank you.*

My dear friends from undergrad, the same group that would get together on Monday nights to watch a day-old episode of Game of Thrones after statistics class while eating the cheapest pizza on God's green earth. Augusto, Well, André, Henrique: *thank you.*

My first advisor, Professor Luiz, who so kindly asked me one day if my English skills were any good, because he had this friend in INDIANA who might be interested in a student like me. Now I realize he had a plan all along.

And of course, said friend in Indiana, who ended up being the best advisor any grad student could have. Seriously guys, I win. Professor Roshanak Nateghi, aka Roshi, was the best advisor I could imagine. I hope the department keeps giving her prizes because she deserves them. *Thank you*, Roshi.

To the professors and members of my committee, thank you for your advice and guidance. Rohini, Tom, Hua Cai, Shweta Singh. Hopefully, we will meet in a coffee shop one day and talk about the weather.

To my friends at Purdue and all the coffee we shared together; Danielle, Jess, Ellen, Sayanti, Renee, Ben, Min Soo, Kendrick, Yue and the rest of the gang, *thank you*. To the Brazilian lost souls and all the beer we shared together; Jose, Danilo, Victor, Andres, Claudio, and others: *thank you* guys.

And to the local boy (and his marble-sculpted-like family) who once I've heard someone describe as "you were a light in the darkness", I also send my thanks. *Thank you*, Jon.

Finally, a thank you to someone who will not understand it and completely ignore this dedication, but she was the best companion in a year of solitude and helped me stay sane: Miss Quiet.

Let us be grateful for all the opportunities.

And also go for them, because "luck is when opportunity meets preparedness."

I am very lucky.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Accurate prediction of electricity demand is a critical step in balancing the grid. Many factors influence electricity demand. Among these factors, climate variability has been the most pressing one in recent times, challenging the resilient operation of the grid, especially during climatic extremes. In this dissertation, fundamental challenges related to accurate characterization of the climate-energy nexus are presented in Chapters 2–4, as described below.

Chapter 2 explores the cost of neglecting the role of humidity in predicting summer-time residential electricity consumption. Analysis of electricity demand in the CONUS region demonstrates that even though surface temperature—the most widely used metric for characterising heat stress—is an important factor, it is not sufficient for accurately characterizing cooling demand. The chapter proceeds to show significant underestimations of the climate sensitivity of demand, both in the observational space as well as under climate change. Specifically, the analysis reveals underestimations as high as 10-15% across CONUS, especially in high energy consuming states such as California and Texas.

Chapter 3 takes a critical look at one of the most widely used metrics, namely, the Cooling Degree Days (CDD), often calculated with an arbitrary set point temperature of 65°F or 18.3°C, ignoring possible variations due to different patterns of electricity consumption across different regions and climate zones. In this chapter, updated values are derived based on historical electricity consumption data across the country at the state level. Chapter 3 analysis demonstrates significant variation, as high as ±25%, between derived set point variables and the conventional value of 65°F. Moreover, the CDD calculation is extended to account for the role of humidity, in the light of lessons learnt in the previous chapter. Our results reveal that under climate change scenarios, the air-temperature based CDD underestimates thermal comfort by as much as $\sim$ 22%.

The predictive analytics conducted in Chapter 2 and Chapter 3 revealed a significant challenge in characterizing the climate-demand nexuses: the ability to capture the variability at the upper tails. Chapter 4 explores this specific challenge, with the specific goal of developing an algorithm to increase prediction accuracy at the higher quantiles of the de-

mand distributions. Specifically, Chapter 4 presents a data-centric approach at the utility level (as opposed to the state-level analyses in the previous chapters), focusing on high-energy consuming states of California and Texas. The developed algorithm shows a general improvement of 7% in the mean prediction accuracy and an improvement of 15% for the 90th quantile predictions.

1. INTRODUCTION

The development and centralization of energy production brought invaluable progress to human societies, along with new challenges. Among these challenges is the need for accurate demand prediction to balance the grid in real time [1]–[3]. Energy storage is still not available at large scale, making the balance between production and demand a delicate dance that planners need to partake in regularly [4]. Demand forecasting is an integral component in grid adequacy planning, with different methods and techniques developed to this objective, including regression analysis, various evaluation benchmarks, and data descriptive models [5], [6]. And for a long time, these models worked. However, in the last decades, the consistently intricate game of demand prediction has seen the introduction of new factors that muddle the waters of established techniques in electricity demand prediction [7]. Between the new factors challenging the fine art of demand prediction, climate change is undoubtedly a heavy champion [8]–[10].

It should be pointed out that intermittent renewable energy generation technologies such as wind and solar also introduced significant uncertainties in energy demand forecasting [11]–[13]. What pushed the rapid development of such technologies is indeed the very factor that changed the game of demand prediction: climate change. It is climate change mitigation that inspired the development of green energy technology and policies, creating the road toward decarbonizing the energy sector [14], [15]. Climate change also poses significant challenges to the cyber-physical energy infrastructure.

Different from abrupt fluctuations in energy demand like those from demographic shifts, new policy implementations, and penetration of distributed renewable energy generation technology, climate change has been slowly poking at the prediction accuracy of energy demand to the point of making some of the widely used practices obsolete. The change in climate is a slow process, but a change nonetheless, and its consequences are taking their toll on energy planning [16]. When the predictions are inaccurate, the consequences are felt throughout the society, even if undemocratically: increasing blackouts, outages, price surges and, ultimately, death [17]–[22].

While there is no miraculous methodology that could bridge the climate-related gaps that energy planning has been ignoring for decades, this dissertation presents essential topics in energy-climate nexus modeling and proposes data-driven methods to increase the accuracy of predicting the climate sensitive portion of energy demand. This introductory section paints the big picture in how such an important infrastructure in American society is showing problematic signs, how the climate-energy nexus is essential in the shifting paradigm for prediction, and introduces the summary of the published (and yet-to-be-published) research papers by the author.

1.1 Climate Change and Extreme Events

An enormous climate change red flag is undoubtedly extreme climate events. In the last 5 years (2016-2020), the losses of these hazardous events cost $606.9 billion dollars and claimed 3,969 lives in the U.S., an uptake of 2.2 times the total cost of the 1990s [23]. Even though climate change was a slow beast to awaken—building its strength in narrowing cycles of development—once fully summoned, the consequences are devastating. Heat waves, a usually less popular climate change apocalyptic cavalier, along with other extreme climate events have been increasing unprecedentedly in frequency and intensity, with record-breaking losses and estimations of 600-1300 yearly deaths [24]–[26]. Heat waves are not an exclusively American phenomenon, resulting in a high mortality rate in Europe among the elderly, breaking record temperatures, and leaving tens of thousands of deaths behind [27]–[29]. This specific characteristic of heat waves, the high mortality among the elderly, increases the concern in the aging U.S. population and their growing need for residential energy to maintain heat comfort [30].

With the summer of 2020 being one of the dreary examples, heat waves are the main drivers for outages that left millions of Americans without energy on the West coast, resulting in health hazards and life losses [31], [32]. Climate change spearheads the reasons why extreme weather events are challenging its own definition of 'unusual, rare climatic event' with escalating appearances [24], [25], [33], [34]. Heat waves have migrated from black swans events to constant presence in summer-time news, with lasting consequences endangering public health and economy. Events such as the increasing occurrence of heat waves and the consequent large adoption of air conditioning take a toll on energy demand, with special attention during the summer, when the cooling demand to counterbalance heat stress soars [17], [35], [36]. Consequentially, energy planners have the necessity to predict such events or at least be prepared when cooling demand spikes.

However, the aging, centralized electric grid was not designed to sustain such high peak demands, or even worse, it requires a complete transformation of the entire infrastructure to sustain the relatively short periods of heat waves when compared to the life span of the entire energy system [1]. Suggestions of energy production decentralization and creation of micro-grids to increase the resilience of the system for extreme events are common in literature, but usually without clear guidelines for implementation [1], [37], [38]. When the system fails, outages are the most feared outcome, with the aforementioned consequences to society. This increase in failures, that have gone beyond closed meeting rooms at energy utility HQs [39], reaching the most vulnerable population—such as the elderly and children [19], [40], [41]—caught once more the attention spotlight of research. Not unrelated, the development of new, data-driven statistical and machine learning methods for prediction brought research back into those same meeting rooms that have been using outdated methods for decades [3], [42]–[44]. This study is just one of the many in this rich niche of research to raise a flag amidst the grid failures that yes, for energy prediction, the game has changed [45]–[47].

1.2 Energy Scope

Similar to other infrastructures in society, many factors influence the consumption of energy power. Socio-economic changes, demographic effects, technological advances, and public policies are examples of other important players that collectively construct final energy demand [2], [3], [47]. Climate is one of the most important factors, and it is essentially different from other factors mainly in its gradual, cumulative effect on energy demand. To analyze the impacts of climate change in energy demand, it is necessary to separate those effects in the final demand to perform an exclusive, climate-energy analysis [48]. The focus of this thesis is solely on the interaction of climate on energy prediction, even though it is recognized the strength of other factors and, in Chapter 4, control variables that represent economic factors are included. Following this line of thought, energy demand had to be selected also for this study. While different sectors form total energy demand, the residential sector is regarded as the most climate-sensitive sector [46], [49]–[51]. This justification is intuitive, as the residential sector, dominated by the purpose of maintaining human comfort, better reflects season variation related to heat stress. This high sensitivity is emphasized during summer, when the increase in air conditioning penetration has a clear impact 0n residential energy demand for cooling [52]. However, it's important to notice that the developed methods can be applied to the other energy sectors such as commercial and industrial.

1.3 Peak Load

As with any other complex problem, energy prediction has different levels of complexity. While predicting energy demand for a comfortable, light-jacket kind of day in mid-April might be an easy task, predicting the highest load demand of a smoldering summer day in the middle of a heat wave is considerably challenging [47], [53]. Nonetheless, still following the line of thought of complex puzzles, this exact hardest challenge of prediction—the highest demand in a certain period of time—dictates the guidelines for designing and maintaining energy generation, transmission, and distribution [3], [47], [54]–[56]. The name of this obstacle: peak load.

While predicting the central tendency (i.e., mean and median) has dominated the methodology in statistical and machine learning prediction [57], catching the highest demand (peak load) and higher quantiles of demand is the holy grail of energy planners since the system is designed to sustain that specific value [53], [58]. Curiously, it is not always the main topic of research [59], but it dominates Chapter 4 of the present dissertation. The importance of predicting beyond the average and accurately forecasting peak load has to be brought to attention since it dictates energy planning design [3], [47], [54]–[56]. For other areas, this might sound strange, as to why there is the need to adapt or transform an entire structure based on a single day or season of high demand, but the already exemplified societal costs enumerate the explanation of this importance. One might suggest that the problem may be solved with accurate prediction of peak load—which has been increasing under the perennial influence of climate change [36], [60]—but a true solution would be to transform the entire system itself, breaking the developed centralized grid into independent, flexible, and adaptive micro-grids, increasing the resilience of the system [37], [38]. This is a wonderful idea with a dreary implementation, with uncountable layers of both technical and bureaucratic impediments [1], [61]. For these reasons, the idea of increased resilience with micro-grids was not pursued in this dissertation, and instead, a new data-driven methodology is presented for peak load prediction (Chapter 4).

1.4 Everything Under the Sun

To describe and analyze the climate-energy nexus, understanding how to quantify climate effects is essential. Different climate measures have been used for this purpose, with surface temperature being in the genesis of many indexes and measurements, such as Cooling Degree Days (CDD), Heating Degree Days (HDD), and others [3], [47], [54], [55], [62], [63]. Even though air temperature's role in heat stress is significant, it has been shown that air temperature by itself lacks the holistic information of human perception of heat [40], [41], [64], [65]. Heat stress does not rely on a single climate quota and refers to a combination of measures to quantify human sensibility to heat, an amalgamation of both temperature and humidity that captures the concept of human discomfort and their capacity to dissipate heat [19], [40], [66]. While this has been a discussed topic, the majority of the works in the climate-demand literature are based solely on air temperature or measures directly derived from air temperature, failing to include the complexity of human heat comfort, which should always take humidity equally into account [3], [47], [54], [55]. However, there is no consensus of which measures of either air temperature or humidity is the most effective for accurate modeling, leading to a multitude of studies with different measures that, sometimes, are chosen simply because it was either the available measure or the most commonly used [67]. In Chapter 2, a study specifically on determining the key factors for heat stress in 48 American states is shown and idea that there are a multitude of climate factors to be explored is introduced. Throughout this dissertation, none of the selection for the climate variables was performed without careful analysis. In Chapter 3, a deep analysis of one of the most wildly used measures, CDD, is presented in detail, with new approaches to the traditional method.

1.5 Summary of Included Papers

1.5.1 The Critical Role of Humidity in Modeling Summer Electricity Demand Across the United States

As discussed, different climate features are responsible for climate effects on energy demand [67]. The first work included in this dissertation explores this specific gap in literature, with the focus on summer-time energy prediction under climate change. While most of the existing literature and projections are based solely on air temperature and its direct derivatives [40], [41], [64], [65], ignoring the fatal combination of air temperature and humidity exposes the system when under extreme heat events. The work presented in Chapter 2 bridges this gap by identifying the key measures of heat stress throughout the contiguous U.S. (CONUS) on a state level. This work shows that predictions based on surface temperature alone underestimate energy consumption in many of the high energy-consuming states, such as California and Texas, with about 10–15% under both present and future climate scenarios. In conclusion, this work demonstrates that air temperature is an important but insufficient measure to capture human heat stress that reflects cooling demand, showing the necessity of adding selected humidity measures for accurate predictions.

1.5.2 The Goldilocks Zone in Cooling Demand: What Can We Do Better?

The second work presented in this dissertation focuses on exploring a widely used metric for cooling demand prediction, Cooling Degree Days (CDD). As it was shown, the increase in intensity and frequency of heat waves are globally surging cooling demand, and the United States is no exception. One of the main climate measures used for predicting this cooling demand is CDD, based on a predefined value of air temperature [35], [68]–[71]. In this work, insights are extracted through analysis of energy demand and climate features observed data over the CONUS region, demonstrating that current estimations based on the arbitrary CDD fail to capture geographically-specific comfort zones, with deviations of ±25%. Following the study developed in Chapter 2, it has amounted to the evidence that air temperature alone falls short to capture human comfort zones through the calculation of heat index CDD, a climate feature that includes both air temperature and humidity. The results for the inclusion of heat index under the effects of projected climate data (accounting for climate change) show a significant underestimation of cooling demand ($\sim 22\%$) when compared to air temperature alone. Considering the high cooling demand during summer, the insights of this work are relevant for the resilience and security of the electric grid under climate change.

1.5.3 A Data-Centric Approach to Increase Prediction Accuracy at the Upper tails: A Case Study of California and Texas.

The final work in this dissertation focuses on a specific problem in energy demand prediction: peak load prediction. It has been shown that the central tendency in energy demand has a different distribution than the higher quantiles of demand, which include the peak load [59]. Based on this problem that was identified with the exploration of the datasets in earlier chapters, the final chapter develops a data-driven methodology to increase peak load accuracy prediction. It was shown that more frequent and intense climatic events have increased electricity demand, and the increase in peak load is even more dramatic, driven by the high penetration of air conditioning and recurrent heat waves. In this chapter, we show a data-driven approach based on the analysis of historical, daily energy demand data for different utilities in California and Texas, leveraging a state-of-the-art, non-parametric Bayesian ensemble-of-trees model to predict the higher quantiles of demand, i.e., the highest peak loads. Results demonstrate that following the developed methodology, prediction accuracy increased 7% on average in general, with an increase of 15% in the 90th quantile. Since the presented methodology is data-driven, it can be expanded to different algorithms and datasets presenting the similar challenge of inaccuracy in high quantiles.

2. THE CRITICAL ROLE OF HUMIDITY IN MODELING SUMMER ELECTRICITY DEMAND ACROSS THE UNITED STATES

Chapter 2 has been previously published in *Nature Communications*. A post-review version is hosted on https://doi.org/10.1038/s41467-020-15393-8.

2.1 Introduction

Accurate predictions of demand is a key challenge in electricity adequacy planning. Climate, technological, and socioeconomic factors are commonly used in predictive models of electricity demand [72], [73] to ensure reliable planning and operation in the electricity sector by adequately balancing supply and demand. However, more frequent and intense climate extremes such as sustained heat waves [74], [75] cause unanticipated changes in load [76], challenging the reliability of electricity demand predictions. This poses a significant risk to the resilient operation of power systems [77]. Specifically, unanticipated higher demand for space conditioning and refrigeration during heat waves has led to rolling outages with serious socio-economic and public health consequences [78]–[83]. In light of the expected increase in frequency and intensity of climate extremes [74], climate-induced outages are an increasing risk to the resilient operation of the electric power infrastructure. [84]–[88].

Quantifying the climate sensitive portion of the electricity demand, referred to as the climate–demand nexus, hinges on effective characterizations of the key measures of 'heat stress'. Heat stress measures refer to a combination of temperature and humidity that capture human discomfort levels and its ability to efficiently dissipate heat to avoid life-threatening conditions [80], [81], [89], [90]. Heat stress is generally associated with high morbidity and mortality rates during heat waves. Despite the rich literature in electricity demand prediction [73], [91]–[94], little prior work has focused on exploring the climate sensitivity of demand with respect to different measures of heat beyond air temperature. While climate science research establishes air temperature as an incomplete measure of the surface heat content [65], [80], [81], [89], [95], the majority of the existing research on climate-demand nexus use air temperature – or features derived from air temperature such as cooling and heating degree days – as key predictors [73], [96]–[98].

We address this gap by leveraging advanced data analytics to characterize the climate sensitivity of residential electricity demand as a function of a range of heat stress measures beyond air temperature, including dew point temperature, discomfort index, heat index, humidex, wet bulb temperature and wet bulb globe temperature (Methods). These measures are specifically chosen because they have been established in literature as most effective in capturing heat stress [99]. However, no consensus has yet been reached on which measures are most predictive of the climate-demand nexus. We isolate the climate sensitivity of residential energy demand (Methods), which has been established as the most climate sensitive sector [86], [87], [100]–[104]. Thus, our results primarily reflect changes in demand due to climate variability. We focus on summer months – i.e., May, June, July, August – as the majority of the US states peak during summer and climate extremes such as heat waves occur in the summer [105].

The central thesis in this paper is that air temperature alone underestimates the climate-sensitive portion of energy demand for cooling. To test this hypothesis, we first comprehensively assess the role of air temperature in capturing the climate sensitivity of summer electricity load at a U.S. national scale; then we identify the measures of heat stress that are most predictive of the climate-demand nexus and finally we quantify the likely underestimation of models based on air temperature alone under current and future climate scenarios.

We use monthly aggregated, state-level electricity consumption data across the contiguous U.S., extracted from the Energy Information Administration (EIA) public reports over the years of 1990–2016 as well as population data from the U.S. Census Bureau, and harmonized climate data from NCEP North American Regional Reanalysis (NARR) [106]. The electricity consumption data is carefully adjusted to remove trends associated with demographic and technological changes to isolate the climate effects on energy demand [107] (Methods).

To quantify the effectiveness of air temperature alone in explaining the climate sensitivity of residential sector energy load, we develop two sets of models at the state-level: air-temperature-only models, which only use surface air temperature a predictor, and selected-features models, where data-driven variable selection is performed to identify the key measures of heat stress considered in this study. Comparing the results of the air-temperature-only and selected-features models, we determine the influence of each measure of heat stress on the overall climate sensitivity of demand. The predictive models are developed using a state-of-the-art, stochastic, non-parametric Bayesian ensemble-of-trees algorithm[108] (Methods) which has been shown to outperform other climate–demand nexus models in terms of predictive accuracy [72], [85], [86], [101], [109].

We conclude that the air-temperature-only models underestimate residential energy consumption, specially for future climate scenarios. With that, we show that near-surface humidity has an equally important role in electricity prediction as air temperature.

2.2 Methods

2.2.1 Observed electricity consumption data

Monthly electricity consumption data was extracted from the Energy Information Administration (EIA) [110] over the years of 1990–2016 at a state level, with a total of 68 months per state (4 months per year for a total of 17 years). The source file includes electricity consumption data in Megawatt hours (converted to kWh), separated by sector: residential, commercial, industrial, transportation and others. The electricity consumption of the residential sector was the focus of the analysis presented in this paper. The residential electricity consumption data was detrended to isolate the climate effects on energy demand from techno-demographic changes such that increases in energy were not attributable to non-climatic factors such as technology shifts and population increase [87], [107]. The residential energy consumption data in each state was first normalized using state-level population data – from the U.S. Census Bureau [111] – to obtain a per capita electricity demand, and then detrended based on the following steps.

$$E(y) = \frac{\sum_{y=1}^{n_{years}} \sum_{m=1}^{12} E(m,y)}{n_{years}} \tag{2.1}$$

Where the total years, n_{years}, range from 1990-2016; m denotes the month and y the year.

Then, an adjustment factor was calculated for each year as the sum of the per capita consumption per month divided by the yearly consumption $E(y)$. This process was repeated for all states.

$$F_{adj} = E(y)^{-1} \sum_{m=1}^{12} E(m, y) \tag{2.2}$$

Finally, the adjusted energy consumption was calculated by dividing the monthly consumption by the adjustment factor.

$$E(m, y)_{adj} = E(m, y)/F_{adj} \tag{2.3}$$

The detrended monthly electricity consumption data (described above) is referred to as observed electricity consumption data. This is a widely used method for trend-adjusting monthly aggregated regional data. For a comparative assessments of different de-trending methodologies please refer to Bessec and Fouquau[112].

2.2.2 Observed climate data

The observed climate data was acquired at a sub-daily time scale for the period of 1990-2016 from the NCEP North American Regional Reanalysis (NARR) [106], [113], [114], aggregated at a monthly level to match the temporal scale of electricity consumption data, and weighted by population so as to give a greater weight to areas with higher population density when aggregating to the state level. The input climate data include air temperature, dew point temperature, discomfort index, heat index, humidex, wet bulb temperature and wet bulb globe temperature. All climate variables use a combination of temperature, humidity and pressure.

Dew Point refers to the temperature which air is saturated with water vapor, calculated in this paper using the equation below:

$$t_d \approx t - \left(\frac{100 - RH}{5}\right) \tag{2.4}$$

Where t_d is the dew point temperature and t is air temperature in Celsius, and RH stands for Relative Humidity in percent (same RH as in the other equations in this section)[115].

Wet Bulb Temperature is the lowest temperature to which air can be cooled by water evaporation at a constant pressure. In this paper, it follows the equation:

$$T_W = T_W - \frac{f(T_w; \boldsymbol{\pi}) - (C/T_E)^{\lambda}}{f'(T_W; \boldsymbol{\pi})} \tag{2.5}$$

Where $\boldsymbol{\pi} = (p/p_0)^{1/\lambda}$ is the Exner function, used for scaling, λ is the inverse of the Poisson constant for dry air (3.504), p pressure (mb), p_0 the reference pressure (1000mb); C is the freezing temperature in Kelvin, and T_E the equivalent temperature, which is the moist potential temperature scaled by λ. For reviewing the complete derivation of the equation, refer to the Appendix A of [99], [116].

Discomfort Index (DI) was developed in the 1950s to calibrate air conditioners and further adapted by the Israeli Defense Force as a main measure of heat stress.

$$DI = \mathit{0.5}T_W + \mathit{0.5}T_C \tag{2.6}$$

Where T_W stands for Wet Bulb Temperature, RH is percentage Relative Humidity and T_C is the Temperature in Celsius. DI is unitless [99], [117], [118].

Wet Bulb Globe Temperature (sWBGT) is the heat stress measure in direct sunlight.

$$sWBGT = \mathit{0.56}T_C + \frac{\mathit{0.393}\mathrm{e}_{RH}}{\mathit{100}} + \mathit{3.94} \tag{2.7}$$

Where $\mathrm{e}_{RH} = (RH/100)\mathrm{e}_{sPa}$; wet bulb temperature used here is unitless and widely used.

e_{RH} represents the vapor pressure in pascals and is calculated from RH in percentage and saturated vapor pressure [99], [119].

Humidex was developed for the Meteorological Service of Canada. It is a unitless [99], [120] index, aiming to explain what the temperature feels like for the human body.

$$HUMIDEX = T_C + \frac{5}{9}\left(\frac{e_{RH}}{100} - 10\right) \tag{2.8}$$

Heat Index (HI) is also called apparent temperature, describing what the temperature feels like to the human body when relative humidity is combined with air temperature.

$$\begin{aligned} HI = &-42.379 + 2.04901523T_F + 10.14333127RH - 0.22475541T_F RH \\ &-6.83783x10^{-3}T_F^2 - 5.481717x10^{-2}RH^2 + 1.22874x10^{-3}T_F^2 RH \\ &+8.5282x10^{-4}T_F RH^2 - 1.99x10^{-6}T_F^2 RH \end{aligned} \tag{2.9}$$

Where (T_F) denotes the air temperature, and HI are measured in degrees Fahrenheit [99], [121].

2.2.3 Projected climate data

Five different Global Circulation Models (GCM) were used in this paper, namely, Geophysical Fluid Dynamics Laboratory Earth Systems Model (GFDL-ESM2M), Hadley Global Environment Model 2 - Earth System (HadGEM2-ES), IPSL Earth System Model for the 5th IPCC report (IPSL-CM5A-LR) [122], Atmospheric Chemistry Coupled version of MIROC-ESM, a Earth System model (MIROC-ESM-CHEM), and the Norwegian Earth System Model (NorESM1-M). The data is considered under the emission scenario that has an end-of-century radiative forcing equal to 8.5 Wm^{-2}; a Representative Concentration Pathway that is characterized by high greenhouse emission levels (RCP8.5) [86], [123]. This global data is approximated by a 0.5 degree spatial resolution ($\sim$ 50 km) [124] and is processed from 1950-2099. The same variables were calculated for the projected data, namely, dew point temperature, discomfort index, heat index, humidex, wet bulb temperature and wet bulb globe temperature. Two periods were extracted from the projected data, namely, the reference period of 1991-2010, and the future period of 2031-2050.

There were two distinct stages in model development and analysis, namely, predictive model development using the observed data and sensitivity analysis using the projected data (Supplementary Figure 5). In the predictive model development stage, electricity consumption data, climate data and population data were aggregated and normalized at the state level, as explained earlier, yielding monthly observed data for 48 states. Using these data sets, the development of air-temperature-only models and selected-features models involved three steps: for the selected-features model, variable selection was performed to identify the key predictors of cooling demand (details in the next section: BART algorithm and modeling process), then, the model hyper-parameters were tuned (through cross validation), and finally a ten-fold cross-validation was performed to assess the predictive accuracy of the final models. For the air-temperature-only models, the same process was followed with the exception of the variable selection step, as there was only one independent variable. After developing these statistical models, the results were processed with the collection of in-sample (training set, 90%) and out-of-sample (test set, 10%) R^2 and RMSE (Root Mean Square Error) values. The comparative assessments between the models were conducted based on the out-of-sample values of errors (RMSE results shown on Supplementary Figure 2 and 3 and R^2 results shown on Supplementary Figures 6 and 7).

In the sensitivity analysis stage, projected climate data from the five GCMs over the reference period (1991-2010) and future period (2031-2050) were used as inputs to the previously developed predictive models. This generated four sets of data, namely, air-temperature, reference: estimates from the air-temperature-only model using the reference period data; air-temperature, future: estimates from the air-temperature-only model using the future period; selected-features, reference: estimates from the selected-features model for the reference period; and finally the selected-features, future, estimates from the selected-features model for the future period. This allowed for calculating the delta change over time between the same models (i.e., calculating the difference between selected-features models' over the periods of 2031-2050 and selected 1991-2010). Similar analysis was also conducted on the 90th quantile of data, as electricity inadequacy during extreme heat episodes is typically associated with the tail of the demand distribution.

2.2.4 BART algorithm and modeling process

We leveraged a non-parametric Bayesian ensemble-of-trees algorithm to characterize the climate sensitivity of electricity consumption, since the algorithm was shown to outperform other climate-demand nexus models in terms of predictive accuracy [72], [85], [86], [101], [108], [109]. The independent variables in the development of the BART (Bayesian Additive Regression Trees) models were heat stress measures, and the response variable was state-level electricity consumption. Models were trained and tested using a 10-fold cross-validation technique [125].

The non-parametric Bayesian sum-of-trees model can be represented using the equation below [108], [126],

$$\mathrm{Y} = \sum_{\mathrm{j}=1}^{m} g(x; T_{\mathrm{j}}, M_{\mathrm{j}}) + \varepsilon, \varepsilon \sim N(0, \sigma^2) \tag{2.10}$$

Here, Y denotes the response variable, g denotes the regression tree function, T_{j} denotes individual tree structures, M_{j} denotes a set of parameters (e.g., mean values) associated with the tree nodes. BART is the sum of m total trees. The decision rules associated with each tree involve binary splitting of the predictor space. So for a given T and M, $g(x; T, M)$ represents a tree with parameters M assigning the binary splits for the terminal nodes of T. ε is the stochastic irreducible error, assumed to be normally distributed with a constant variance chipman$_bart:_2$ 010.

The BART algorithm uses a prior over the sum-of-trees parameters to regularize the trees, i.e., to weaken the effect of individual trees with the goal for better prediction when adding all decision trees. A prior is imposed on all parameters, $(T_{\mathrm{j}}, M_{\mathrm{j}})$ $[\forall \mathrm{j} = 1, ..., m]$ and σ. This regularization is made to keep individual trees effects from being unduly influential over the entire sum-of-trees. Without a regularization, large tree components would overwhelm the sum-of-trees and would diminish the benefits of leveraging additive tree structures. The prior regularization can be mathematically represented as:

$$\mathrm{p}(M_{\mathrm{j}}|T_{\mathrm{j}}) = \prod_{\mathrm{i}}^{m} p(\mu_{\mathrm{ij}}|T_{\mathrm{j}}), \quad where \quad \mu_{\mathrm{ij}} \in M_{\mathrm{j}} \tag{2.11}$$

Where tree components T_j and M_j are assumed to be independent of each other and of σ.

As mentioned earlier, to assess the role of air temperature in capturing the climate sensitivity of electricity consumption compared to other measures of heat stress, two models were developed for each state: the air-temperature-only model, and the selected-features model. Air-temperature-only models were constructed with a single input variable of surface temperature and electric demand as the response variable. To develop the selected features a data-driven variable selection algorithm was harnessed. Variable selection: Specifically, the variable selection algorithm involved identifying the variables most frequently used as decision/splitting rules in the sum-of-tree model [108]. The calculation was then averaged over 100 model-iterations to achieve stable estimates throughout the sum-of-trees distribution. This is a reasonably robust approach for variable selection using the Bayesian ensemble-of-trees algorithm **blaikapelner_bartmachine:_2013**, [108].

2.2.5 Pseudo-Code

Feature Selection is performed for 48 states before running BART. Code performed in parallel cores.

Algorithm 1 Bart Machine Model Fit

for i $\in \{1 : 48\}$ **do**

    for $k \in \{1 : 20\}$ **do**

        $BartMachine_k, RMSE_k, R^2_k$

    end for

    $RMSE_i = mean(RMSE_k) \sim k \in 1 : 20$

$R^2_i = mean(R^2_k) \sim k \in 1 : 20$

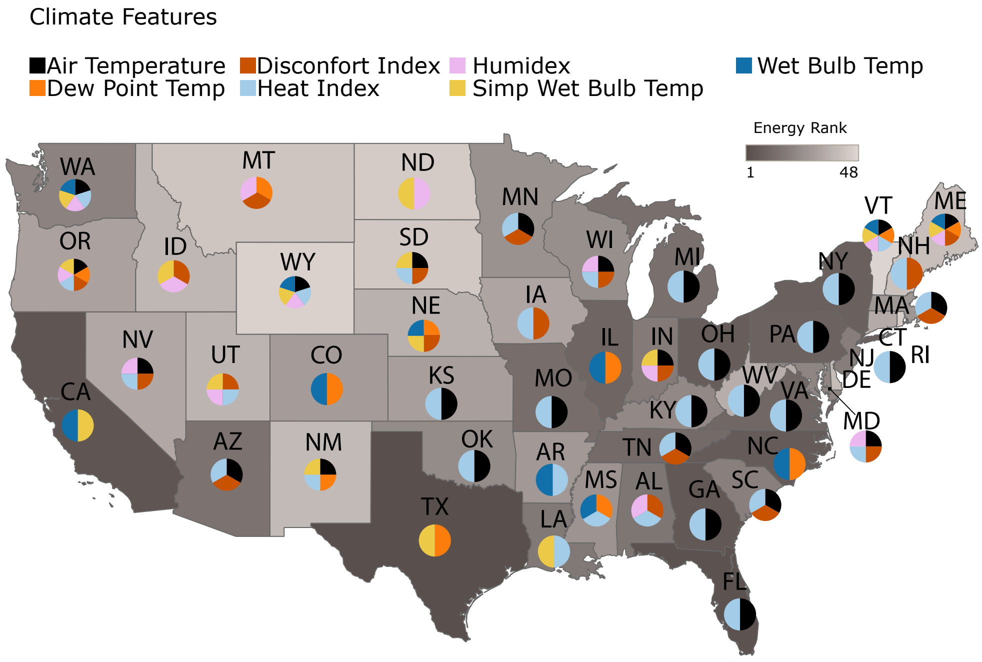
end for

2.3 Results

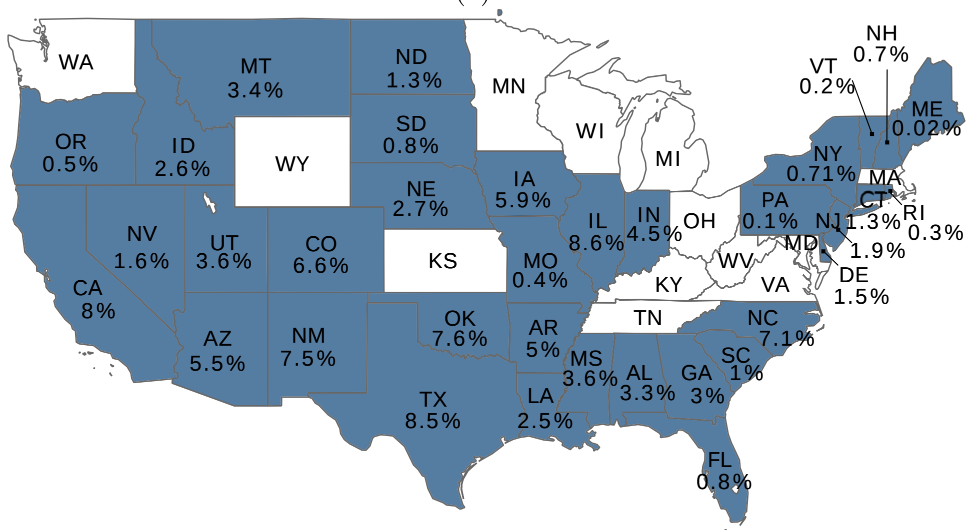
2.3.1 Climate features identification

Our results reveal substantial improvements in prediction accuracy in 35 states for the cooling demand as a function of both temperature and humidity over the air-temperature-only model. Figure 1a illustrates the key predictors of the climate-sensitive portion of electricity demand across the 48 US states as identified by the selected-features model. A darker shade represents states with higher energy consumption and the pie charts depict the selected climate variables for each state. We find significant variability in the identified key measures of heat stress across the United States, illustrating the importance of conducting region-specific analysis in characterizing the climate–energy demand nexus. The most energy-intensive states such as California and Texas, as well as many of the summer-peaking southern states (Supplementary Figure 1 and Supplementary Note 1), benefit from the inclusion of measures of heat stress beyond air temperature – such as dew point temperature, wet bulb temperature, wet bulb globe temperature and heat index (Methods) – in estimating the climate sensitivity of demand. This underscores the importance of accounting for near-surface humidity combined with temperature to better model human comfort levels and response to heat which translates to residential cooling demand.

The highlighted states in Figure 1b represent the extent in which the selected-features models outperformed air-temperature-only models in terms of predictive accuracy. Comparing the air-temperature-only models and selected-features models reveal that 35 states benefit from including measures of heat stress beyond air temperature, based on the comparison of Root Mean Square Error (RMSE) improvement in prediction accuracy (Figure 1b). In addition to comparing model performances for the average consumption levels (i.e., 50th quantile predictions, Figure 1b), we also compare the 90th quantile predictions (Supplementary Figure 2 and 3). In this way, the effects of extreme heat events can be deduced from the higher quantiles of the aggregated electricity demand data. In other words, since heat waves typically coincide with the energy peaking summer months, the results for the 90th quantile RMSE improvement (Supplementary Figure 3) reflect months with more intense and frequent heat events. We observe significant model improvements, in terms of out-of-sample RMSE, for a number of high-energy consuming states of the order of 8% (7.7%) in California (CA), 8.5% (21.1%) in Texas (TX), 8.6% (26.1%) in Illinois (IL) and 7.1%(9%) in North Carolina (NC), for the 50th quantile predictions (and 90th quantiles predictions-Supplementary Figure 3). The observed improvements are substantial. For instance, in a state such as Texas, the selected-features model's average improvement of 8.5% in a given summer month such as August 2016 is equivalent to 1,498,968 MWh of energy consumption, which could sustain the residents of Austin-TX for more than four months [111], [127]–[129]. In the same year and month, the 8% median increase in the state of California could sustain almost 1.5 million of Californians households [130].



(a)

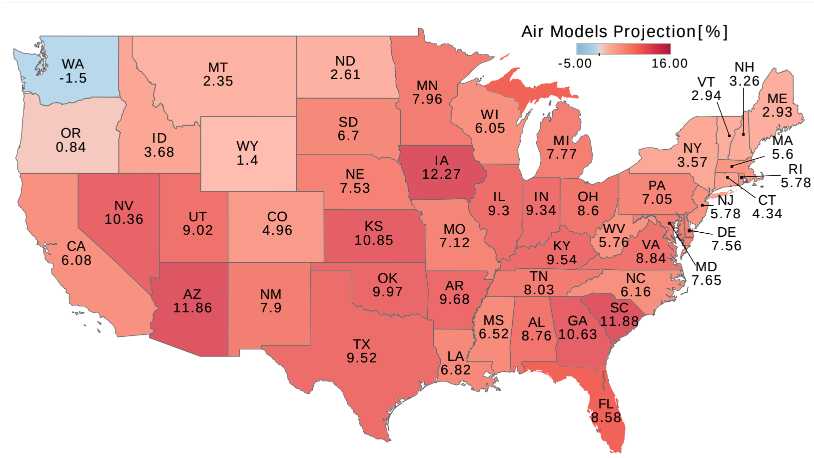


(b)

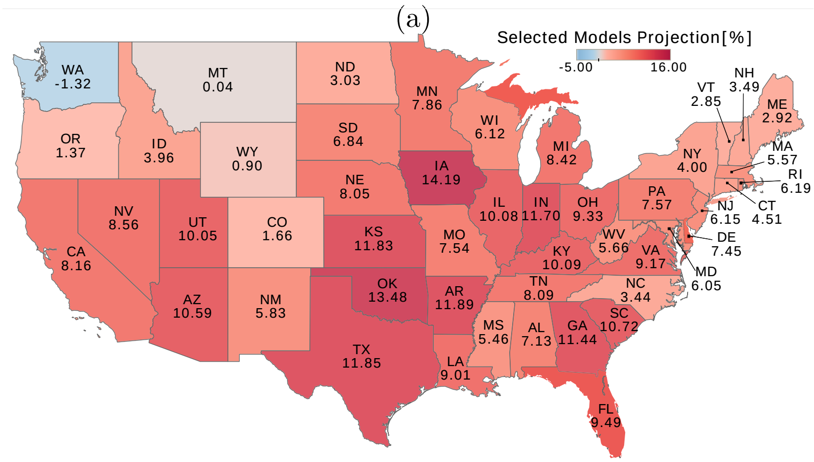
Figure 2.1. a States are shaded according to total energy consumption intensity (scale bar goes from 1 to 48, with darker tones representing higher intensity states in terms of energy consumption). The pie charts illustrate the state-level key measures of heat stress for predicting the climate-sensitive portion of electricity demand. The majority of states have up to two most important features. It is important to point out that climate sensitivity is not an exclusively geographical characteristic, and considering the population density weight method for the variables calculation (Methods), it is possible to expect neighboring states to not necessarily have the same selected climate variables. Non-summer peaking states with many selected features (e.g., Washington (WA) and Oregon (OR)) reveal relative climate-insensitivity in comparison (Supplementary Note 1). Note that the top energy consuming states of California (CA) and Texas (TX) do not include air temperature as the key predictors of the climate-sensitive portion of demand. b Highlighted states show percentage improvement in predictive accuracy when comparing the selected-features models to the air-temperature-only models (comparison based on out-of-sample RMSE).

2.3.2 Sensitivity analysis under future climate scenarios

To quantify the extent to which the air-temperature-only models underestimate the climate sensitivity of demand under different climate scenarios, we use projected climate data extracted from five widely used Global Circulation Models (GCM).



(a)



(b)

Figure 2.2. Increase in the climate-sensitive portion of the demand under climate change (scale bars go from -5 to 16 percent) based on: (a) the air-temperature-only models, and (b) the selected-features models. Note the significantly higher demand increases for the selected-features models (32 states showed higher values for the selected-features models), supporting the initial hypothesis that the air-temperature based projections underestimate the climate sensitivity of demand.

The GCM-extracted climate data are used to train state-level ‘air-temperature-only and selected-features models to predict for the reference period of 1991-2010, and the future period of 2031-2050 (Methods). Since the objective here is to evaluate the changes in demand pattern as a result of future climate change (i.e., to quantify the relative increase over the periods, as opposed to absolute values), reference period data from the GCM is used instead of the historical data to remove bias induced by the use of data from different sources in the sensitivity analysis. In other words, input data generated from the same source (i.e., the GCMs) over the two time periods of reference (1991-2010) and future (2031-2050) are used in the developed predictive framework to track the relative increase in the climate-sensitive portion of demand as estimated by the air-temperature-only versus selected-features models.

The 1991-2010 reference period was chosen to match the period of the observed data as closely as possible and, at the same time, the same length as the future period (i.e., 20 years). While the common practice in climate science research is making projections until 2100, that timeline is not appropriate in the energy sector, given the life span of the existing energy infrastructure [131]–[133]. The future period of 2031-2050 was chosen since energy infrastructure planning has the 2050 year as a comparable reference target in most energy projection and planning reports [134], [135].

The selected-features models project a significantly higher increase in the climate sensitive portion of demand over time compared to the air-temperature-only models, when comparing the relative increase of each model (Figure 2). The projections based on the selected-features models for the high-energy consuming states of Texas, California and Florida show a relative demand increase of about 12%, 8% and 10%, respectively, over the reference period (Figure 2).

Figure 3 depicts the relative increase of projected demand by the selected-features model over the air-temperature-only model for the average consumption levels and 90th quantile predictions for the top seven energy-consuming states that presented an improvement in prediction accuracy in the historical period. Figure 3 enables the comparison between the ratios of the projections based on the selected-features over the air-temperature models for 90th percentile as well as the average values, as shown in the equation below. Complete results for all states are shown on Supplementary Figure 4.

$$\frac{\text{Selected features future}_{2031-2050} - \text{Selected features reference}_{1991-2010}}{\text{Air-temperature-only future}_{2031-2050} - \text{Air-temperature-only reference}_{1991-2010}}$$

This comparison reveals that the underestimation in energy demand attributable to the air-temperature-only models is substantially more significant for the projections of higher demand values that are typically associated with periods of intense heat stress months – where the resilient operation of the grid is most challenged and unexpected demand spikes lead to supply inadequacy and thereby increase the risks of morbidity and mortality.

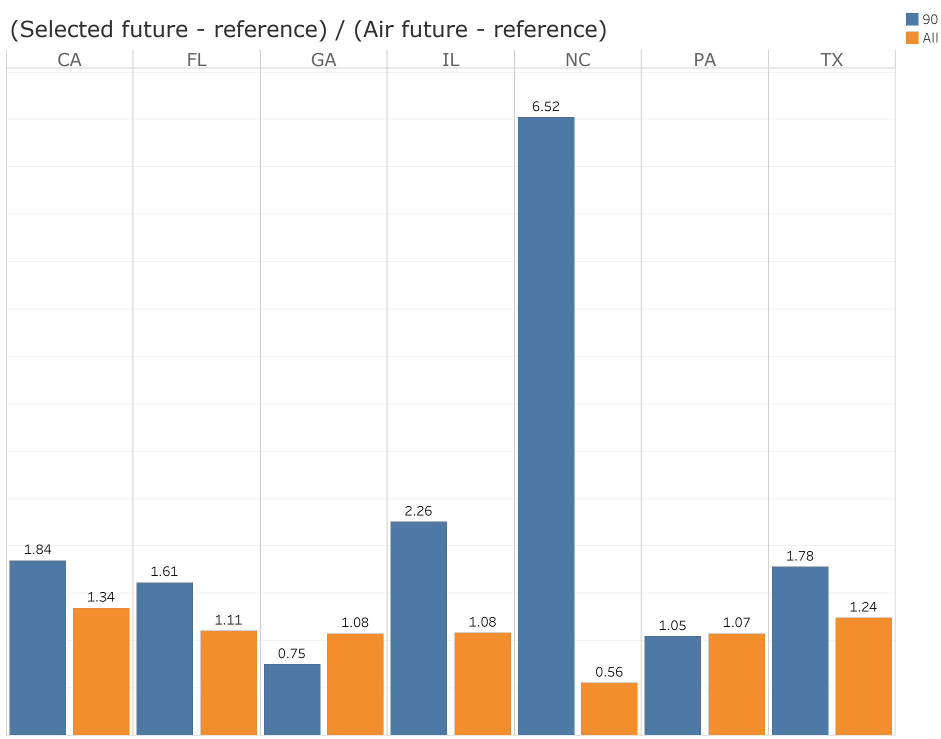


Figure 2.3. Seven out of the top ten energy consuming states in the US that showed improvement with selected-features model. The relative increase of projected demand by the selected-features model over the air-temperature-only model for the average consumption levels and 90th quantile predictions for the top energy-consuming states.

2.4 Discussion

The existing projections of climate-induced demand increase based on rising air temperatures alone ignore the fact that rising temperatures are associated with increased humidity – a lethal combination that increase morbidity and mortality in the absence of adequate cooling capacity during extreme heat events. We show that air temperature is a necessary but not sufficient measure to characterize residential space cooling demand during summer times. Humidity levels are also critical in capturing what truly reflects heat sensation for the human body [80] . Ignoring the role of humidity leads to underestimating the climate sensitivity of demand, challenging the resilient operation of power systems – especially under future warming scenarios where summertime energy production will be further constrained [136]. Inadequate supply to meet rising demand will have significant socio-economic costs due to adverse health effects – particularly among the most vulnerable population – and warrants rapid and costly investments in energy infrastructure expansion as well as adaptation measures [96].

We propose a data-driven framework to quantify the extent of underestimation of the climate-sensitive portion of cooling load, attributable to the use of air temperature alone as a measure of heat stress. Our results reveal a significant increase in predictive accuracy (of the order of 8% for high energy consuming states such as CA and TX) in characterizing the climate-sensitive portion of demand through a more holistic inclusion of measures of heat stress. The results based on projecting demand under future climate scenarios are consistent, in that models based on air temperature alone show a systematic underestimation of the climate-sensitive portion of demand. The underestimation is particularly pronounced at the upper tail of the demand distribution, suggesting that the projected increase would be almost twice as large when considering measures of heat stress beyond air temperature in high energy states such as California, Florida and Texas.

It is important to note that our results pertain to only the climate-sensitive portion of the residential cooling demand. To project changes in total residential cooling demand under future climate scenarios, additional information regarding economic and population growth as well as technology and demographic changes need to be considered. What we wish to highlight here is that, under similar cooling technology, the existing projections based on air-temperature alone substantially underestimate the anticipated demand increase under future warming scenarios which could lead to inadequate and ineffective investments in capacity expansion and demand response programs as well as adaptation measures. Furthermore, while this study is focused on the residential energy demand, the methodology can be extended to other climate-sensitive sectors [137].

3. THE GOLDILOCKS ZONE IN COOLING DEMAND: WHAT AN WE DO BETTER?

Chapter 3 has been submitted to *Earth's Future* and is currently under review.

Plain Language Summary

Hotter summer days and more frequent and intense heat waves are causing a sharp rise in demand for air conditioning across the globe. Accurate estimation of demand for space cooling is an integral component of resilient planning, operation, and management of the grid. One widely used metric for characterizing this demand is the Cooling Degree Days (CDD), which is calculated based on the difference between the daily temperature mean and a pre-defined base temperature that represents a 'comfort zone'. In this paper, we analyze historical data on climate and energy demand, and find that the most frequently used base temperature of 65°F in CDD calculations leads to mis-characterizing geographically specific 'comfort zones' across the U.S. This can cause significant under- or over-estimations of energy cooling demand. Moreover, we extend the CDD calculations to also account for the role of humidity and demonstrate the cost of ignoring humidity in CDD calculations under present and future climate conditions.

3.1 Introduction

The thermal comfort of societies is critical not only for human health and well being but also for achieving a high-sustainability future. Despite the direct linkages between cooling demand and each of the 17 Sustainable Development Goals (SDGs), the unprecedented global increase in demand for cooling has been largely absent from today's sustainability debates [138]. Under current socio-economic and climatic conditions, three-quarters of the global population will experience health risk due to exposure to extreme heat events [139], with significant equity and justice implications. The demand for space cooling is expected to witness a threefold increase by 2050 [140]; the inability to meet this rising demand sustainably is bound to widen the energy poverty gap and increase GHG (greenhouse gas) emissions, exacerbating climate change and its impacts on modern society.

Air conditioning is touted as an integral component of modern living and a testament to human civilization's progress [141]. Moreover, it is an important driver of summer-time peak load—the highest energy demand in a given period—which often sets the key operational and planning parameters in energy infrastructure management [3], [47], [54]–[56]. With increased intensity and frequency of heat waves and accelerated adoption of air conditioning, access to accurate estimates of cooling demand (during both peak and off-peak hours) has become an important pillar in energy systems planning [9], [24], [142], [143]. Accurate characterization of summer-time peak load is particularly important for residential customers, which represent the most climate-sensitive segment of the energy sector [49]–[51], [144], [145].

Cooling Degree Day (CDD) is a practical and widely used measure for quantifying summer-time space cooling demand in energy planning [35], [146]. CDD represents the number of degrees a day's average temperature exceeds a pre-specified set-point temperature, and any value that exceeds this base temperature is assumed to trigger demand for cooling. CDD's set-point temperature represents a comfort zone—aka a 'Goldilocks zone' for human thermal comfort, where it is neither too cold nor too hot. The selected comfort zone temperature is often arbitrarily set at 65°F (18.3°C) in global and regional energy planning studies [35], [52], [68]–[71], [147]. More specifically, while in certain applications such as building-level thermal comfort studies [148] empirically derived base temperatures have been used, in studies related to energy infrastructure planning—the focus of this paper—CDD's set point temperature is almost always set at 65°F (18.3°C) [35], [52], [68], [69], [71].

There are two fundamental caveats to the approaches that calculate CDD based on the generic set-point value of 65°F for sustainability and resilience analytics in energy infrastructure planning and management. Firstly, the set-point value of 65°F was derived decades ago, with no consideration of climate change and thus might no longer be a representative value under present and future climate conditions. Secondly, previous studies have shown that air temperature is a necessary but not sufficient measure of heat stress [36], [67], [149], [150]. However, CDD does not take humidity into account [146], rendering its effectiveness in capturing human thermal comfort questionable. In the light of the recent record-breaking blackouts last summer [31] along with increased frequency and intensity of heatwaves [151], the energy sector must address these shortcomings to mitigate the growing threats of climate change and enhance the security, sustainability, and resilience of the grid. Otherwise, incomplete and inaccurate understandings of how human thermal comfort relates to cooling demand will hamper urgent transformations needed to unlock sustainable pathways, and will likely increase the risk of path-dependent trajectories in the energy sector.

We address these fundamental gaps by deriving geographically specific CDDs and extending the calculation of CDD to also account for humidity. Specifically, we first derive geographically specific CDDs for each state[1], using summer-time (May to September) residential energy consumption data (1990-2016) to establish region-specific optimal set-point temperatures. By comparing these values with CDDs based on 65°F set-point temperature, we assess the divergence in values throughout the American territory. We discuss the implications of the over- or underestimation of the newly calculated CDDs for energy planning under both present and future climate conditions.

Additionally, to account for the critical role of humidity, we go beyond air temperature in calculating CDD. In particular, we extend the CDD method to heat index—a widely used climate measure for human heat comfort that includes humidity [36], [67], [152], [153]—and harness CMIP5-GCM climate scenarios to make projections under climate change. Our results demonstrate a considerable deviation of the optimal set-point temperatures from the base temperature of 65°F (18.3°C) in most states, with an average deviation of 10%. In addition, the projected heat index-based CDDs show a considerable increase in value as compared to air temperature, with an average of 22% higher values. Our findings reveal that a unilateral focus on air temperature-based CDDs with a generic set-point temperature in energy systems planning undermines the resilience of the grid under climate change, especially during extreme heat events.

The structure of the paper is as follows. Details of the data collection, data processing, and methodology are summarized in Section 2. Results are presented in Section 3. The paper concludes by summarizing findings and discussing the significance of results in Section 4.

[1]↑While state boundaries do not always coincide with climate boundaries, our state-level analysis is motivated by providing insights that are relevant to state-level policymakers and energy planners.

3.2 Data and Methods

3.2.1 Observed Climate Data

The observed climate data is acquired at a sub-daily (3 hourly) time scale for the period of 1990-2016 from the NCEP North American Regional Reanalysis (NARR) at a 32 kilometer spacial resolution [106], [113], [114]. Data is aggregated at a monthly level to match the chronological scale of electricity consumption data and weighted by population density when aggregating to the state level. Specifically, the 2010 UN-adjusted Gridded Population of the World dataset (Version 4) is used for this work, collected from the Socioeconomic Data and Applications Center (SEDAC; http://sedac.ciesin.columbia.edu). This procedure to give higher weight where population is concentrated when averaging state level data is in line with previous studies in residential electricity demand [59], [154].

3.2.2 Projected Climate Data

While analyzing observational data is essential for understanding past variability in historical events, they provide limited knowledge for anticipating the future, especially under non-stationary conditions. Using projected climate data is essential for characterizing the growing effects of climate variability and change on the energy sector [36], [47], [155]. To extend our analysis into the future such that our findings are relevant for energy planning, projected climate data are acquired for the period of 2031–2050. This timeline is chosen due to the fact that the year 2050 is consistently used as a target year for mid-term planning in energy reports [156], [157]. This timeline is practical as it allows for considering climate change effects on the sector without having to consider significant transformations to the architecture of the electric grid.

The projected climate data used in this paper are derived from five different Global Circulation Models (GCM), namely, Geophysical Fluid Dynamics Laboratory Earth Systems Model (GFDL-ESM2M), Hadley Global Environment Model 2 - Earth System (HadGEM2-ES), IPSL Earth System Model for the 5th IPCC report (IPSL-CM5A-LR) [157], Atmospheric Chemistry Coupled version of MIROC-ESM, a Earth System model (MIROC-ESM-

CHEM), and the Norwegian Earth System Model (NorESM1-M). The data are considered under the emission scenario that has an end-of-century radiative forcing equal to 8.5 Wm^{-2}—Representative Concentration Pathway that is characterized by high greenhouse emission levels, RCP8.5, [8], [123]. For the state level averaging, data is approximated by a 0.5 degree spatial resolution ($\sim$ 50 km) [124] and weighted by population.

3.2.3 Electricity Data

Similar to the temporal resolution of the observed climate data, monthly electricity data sales data are used in this work. Data are acquired from the U.S. Energy Information Administration [158] over the years of 1990–2016 at a state level for the residential sector. To isolate the climate effects from the electricity data—which are influenced by various factors such as technological changes, policy implementation, demographic shifts, etc. [2], [3], [47]—we de-trended the raw, state-level electricity data. We leveraged a well-established de-trending method for isolating the climate effects [48], which is widely-used in the energy research literature [9], [63], [159]–[162]. Electricity demand data are initially normalized by the state-level population to obtain a per capita value of consumption. The de-trending process involves the following steps:

$$E(y) = \frac{\sum_{y=1}^{n_{years}} \sum_{m=1}^{12} E(m, y)}{n_{years}} \tag{3.1}$$

Where the total years, n_{years}, range from 1990–2016; m denotes the month and y denotes the year.

An adjustment factor is calculated per year, and it is the sum of the monthly per capita demand divided by the yearly consumption $E(y)$.

$$F_{adj} = E(y)^{-1} \sum_{m=1}^{12} E(m, y) \tag{3.2}$$

The final de-trended demand is obtained by dividing the monthly consumption by the previous calculated adjustment factor.

$$E(m, y)_{adj} = E(m, y)/F_{adj} \tag{3.3}$$

3.2.4 CDD Calculation

Once climate and electricity data are aggregated and available, the next step is CDD calculation. Daily CDD is calculated as shown below in Equation 3.4.

$$CDD_{daily} = \begin{cases} 0, & T_d < T_b \\ T_d - T_b, & T_d > T_b \end{cases} \tag{3.4}$$

Where T_d represents daily average temperature and T_b represents the base temperature/set-point temperature selected for the CDD calculation. CDD is usually aggregated to annual or monthly levels by summing the respective daily values.

While T_b is often arbitrarily set to 65°F (18.3°C) [35], [71], we leveraged the well-established energy signature method [48], [62], [163]–[165] to derive geographically specific CDD set-points for all 48 CONUS states. The analysis is done by examining scatter plots of energy consumption versus climate variables to select a vertex that reflect cooling sensitivity, as characterized by a sharp increase in demand at a certain climate threshold value. The energy signature method is performed in three steps:

1. Iteratively process the data to select relevant intervals, conducive to identifying the sensitivity points (or base values/set-points);

2. Fit piece-wise constant regression models to each region.

3. Repeat steps 1 and 2 until distinct vertex points are detected.

Considering the uncertainty associated with this method, confidence intervals with 10,000 bootstrap repetitions are calculated for each base value. At the end of the process, the 48 CONUS states have CDD base values for air temperature and heat index. An example of the energy signature method is illustrated in Figure 4.2.

The geographically-specific CDD based values are then compared against the widely-used 65°F (18.3°C) value. The deviations are spatially illustrated in Section 3. Reduced form equations are then used to characterize the implication of the discrepancies between the derived and widely-used set point temperature of 65°F (18.3°C) in-terms of energy demand.

3.3 Extending the CDD Calculation to Include Humidity

To extend the CDD analysis under climate change to also account for humidity humidity, heat index-CDD was calculated using the Energy Signature method discussed previously for the 2031-2050 time period, as illustrated in Figure 4.2(b) and 4.2(d). Heat index (HI), also called apparent temperature, describes what the temperature feels like to the human body when relative humidity is combined with air temperature [67], [166]. Characterizing the climate-sensitivity of energy demand requires accounting for the synergistic effects of surface temperature and humidity on human body and; accounting for the role of humidity, therefore, is necessary for modeling demand. Heat index is calculated following the equation bellow:

$$\begin{aligned} HI = &-42.379 + 2.04901523\, T_F + 10.14333127\, RH - 0.22475541\, T_F RH \\ &-6.83783x10^{-3} T_F^2 - 5.481717x10^{-2}\, RH^2 + 1.22874x10^{-3} T_F^2\, RH \\ &+8.5282x10^{-4} T_F\, RH^2 - 1.99x10^{-6} T_F^2\, RH \end{aligned} \tag{3.5}$$

Where (T_F) denotes the air temperature, RH denotes relative humidity and HI are measured in degrees Fahrenheit.

3.4 Characterizing Air Conditioning Prevalence and Affordability

CDD analysis has other applications beyond the direct use of the CDD index. CDD is used as an input to different measures of climate comfort, such as cooling penetration (PNT), and to calculate the ratio of households that could afford air conditioning (Smax). We extended our detailed CDD analysis to these two other indexes due to their use related to human heat comfort [167], [168]. PNT is calculated as the following equation [167]. It represents the percentage of homes in a certain area that have air conditioning.

$$PNT = \begin{cases} 26.33 \ \ln CDD - 81.69, & 0 < CDD < 920 \\ 97.3, & CDD > 920 \end{cases} \tag{3.6}$$

CDD is the summation of annual CDD.

Smax represents the fraction of households in a certain area that would acquire AC if they could afford it [168] and is calculated as shown below.

$$S_{max} = 1 - 0.949\mathrm{e}^{-0.00187CDD} \tag{3.7}$$

CDD here denotes the summation of annual CDD.

3.5 Results

This section starts by first summarizing the results associated with deriving geographically specific CDDs. It is then followed by extending the CDD calculation to also account for humidity, discussing the associated implications under present and future climate conditions.

3.5.1 CDD Base Value Heterogeneity Across the CONUS

To test the hypothesis of whether 65°F (18.3°C) adequately captures thermal comfort across the CONUS, we leverage the energy signature method [164], [165], [169], [170] discussed in the previous section. Implementing the energy signature method involved using the average monthly residential energy consumption data from 1990 to 2016 [171] together with air temperature data for the same period [172]. The differences between the 65°F (18.3°C) and derived optimal set-points are depicted in Figure 3.2(a), with states shaded in orange (blue) representing CDDs with higher (lower) than 65°F (18.3°C) set-point temperatures. Figure 3.3(a) illustrates this same variation as a scatter plot of CDDs with fixed and regionally varying threshold points, such that states farther from the reference (1:1) line show a greater deviation of the geographically specific set-point temperature from the 65°F

(18.3°C). The state of Washington is excluded from Figure 3.2 owing to the relative climate insensitivity of its summer-time demand during the study's time span [36], [173], [174](also see Supplementary Figure 1).

There are significant deviations of the derived base temperature from 65°F (18.3°C), with 30% of the CONUS states showing absolute variation higher than 10% (6.5°F). In Southern states, the optimal set-point temperature is significantly higher than the conventional 65°F base value. For instance, Texas (TX) and Florida (FL) show notable deviations from 65°F, with significant implications for the states' energy planning, given their high population and energy consumption, especially during hot summers. To quantify the implications of these deviations from the commonly used set-point temperature for cooling demand, we harness state-specific reduced form equations established via regressing summer-time energy demand on CDD.

Figure 3.2(b) depicts the implication of the derived CDD variable set-point temperature for the climate-sensitive portion of cooling demand—reported in-terms of the percentage shift in state-level energy consumption, with variations up to 40%. This result demonstrates that in states with negative variations (shaded in gray) the conventional set-point temperature overestimates the climate-sensitive portion of the energy demand. The overestimation has a higher absolute variation, as seen in states like Colorado (CO, -32.8%) and Maine (ME, -38.1%). Conversely, in states with positive variations (shaded in red) the conventional (fixed point) approach underestimates the climate-sensitive demand. While these underestimations are lower in absolute value, they have significant implications in key energy-intensive southern states such as Florida (FL, 9%) and Georgia (GA, 8.9%). Figure 3.3(b) further illustrates this point for the nine states: Florida (FL), Georgia (GA), Louisiana (LA), Kentucky (KY), North Dakota (ND), Montana (MT), Maine (ME), New Hampshire (NH) and Oregon (OR), showing the values for over- (bars in blue) and under-estimations of energy demand (bars in orange). Here the values refer to the differences in per capita energy estimates in MWh—estimated based on the relative differences in consumption shown in Fig. 3.2(b) and considering the respective state populations [175]. Notably the large differences between the fixed and updated baseline methods can be observed for the states of Florida (underestimation by more than 800 K MWh) and Oregon (overestimation by 500 K MWh).

The states where the conventional approach leads to an underestimation of cooling demand present serious challenges to energy planning. More specifically, even a small deviation from forecasted and/or anticipated demand in these states can prove costly not only to energy infrastructure planners and operators but also the consumers. For example, a 9% variation in Florida—in terms of its June 2016 demand—is equivalent of maintaining 55,000 Floridian households' energy for an entire year. In a state where air conditioning takes almost a third of the summer-time energy consumption, this is a serious cause for concern in energy planning [158].

Besides the significant implications of access to geographically specific CDDs for demand forecasting, it has serious consequences for other key elements in energy planning, namely, estimating air conditioning adoption rates, since CDD is the base for other indexes calculations. For example, the use of generic CDDs in calculating Cooling Penetration (PNT) [167] and the fraction of households that would acquire AC if they could afford it (S_{max}) [168] would yield misestimations as high as 9% and 17%, respectively, as we can see in Figure 3.4 for observed values between 1990 and 2016. PNT is also significantly affected for projected CDD and humidity-based CDD, as seen in Supplementary Figure S2 and S3 (up to 28% change for air CDD and a max of 7% in heat index CDD—total average of 5% and 2%, respectively). S_{max} has a greater variation for projected data, shown in Supplementary Figures S4 and S5, with an average of 9% change for air temperature CDD and 6% for heat-index based CDD estimates. Compared to the PNT estimates, S_{max} has a higher variation partly due to nature of its estimation procedure that do not include any threshold limits (Equation 3.7). Nevertheless both variables (PNT and S_{max}) show the overwhelming underestimation in the projected CDD estimates (states in blue; Fig. 3.4, see also Supplementary Fig. S4 and S5), which, as presented, are a source of great concern in energy planning.

3.5.2 On the Role of Humidity

Considering the significant challenges posed by climate change, not only in terms of increased frequency and intensity of extreme heat events over time [47], [157], [176], [177], but also the growing importance of humidity in shaping future air conditioning demand [35],

[36], [164], we analyze the projected changes in CDDs based on air temperature and contrast them with a similar measure based on heat index, which accounts for both air temperature and humidity. We harness the climate projection data-set of five CMIP5-GCMs under the RCP8.5 for the period of 2031-2050.

Heat index-based CDDs are calculated using the same signature method that is used for calculating air temperature-based CDDs. In other words, we estimate the geographically varying optimal heat-index values based on electricity consumption data. For conducting projections under climate change, we use the 2031-2050 time period to be consistent with the time-span most commonly used in mid-term energy planning reports [156], [157], while still accounting for climate change effects.

Figures 3.5(a) and 3.5(b) demonstrate how heat index-based CDD compares with air temperature-based CDD for the geographically-specific (updated CDD) and the conventional 65°F (18.3°C) set-point values, respectively. There is a greater variation across states when using the conventional set-point temperature, suggesting that ignoring humidity in CDD calculations will likely lead to underestimating human heat comfort – varying between ∼22% and ∼36% variations as depicted in Figure 3.5(a) and Figure 3.5(b). States with high energy consumption such as Texas (TX, 20.9%) and Florida (FL, 70.5%), present significant underestimation of CDD projection when compared with the geographically-specific CDD. Throughout the CONUS, the variation is overwhelmingly positive (shades of orange) – a clear sign of underestimation in projected energy demand when comparing the conventional air-based CDD approach with the heat index based estimates.

Figure 3.5(c) and 3.5(d) illustrate the same information, but in the form of scatter-plots, with an average of 22% of projected (2031-2050) underestimation comparing the updated CDD, including top energy-consuming states like California (CA) and Florida (FL). It is clear that most states are shaded in blue – higher values for the heat index CDD – showing once more the significant potential for underestimation when only focusing on air temperature CDDs, either the updated values or the convention fixed set-point values. The same analysis was conducted for observed data (1990-2016), shown in Supplementary Figure S6, that even though the underestimation (states in blue) are not as acute as when considering the future effects of climate change (Fig. 3.5(c) and Fig. 3.5(d)), still 56% of the states present underestimation. Overall these results emphasise the importance of accounting for the humidity related heat-stress measures in estimation of CDD and climate-sensitive portion of energy demand projections.

3.6 Discussion and Concluding Remarks

Increased demand for cooling has been identified as a critical blind spot in today's sustainability discourse [138]. Inadequate characterization of human thermal comfort poses significant challenges to the security and resilience of the grid and present obstacles to achieving SDGs [35], [50], [149]. Despite its widespread use in characterizing thermal comfort, CDD is not a universally reliable proxy for cooling energy demand.

Here, we examine the consequences of calculating CDD based on a generic set-point temperature of 65°F (18.3°C) in energy infrastructure planning. Specifically, we use observed trends in summer-time energy demand to derive geographically specific comfort-zone temperatures across CONUS, and demonstrate the degree to which generic CDDs over- or underestimate demand for cooling by disregarding geographical heterogeneity in thermal comfort across the country. Moreover, by extending the calculation of CDD to also account for humidity, we demonstrate the degree to which current approaches fall short in capturing human thermal comfort under present and future climate conditions.

As the world gets hotter and the demand for cooling energy soars, utilities face unprecedented challenges in reliably balancing the grid, especially during the more frequent and prolonged heat events [24], [36], [47], [52]. We demonstrate that relying on conventional CDD for energy projections and ignoring the critical role of humidity will be costly for both utilities and customers. Credible projections of demand, both in the near term and future, allow policymakers and utilities to develop more sustainable and proactive plans. For instance, policy levers such as carbon tax and demand-side management can decelerate the adoption of AC units, increase the share of renewable generation and incentivize investments in energy-efficient appliances. Additionally, passive cooling designs and nature-inspired construction methods can lower the temperature in buildings and mitigate the soaring demand for cooling. Such design solutions include the use of shades, enhanced wind circulation, green rooftops, evaporative cooling, glass modifications, and bio-inspired cooling technologies [178]–[180]. Higher vegetation in the urban environment has also been shown to have a modulating effect during extreme heat events [181]–[183].

In summary, our study underscores the value of leveraging the observed trends in energy demand in deriving optimal, regionally-specific comfort zone levels for calculating CDDs. Moreover, we demonstrate that disregarding humidity leads to underestimating demand under climate change, with considerable implications for the security of the grid. These insights contribute to pushing the sustainable development agenda and efforts in delivering sustainable cooling to society.

3.7 Figures

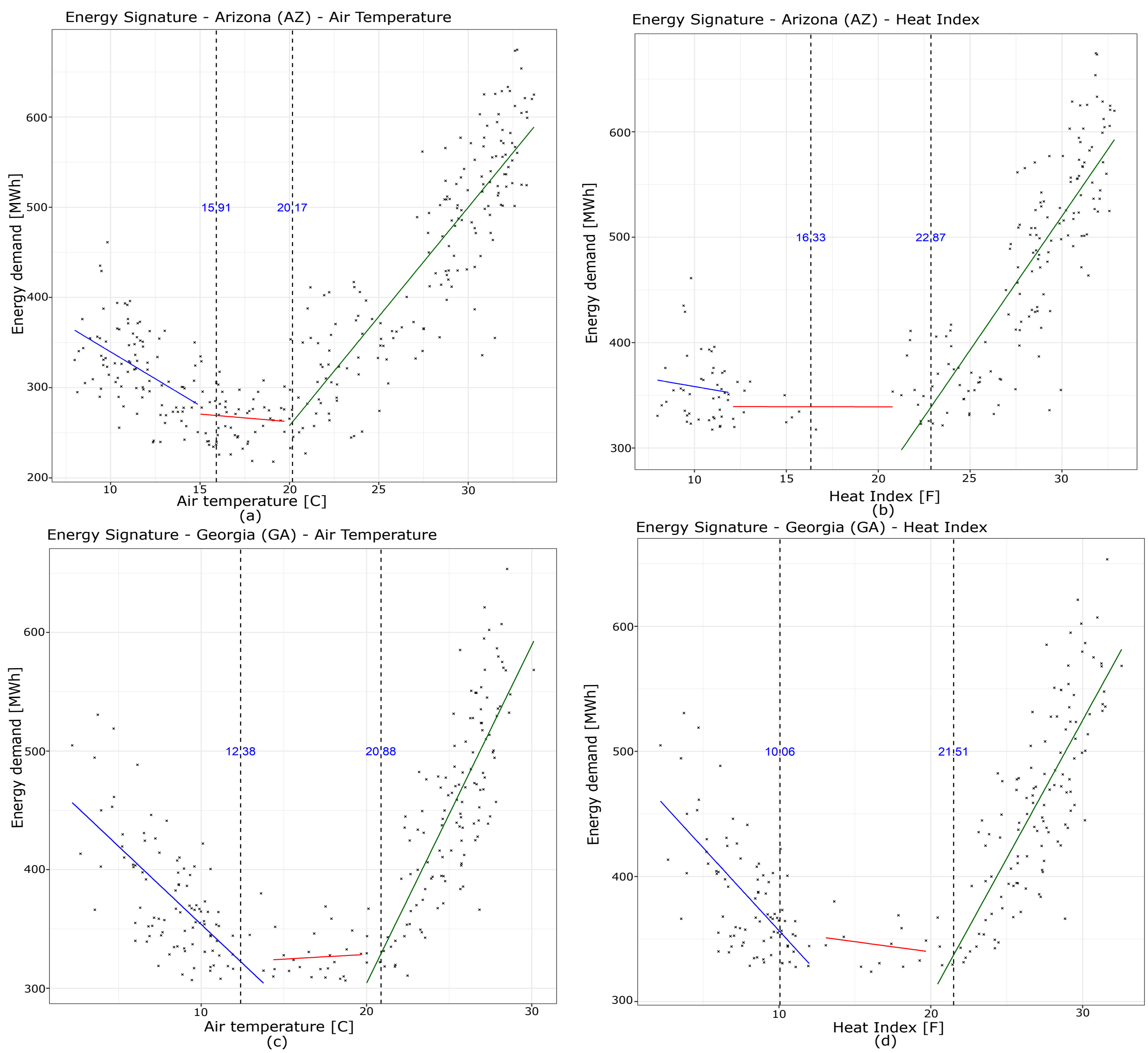


Figure 3.1. An example of the Energy Signature Method conducted for the state of Arizona (AZ) for air temperature CDD (a) and heat index CDD (b); for the state of Georgia (GA) for air temperature CDD (c) and heat index CDD (d). In blue, the determined heating and cooling set points for each state and variable. The three regression lines are identified in the figures, and the base points are the intersection of said lines.

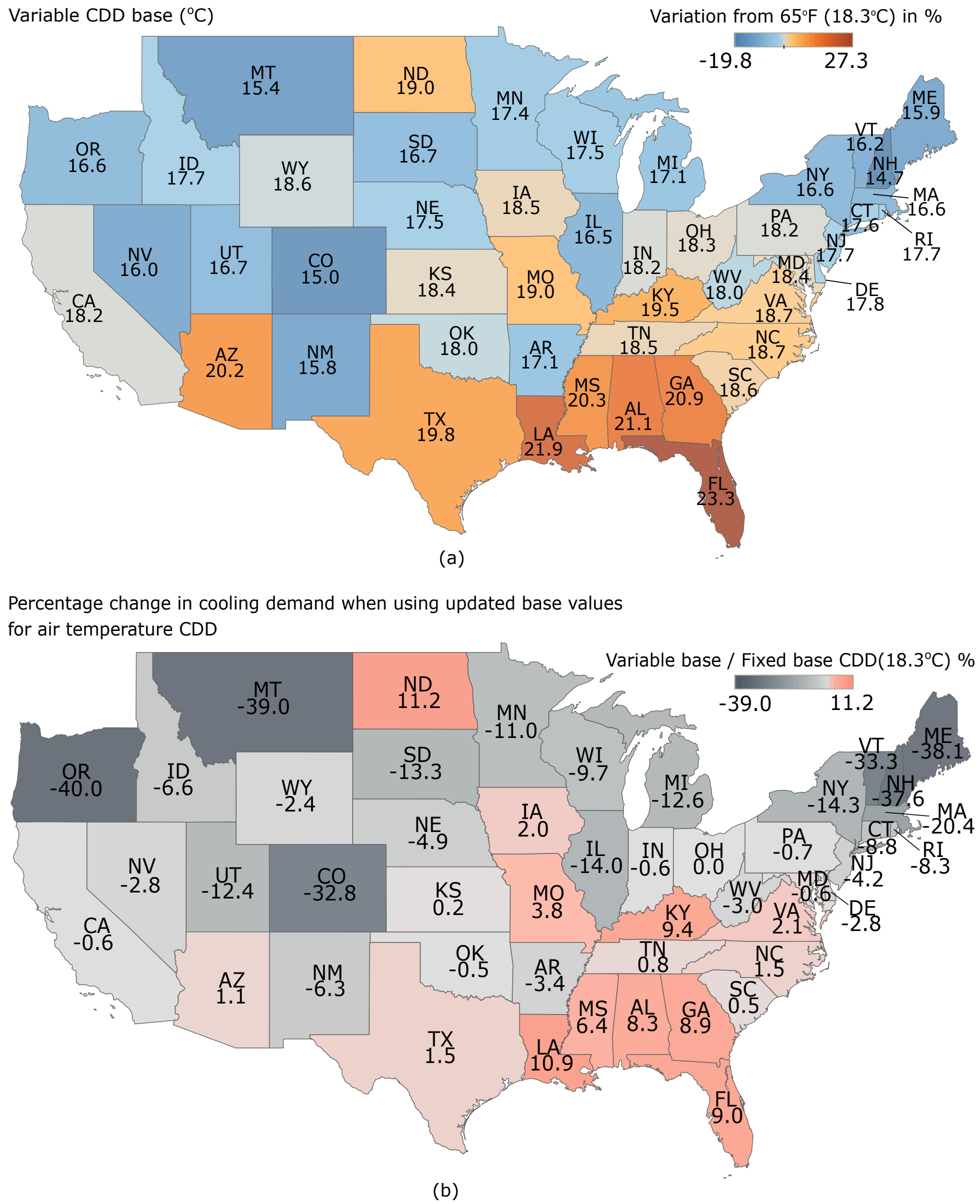
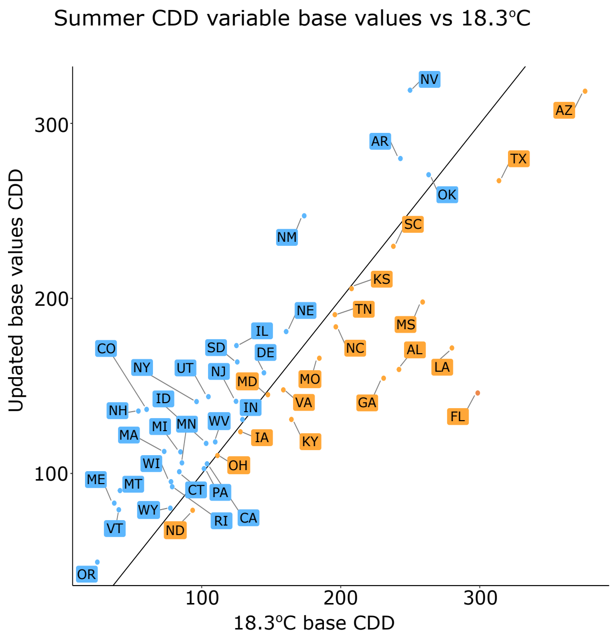
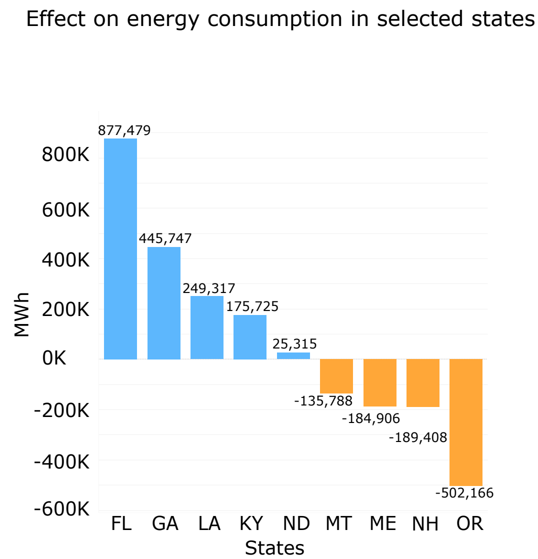


Figure 3.2. (a) The derived CDD air temperature set-points for the CONUS states. In orange (blue), the darker the state color, the greater its positive (negative) variation from the traditionally used 65°F (18.3°C) set-point.(b) Percentage change in the climate-sensitive portion of residential cooling demand in all 48 CONUS states when using the updated set-point for air temperature CDD.



(a)



(e)

Figure 3.3. (a) Comparison of the variable set-points (y-axis) with the fixed 65°F (18.3°C) set-point (x-axis). (b) The resulting impact on energy consumption in thousands of MWh for the states of Florida (FL), Georgia (GA), Louisiana (LA), Kentucky (KY), North Dakota (ND), Montana (MT), New Hampshire (NH) and Oregon (OR).

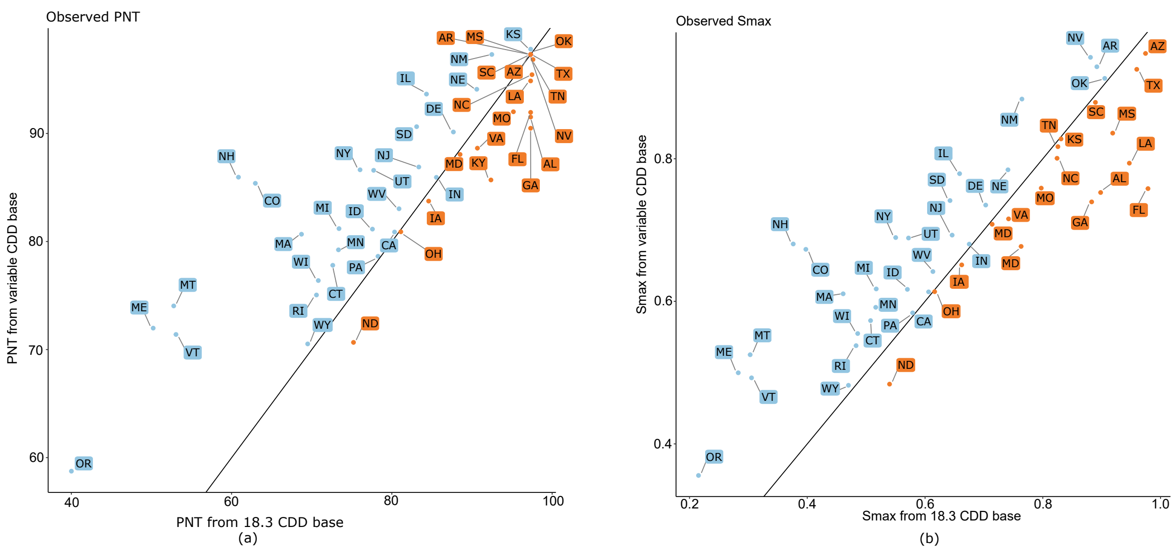


Figure 3.4. (a) Percentage variation in PNT average (1990-2016) for the 65°F (18.3°C) air temperature set-point and the updated/variable set-point values. (b) Similar to (a), but for Smax values. (1:1) line for reference.

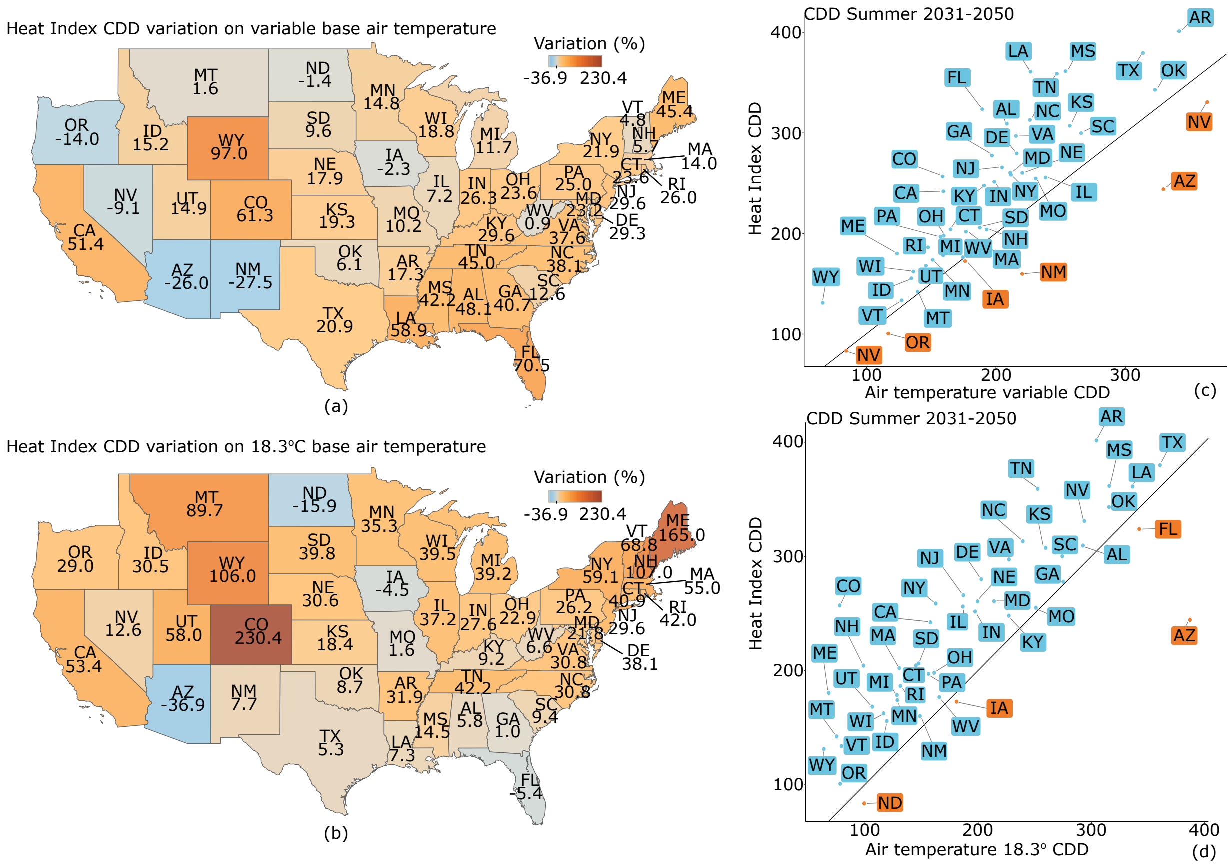


Figure 3.5. (a) Percentage variation in CDD average (2031-2050) between heat index and the updated set-points for air temperature. (b) Similar as (a), but for the 65°F (18.3°C) air temperature set-point. (c) Comparison of heat index summer CDD (2031-2050) and the updated set-points for air temperature. (1:1) line for reference. States in blue have a higher value for heat index CDD; states in orange, lower, when compared to the updated set-points for air temperature. (d) Similar as (c), but for the 65°F (18.3°C) air temperature set-point.

4. A DATA-CENTRIC APPROACH TO INCREASE PREDICTION ACCURACY AT THE UPPER TAILS: A CASE STUDY OF CALIFORNIA AND TEXAS

4.1 Introduction

Since the development and centralization of energy production, followed by privatization and certain levels of deregulation, accurate prediction of daily energy demand has been essential for the resilient functioning of the electric grid [1]–[3]. Given the lack of large scale technologies for energy storage [4], access to accurate prediction of demand is a key pillar in reliable operation of the electric power infrastructure [7], [36].

To develop models that accurately predict load, it is imperative to understand the different factors that influence energy demand. Socio-economic characteristics [184], technological enhancements [185], new policies, and market fluctuations [186] all have a role in shaping the demand curve, but climate variability and change have been the main culprits behind the unprecedented increases in peak load in recent times [36], [47]. With the increase in frequency and intensity of extreme climate events, such as heat waves and droughts [24], [25], summer-time cooling demand is breaking records [35], [50]. Additionally, cooling demand is expected to witness a threefold increase by 2050 under climate change, exposing the population to health hazards if the rising demand for air conditioning is not met [187], [188]. Accurate predictions of the climate-sensitive portion of the peak load (particularly its upper-tail) under climatic extremes has significant equity and justice implications. This is because residential demand is the most climate-sensitive sector and the elderly, children, low income population, and communities of color are the most sensitive to heat stress [49]–[51], [144], [145]. When the unforseen, surging cooling demand triggers power outages during extreme heat waves, the public health consequences are devastating, with thousands of deaths in Europe and America, specially in the marginalized segments of the population [27]–[30].

The literature in energy demand prediction, however, mainly focuses on estimating the central tendency of the demand distribution—i.e., mean and median— and often does not consider the electricity demand distribution in its entirety [188]–[191]. But, predicting the highest demand in a certain period of time (i.e., the peak load) is of extreme importance for energy planning, and the upper tails of the peak load distribution should inform the design of reserve margins to ensure the resilient operation of the grid [3], [47], [54]–[56], [59].

Our goal in this study is to present a data-centric algorithm to increase the accuracy of the peak load predictions at the higher quantiles. The prediction model used in this chapter is based on the non-parametric Bayesian ensemble-of-trees algorithm (BART machine, see Methods). Here, we focus on modeling the daily summer-time peak load in the residential sector in California (1990 to 2016) and Texas (1990-2015). California and Texas are selected for this study due to their large population and economy, as well as being the top 2 energy-consuming states in the U.S. [192]. Given that these states are summer peaking, we focused on only using the summer time peak loads (i.e., May to August). The selected utilities to demonstrate the developed method include the Los Angeles Department of Water and Power (LADWP), San Diego Gas & Electric (SDGE), and Pacific Gas & Electricity (PGE) in California as well as the Electric Reliability Council of Texas (ERCOT) and El Paso Electric Company in Texas.

While achieving higher accuracy of prediction can be achieved via increasing the complexity of the predictive models—for example via leveraging meta algorithms such as bagging, boosting and stacking [193]–[195]—in this study, we leveraged a data-centric approach. We chose this path in order to address the problem in a more simple manner, utilizing less computing time and resources. To evaluate the performance of the developed algorithm, two sets of models were developed: a 'base model' that simply used the raw data, and a 'proposed model' that used the augmented data as input. Our results revealed an average increase of 7% in RMSE in the mean prediction. For the 90th quantile, where extreme peak loads are concentrated and the prediction is historically poor, we had an average improvement of 15%.

4.2 Identifying The Peak Load Prediction Problem

The residential peak load demand data is collected at the daily scale (i.e., highest hourly demand of the day) during 1990 to 2016 in California. and 1990 to 2015 for Texas. Only the peaks loads during the summer months (i.e., in May, June, July and August) were selected [192]. Moreover, daily and monthly electricity price data as well as dummy variables indicating weekdays and holidays were included in the model as control variables [196]. For Texas, no price data was available. Climate data is aggregated to the equivalent geographic locations of each of the utility's peak load data, using population density as a weighting factor [59]. After data collection, the final datasets with different climatic variables, peak load and price data (for California) are generated.

Our predictive analytics conducted in the previous chapter revealed models' shortfall in capturing the data variance in the upper quantiles, as evident in Figure 4.1 which depicts the model's lower out-of-sample performance at the higher quantiles (5800-7000 MWh) for LADWP in California. It is clear from the figure that, in general, the model does not perform poorly, with 80.89% coverage of predicted values. However, the quantiles and area where the model under-performs is very specific: the high quantiles. This is the motivation for this study: to increase model's overall performance, with a special focus on improving accuracy at the higher quantiles.

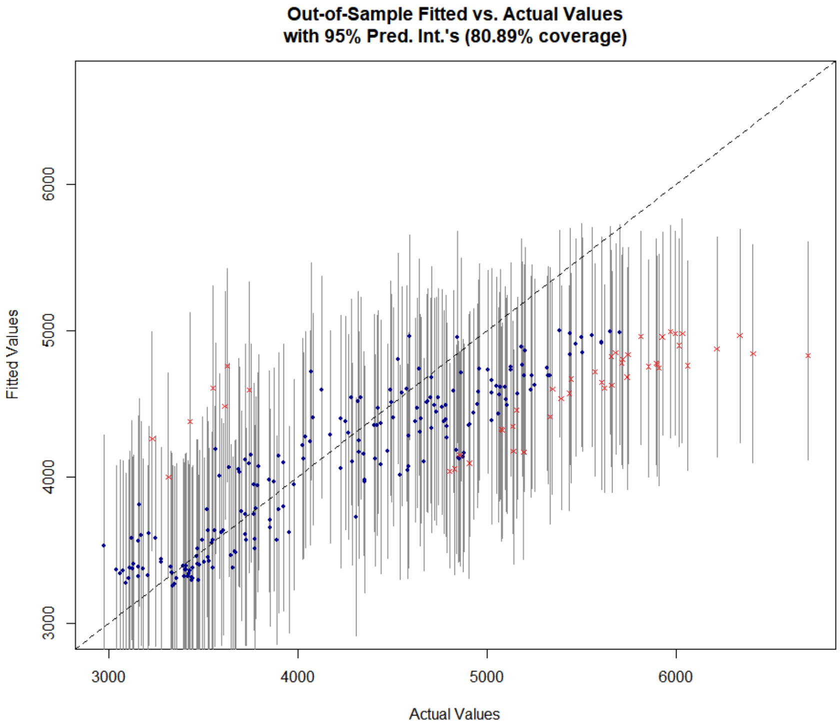


Figure 4.1. LADWP dataset: predicted (fitted) values vs test (actual, unseen, hold-out) set, base BART Machine model, without data treatment. Note the lower performance in the higher quantile, above 5800 MWh.

4.3 Data-driven Approach

Many methods exist that can increase the accuracy of predictive models, with varying levels of complexity [3], [47], [54]–[56], [193]–[195]. In the present study, a data-centric approach is developed. The ' base model' (i.e., the model that leverages the raw data) often falls short of capturing the variability at the higher quantiles since the more extreme data in the test set are often not available in the training set and thus cannot inform the model training. This scenario will likely become even more extreme under climate change. To increase the chance of capturing this 'unseen' extreme demand, we developed a data-centric algorithm that include three key components: (i) a time series forecasting component; (ii) an adjustment to the 'multiplication factor' phase; and (iii) the incorporation of the augmented data—i.e., the forecast times the multiplication factor—in the original training data set. The flowchart of the process is illustrated in Figure 4.2.

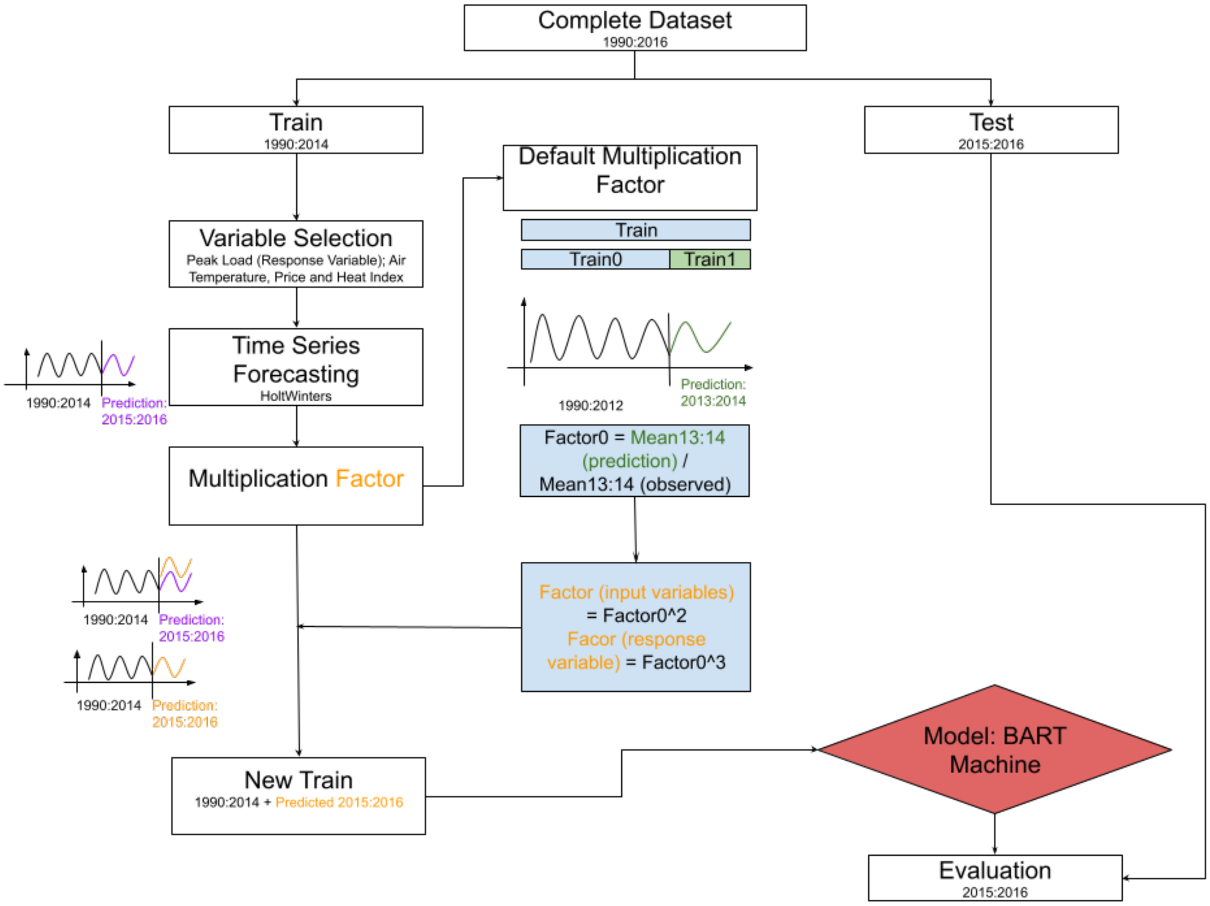


Figure 4.2. Flowchart of the developed methodology.

Initially, variable selection is performed on the original train set (chronologically divided, 1990-2014). The selected variables are converted into time series and used to forecast the same chronological space for the test set (2015-2016) with the Holt-Winters forecasting procedure with an additive seasonal effect (Methods). However, this simple forecast underestimates the increasing tendency for the peak load variable (response variable). Aware of the under-performance of the forecast time series, this forecast is multiplied by a factor. To determine an initial value for this multiplication factor, the train set is once more divided (1990-2012—Train0—and 2013-2014—Train1), and used for the forecasting (Holt-Winters, Methods) of 2013-2014 partition (Train1 in Fig. 4.2), values that we have access in the original train set for comparison. The ratio between the 2013-2014 forecast mean and the 2013-2014 observed data mean is the base for the multiplication factor, gauging how below the actual data our initial forecast is. If this ratio is bellow 1, it shows that in that specific period of the train set, there is a decreasing tendency between the observed data mean and the forecast values. In this case, if the variable is the response variable (peak load in this study), we invert the ratio to have a value above 1, because we identified the problem as an underestimation of the response variable, thus we need to make sure to create higher values in the train set, not lower. If the variable is not the response variable, no multiplication factor is used by converting the multiplication factor to 1, i.e., if there is a declining tendency for the independent variable, no multiplication is performed for that variable in the method.

In the case of a ratio above 1, the ratio is squared to be the multiplication factor for the independent variables, and the cube of the ratio is used as the default multiplication factor for the response variable (peak load). The final value for the ratio exponent for the response variable is iterated based on RMSE final results, but 3 is the default value. The reason why the ratio is multiplied by an exponent comes from the analysis of the time series forecast: since the final use of the multiplication factor is on the further forecast time period of 2015-2016 (test set time window), it is assumed that the ratio increases when comparing the forecast with the actual value, thus the use of an exponential value (as in an 'interest' over the forecast values).

Finally, with the initial values for the multiplication factor (this value can be further adjusted interactively), the original time series forecast for 2015-2016 is multiplied by the multiplication factor. The multiplied forecast values are combined with the original train set, and a new train set is achieved. A BART Machine (Methods) model is leveraged with the new train set.

Since this method is focused on the data, different algorithms besides BART Machine can be applied in this final phase of the methodology. This non-parametric Bayesian ensemble-of-trees algorithm [197] is chosen in this study due to the fact it has been shown to outperform other climate–demand nexus models in terms of predictive accuracy [2], [8], [9].

4.4 Methods

4.4.1 BART Machine Algorithm

We leveraged a non-parametric Bayesian ensemble-of-trees algorithm to characterize the climate sensitivity of electricity consumption, since the algorithm was shown to outperform other climate-demand nexus models in terms of predictive accuracy [2], [8]–[10], [142], [197]. The models' hyperparameters were tuned using a 10-fold cross-validation technique [57].

The non-parametric Bayesian sum-of-trees model can be represented using the equation below [197], [198],

$$\mathrm{Y} = \sum_{\mathrm{j}=1}^{m} g(x; T_{\mathrm{j}}, M_{\mathrm{j}}) + \varepsilon, \varepsilon \sim N(0, \sigma^2) \tag{4.1}$$

Here, Y denotes the response variable, g denotes the regression tree function, T_{j} denotes individual tree structures, M_{j} denotes a set of parameters (e.g., mean values) associated with the tree nodes. BART is the sum of m total trees. The decision rules associated with each tree involve binary splitting of the predictor space. So for a given T and M, $g(x; T, M)$ represents a tree with parameters M assigning the binary splits for the terminal nodes of T. ε is the stochastic irreducible error, assumed to be normally distributed with a constant variance chipman2010.

The BART algorithm uses a prior over the sum-of-trees parameters to regularize the trees, i.e., to weaken the effect of individual trees with the goal for better prediction when adding all decision trees. A prior is imposed on all parameters, $(T_{\mathrm{j}}, M_{\mathrm{j}})$ $[\forall \mathrm{j} = 1, ..., m]$ and σ. This regularization is made to keep individual trees effects from being unduly influential over the entire sum-of-trees. Without a regularization, large tree components would overwhelm the sum-of-trees and would diminish the benefits of leveraging additive tree structures. The prior regularization can be mathematically represented as:

$$\mathrm{p}(M_{\mathrm{j}}|T_{\mathrm{j}}) = \prod_{\mathrm{i}}^{m} p(\mu_{\mathrm{ij}}|T_{\mathrm{j}}), \quad where \quad \mu_{\mathrm{ij}} \in M_{\mathrm{j}} \tag{4.2}$$

Where tree components T_{j} and M_{j} are assumed to be independent of each other and of σ.

The presented data-driven methodology of this paper depends on model variable selection. Specifically, the variable selection algorithm involved identifying the variables most frequently used as decision/splitting rules in the sum-of-tree model [197].

4.4.2 Climate Data

Observed climate data is acquired at a sub-daily time scale for the period of 1990-2016 from the NCEP North American Regional Reanalysis (NARR) [172], [199], aggregated at a daily level to match the temporal scale of electricity consumption data, and weighted by population so as to give a greater weight to areas with higher population density when aggregating to the state level. The input climate data include air temperature, dew point temperature, discomfort index, heat index, humidex, wet bulb temperature, and wet bulb globe temperature. All climate variables use a combination of temperature, humidity and pressure.

Initially, many climate variables are used for the variable selection. A full extend of all the heat measures used before variable selection is presented here:

Dew Point refers to the temperature which air is saturated with water vapor, calculated in this paper using the equation below:

$$t_d \approx t - \left(\frac{100 - RH}{5}\right) \tag{4.3}$$

Where t_d is the dew point temperature and t is air temperature in Celsius, and RH stands for Relative Humidity in percent (same RH as in the other equations in this section)[200].

Wet Bulb Temperature is the lowest temperature to which air can be cooled by water evaporation at a constant pressure. In this paper, it follows the equation:

$$T_W = T_W - \frac{f(T_w; \boldsymbol{\pi}) - (C/T_E)^{\lambda}}{f'(T_W; \boldsymbol{\pi})} \tag{4.4}$$

Where $\boldsymbol{\pi} = (p/p_0)^{1/\lambda}$ is the Exner function, used for scaling, λ is the inverse of the Poisson constant for dry air (3.504), p pressure (mb), p_0 the reference pressure (1000mb); C is the freezing temperature in Kelvin, and T_E the equivalent temperature, which is the moist potential temperature scaled by λ. For reviewing the complete derivation of the equation, refer to the Appendix A of [67], [201].

Discomfort Index (DI) was developed in the 1950s to calibrate air conditioners and further adapted by the Israeli Defense Force as a main measure of heat stress.

$$DI = 0.5T_W + 0.5T_C \tag{4.5}$$

Where T_W stands for Wet Bulb Temperature, RH is percentage Relative Humidity and T_C is the Temperature in Celsius. DI is unitless [67], [202], [203].

Wet Bulb Globe Temperature (sWBGT) is the heat stress measure in direct sunlight.

$$sWBGT = 0.56T_C + \frac{0.393\mathrm{e}_{RH}}{100} + 3.94 \tag{4.6}$$

Where $\mathrm{e}_{RH} = (RH/100)\mathrm{e}_{sPa}$; wet bulb temperature used here is unitless and widely used.

e_{RH} represents the vapor pressure in pascals and is calculated from RH in percentage and saturated vapor pressure [67].

Humidex was developed for the Meteorological Service of Canada. It is a unitless [67], [204] index, aiming to explain what the temperature feels like for the human body.

$$HUMIDEX = T_C + \frac{5}{9}\left(\frac{\mathrm{e}_{RH}}{100} - 10\right) \tag{4.7}$$

Heat Index (HI) is also called apparent temperature, describing what the temperature feels like to the human body when relative humidity is combined with air temperature.

$$\begin{aligned} HI = -42.379 + 2.04901523T_F + 10.14333127RH - 0.22475541T_F RH \\ -6.83783x10^{-3}T_F^2 - 5.481717x10^{-2}RH^2 + 1.22874x10^{-3}T_F^2 RH \\ +8.5282x10^{-4}T_F RH^2 - 1.99x10^{-6}T_F^2 RH \end{aligned} \quad (4.8)$$

Where (T_F) denotes the air temperature, and HI are measured in degrees Fahrenheit [67], [166].

4.4.3 Time Series Forecast

Holts-Winter time series forecasting techniques is used widely in literature [205]–[207]. The additive Holt-Winters prediction function, used in this paper, is characterized as [208], [209]:

$$\hat{Y}[t+h] = a[t] + hb[t] + s[t-p+1+(h-1)modp] \quad (4.9)$$

This for a time series with period length p, and where $a[t]$, $b[t]$ and $s[t]$ are given by:

$$a[t] = \alpha(Y[t] - s[t-p]) + (1-\alpha)(a[t-1] + b[t-1]) \quad (4.10)$$

$$b[t] = \beta(a[t] - a[t-1]) + (1-\beta)b[t-1] \quad (4.11)$$

$$s[t] = \gamma(Y[t] - a[t]) + (1-\gamma)s[t-p] \quad (4.12)$$

4.5 Results

An initial analysis on each respondents' train and test sets quantiles and distributions is shown on Figure 4.3, Figure 4.4 (boxplots), Figure 4.5 and Figure 4.6 (distributions). Texas' respondents (and part of New Mexico, since one of the Texas' respondents acts beyond the Texas' border) are subdivided in 9 groups, namely TX1 (El Paso Electric Company), TX2 (ERCORT Coast), TX3 (ERCOT East), TX4 (ERCOT Far West), TX5 (ERCOT North), TX6 (ERCOT North Central), TX7 (ERCOT South Central), TX8 (ERCOT South), and TX9 (ERCOT West) [210]. Three respondents from California are shown in this analysis, the Los Angeles Department of Water and Power (LADWP), San Diego Gas & Electric (SDGE), and Pacific Gas and Eletric (PGE). In the analysed datasets there is a clear wide range of the energy peak demand quantiles, from 600 to 26000 MWh, showing how the different regions act in varying scales of daily highest demand. This finding corroborates the analysis process to be executed as one leveraged model per respondent, and not a single model for the entire state or region.

The discrepancies between train and test sets highlight the possible problem of inaccuracy for the high quantiles' prediction, as we can see for LADWP in Figure 4.3 and Figure 4.1. The test set displayed in green in Figure 4.3 presents higher quantiles than the train set in orange, showing that the test or hold out set has unseen, greater peak loads when comparing to the train set. The machine learning model, fit to the train set, underestimates for the highest demand on the test set, as shown on Figure 4.1. This variability between train and test set is one possible pointer for the under-performance of the high quantiles prediction.

The presented algorithm in this work is, in a way, a path to 'interpolate' values between the known train set and the unseen test set. The final goal is for each respondent to be prepared for the increasing values in peak load, creating data that mimics the quantiles found in the test set (which continues to increase in part due to climate change effects). However, it is important to highlight that the original test set is not used in the development of the model, only in its evaluation, following good machine learning practices to avoid data leakage—i.e., not to use the hold out set to fit the model and sub-sequentially use the test set to evaluate the model.

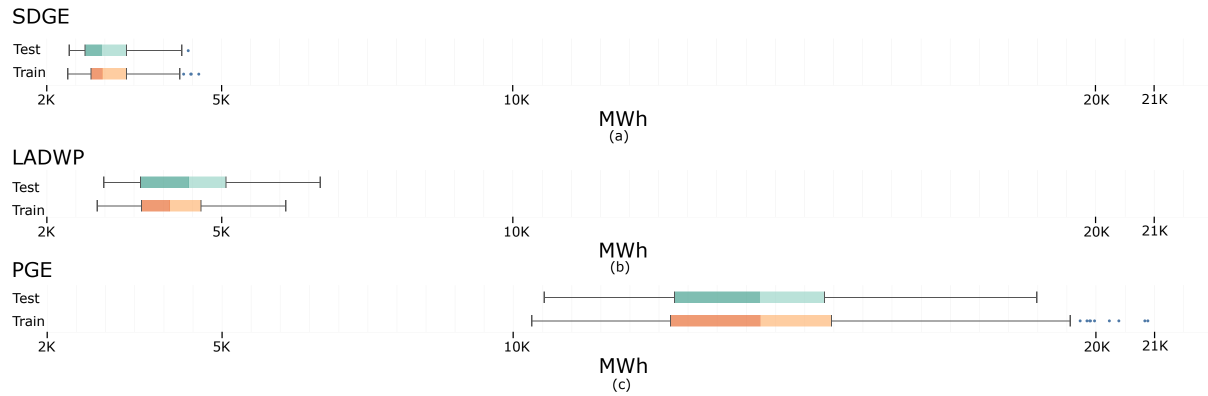


Figure 4.3. Train and test set boxplots for each of California's respondents. In green, the test set quantiles. In orange, the train set quantiles.

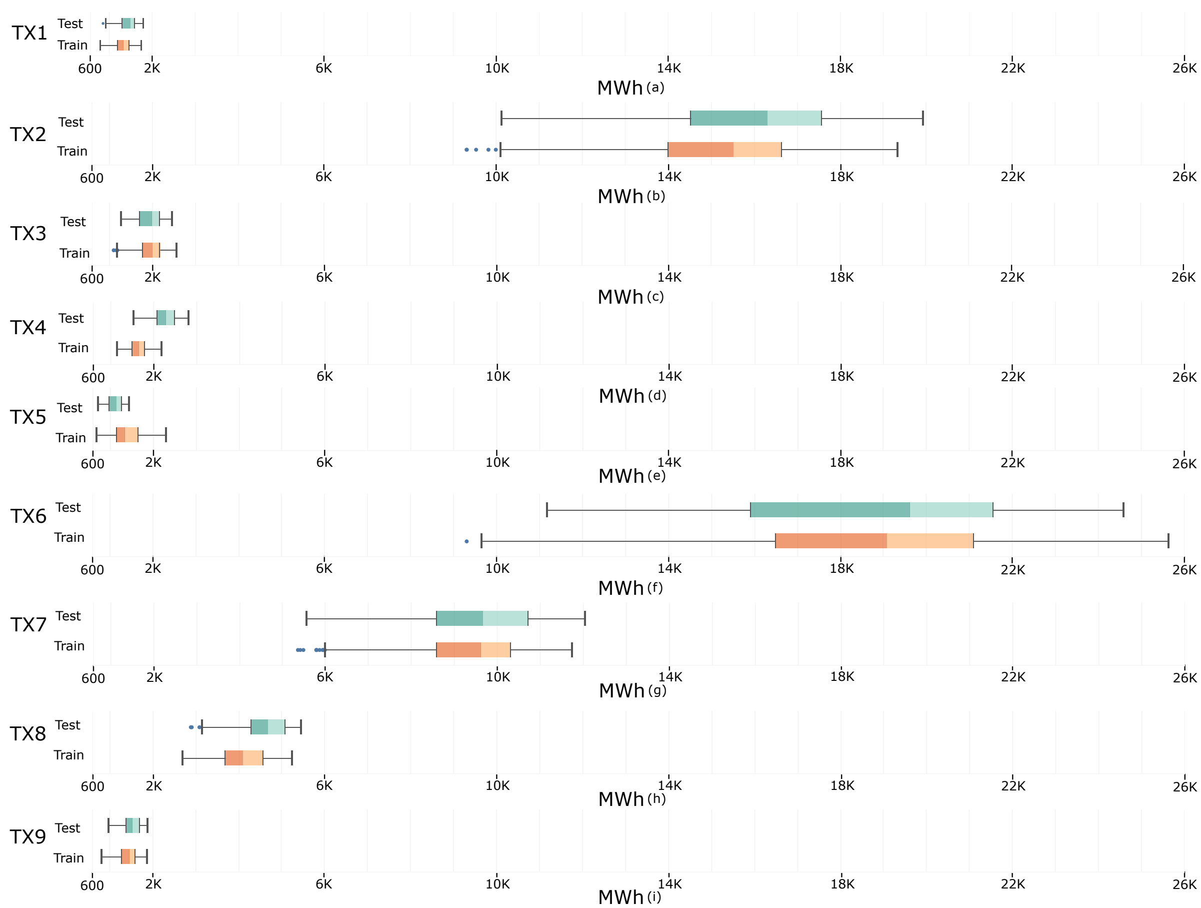


Figure 4.4. Train and test set boxplots for each of Texas's respondents. In green, the test set quantiles. In orange, the train set quantiles.

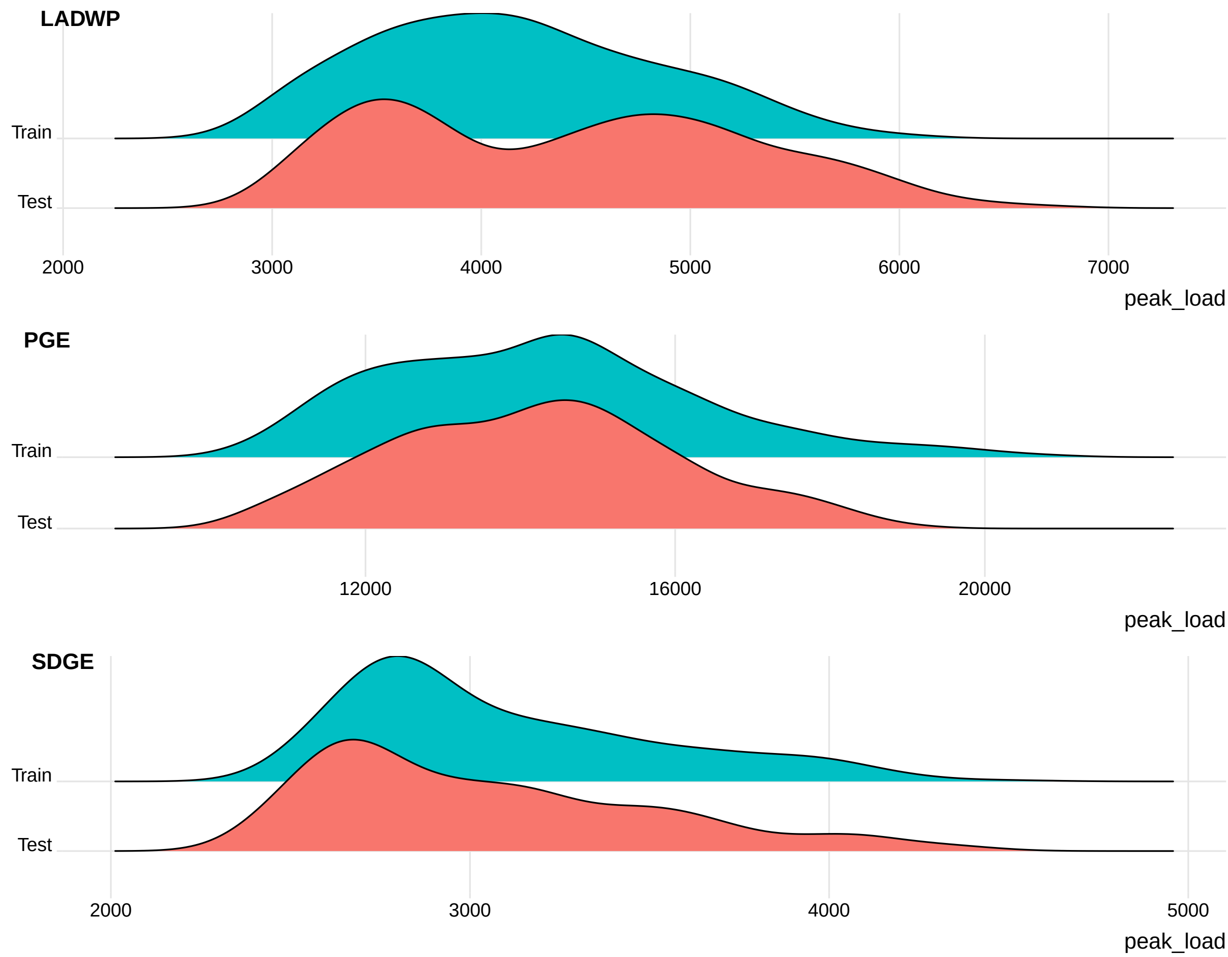


Figure 4.5. Train and test set distributions for each of California's respondents.

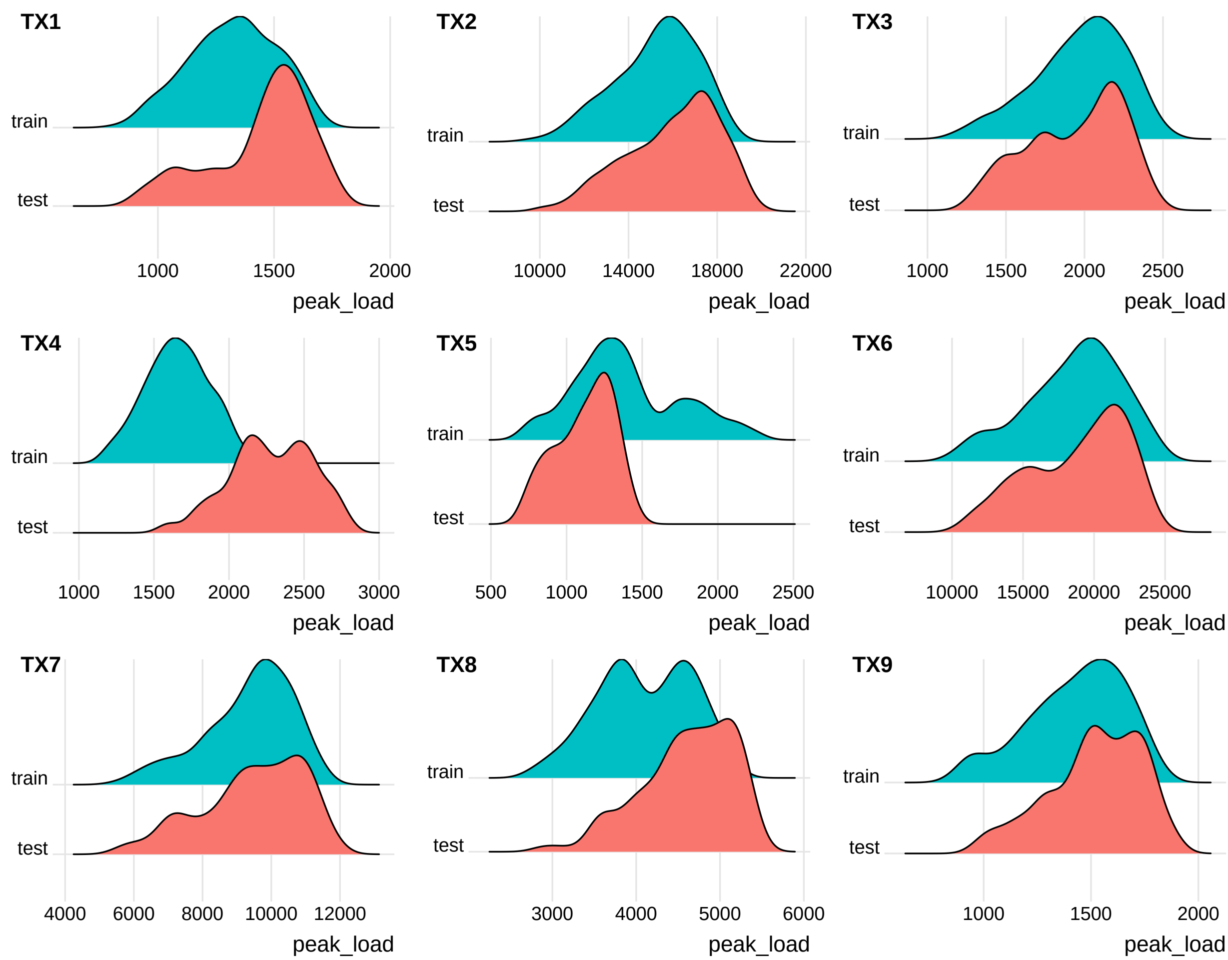


Figure 4.6. Train and test set distributions for each of Texas' respondents.

The results for the altered models are shown in Figure ?? and Figure ??, with the comparison of the test set RMSE from the base model (without the data treatment) with the RMSE from the model leveraged with the alternated data from the described methodology. An average of 7% improvement is shown in the general (central tendency) of the models. Respondents that under-perform, or do not shown significant change such as PGE and TX3, do not present great variability between the train and test sets in Figure 4.3 and Figure 4.4. In this case, the model performs satisfactorily without the data-driven method possibly due to its lack of great variability between historical and the tested future data.

However, it is expected a certain heterogeneity in the results between the datasets, since they represent different regions with different population density and consumption patterns.

In other words, even respondents with less variability between train and test, like SDGE, can show great improvement with the method.

Respondents TX8 and TX9 also perform significantly well, and they all represent areas with average daily consumption bellow 6K MWh, representing smaller respondents in the national scale. TX8 and TX9, as seen in Figure 4.4, show the greater quantiles for the test set (green) when compared to the train set (orange), another example where historical data does not encompass future peak load demand, and the presented method worked satisfactorily to bridge this gap. In terms of scale, TX2, a respondent with high peak load quantiles, had 12% increase in accuracy for both general RMSE and the 90th quantile RMSE, which shows that the data-driven model is not limited to respondents with lower peak load demand.

The result for the 90th quantile improvement is 15% on average for all respondents, higher than the general improvement, which is 7% on average. This demonstrates the data-driven method is efficient in its main goal to improve the prediction of higher quantiles. SDGE, specifically, has a significant improvement of 55% for the 90th quantile (Fig. 4.9), followed by TX8 (31.1%) and TX9 (27.3%), respondents with peak load quantiles bellow 6000 MWh.

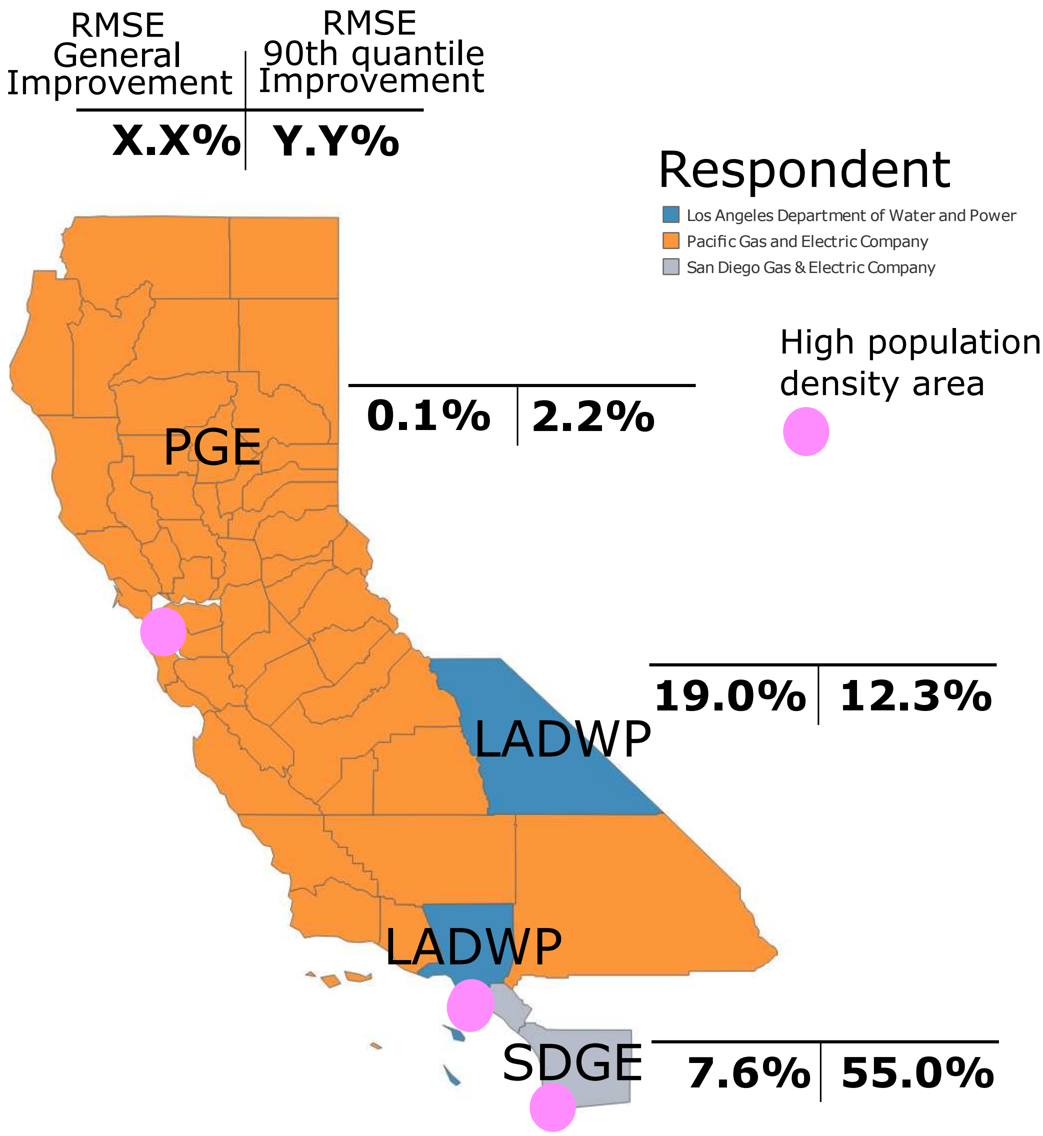


Figure 4.7. Results for the state of California.

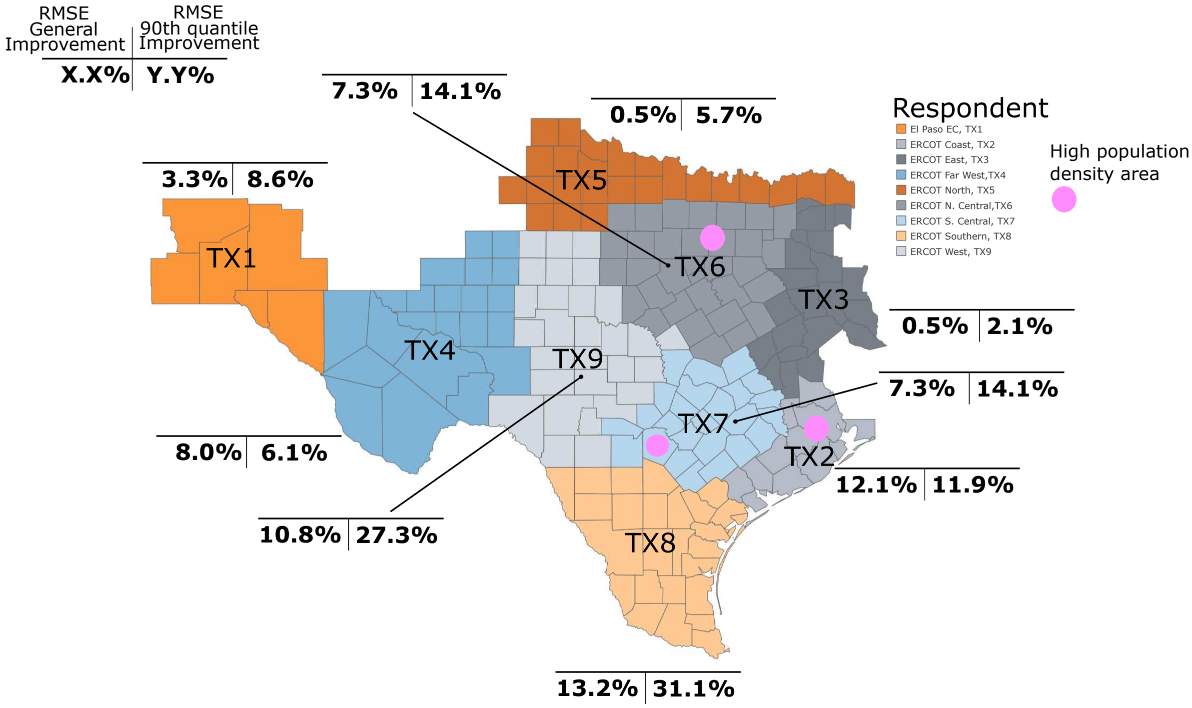


Figure 4.8. Results for the state of Texas (and parts of New Mexico, since the respondent TX1 acts beyond the Texas' border).

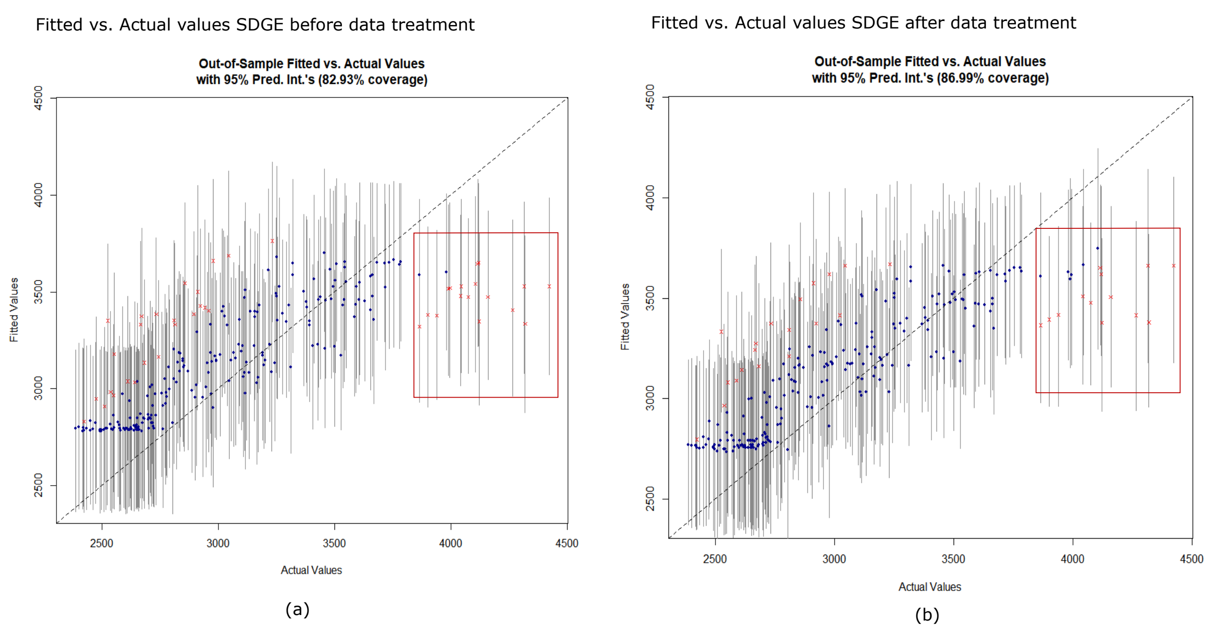


Figure 4.9. Fitted vs. actual values for the respondent SDGE before (a) and after (b) the data-driven treatment. Highlighted in the red rectangle the higher quantiles of the data.

4.6 Discussion and Concluding Remarks

Accurate prediction of energy demand is essential for the regular function of the energy infrastructure, a vital organ in our society. Peak load prediction is specially important due to its role in energy planning design and maintenance [7], [36]. In this work, we present a data-driven methodology to manipulate daily peak load historic data to increase accuracy focused on the higher quantiles, i.e., the highest peak loads of the historical data—which usually represent events related to extreme climate with disastrous consequences when the demand is not met [187], [188].

Implications of this work include accurate information for police makers and utilities to develop sustainable and resilient plans when working on energy infrastructure. Considering the drastic effects of climate in the energy demand [49]–[51], [144], [145], studies like the present one can influence policies such as carbon taxes, adoption of renewable sources, and incentive of energy efficiency cooling appliances and other strategies aiming for climate change mitigation [211]–[213].

This study is also an alert on possible climate effects on energy demand, with significant socio-economics costs to society in consequence of a lack of preparedness for the surge in peak load. Health hazards, particularly for the most sensitive population (the elderly, children, and low income population), must be taken into account when designing the electric grid expansion, adaptation, and transformation measures related to the increase in demand [214].

The specific data limitation treated in this work is usually characteristic of utilities that act in limited geographic areas or with lower peak load demand distribution (Fig 4.6), but not limited to it, as we saw in the TX2 (ERCORT Coast) region. For example, a respondent with extensive geographic locations and higher demand in general (including peak loads), has a higher chance of performing satisfactorily for unseen data due to its high variability, such as PGE in California and TX3 (ERCOT East) in Texas. However, from the 3,300 electric utilities in America, 200 are responsible for the majority of the demand, leaving thousands of utilities in similar situation as LADWP and SDGE, as in concentrated clientele and less variation in historic data [215], thus more susceptible to climate extremes. Under the increasing influence of climate change [24], [25], smaller utilities are going to see frequent unprecedented demands, targeting millions of residencies. In addition to the climate pressing demand, smaller respondents have less flexibility to deal with unprecedented charges and are easily targeted with economic insecurities and even bankruptcy when hazardous events evoke high demand and price surges, such as recently in Texas [216], [217]. With that, this study has vital implications for the many utilities working in this focused scale.

Finally, the presented findings can be expanded to other types of datasets with the similar problem of failing to predict unseen, high quantiles. Additionally, the same data-driven methodology can be expanded to other algorithms applied in different research areas, since this method alters solely the input data.

4.6.1 Other Attempts

The presented problem—prediction inaccuracy for higher quantile—is not, by far, a simple research topic. While working on this chapter, other attempts were developed to approach the issue. In light of research transparency, here is a brief discussion on earlier attempts developed before the presented methodology was achieved.

An initial attempt involved data perturbation. The goal was to perturb the original train set in such a guided way to include values that would be evaluated in the test set (idea that persisted throughout the different approaches). This was specifically attempted through oversampling high quantiles and under-sampling lower quantiles in the train set. Without satisfactory results, a new approach followed: an unbalanced class approach. That is, create a class for ‘low’, ‘median’, and ‘high’ quantiles and treat the ‘high’ quantile class as the unbalanced class. Attempts to balance the dataset were tested, building on the idea of over and under-sampling the train set with different algorithms, such as SMOTE and ROSE [218], [219].

Following the idea of creating a new classification label, a hybrid approach was attempted. It consisted of two steps: a neural network to predict the labels of the train set and create a new variable to classify each entry, followed by the usual fitting of the model with the additional created label as a new variable. Besides the mentioned BART algorithm, quantile regression random forests were implemented to analyse the effectiveness of the new methods on different quantiles. Ultimately, it was decided to focus the testing on one algorithm, and BART Machine was chosen due to aforementioned reasons of its high performance for prediction on the climate-energy demand data.

After these attempts that were a mix of algorithm-centric and data-centric, a decision was made to focus solely on the data-centric approaches and to use the BART algorithm to evaluate the results. Another data-centric approach was attempted through wavelength transformations, which had been shown to increase accuracy for energy demand prediction [220]. It did not yield satisfactory results, leading to the development of the methodology presented in this chapter.

4.6.2 Future Work

Considering the complexity of this problem, is its not expected for a single methodology to offer a complete solution. Thus, the presented methodology in this chapter can be further improved. One area for improvement is how the multiplication factor would vary over time when considering, for example, projected data. The results shown in this chapter are related to observed data, however, in previous chapters is was presented that there can be divergences in the analysis of historical data and projected (future) data. How the default multiplication factor in the presented method would be calculated in that new, projected time window, could change over time. An initial hypothesis would be that the values for higher quantiles of peak load would increase, as seen in the historical data trend. That would culminate with possible higher values for the exponent of the multiplication factor.

Another area that can be expended in this study is the inclusion of other datasets. Not only climate and energy related data, but other types of datasets that present a similar challenge to increase prediction accuracy at the upper tails.

5. CONCLUSION

The challenge of modeling the electricity demand-climate nexus remains a hot topic in research. This dissertation has made foundational contribution to the literature on the topic by (i) quantifying the cost of ignoring humidity in demand prediction under present and future climate conditions, (ii) characterizing the value of deriving geographically-specific set-point values in calculating CDDs, and demonstrating the utility of extending the CDD calculation to include the role of humidity under present and future climate conditions, and (iii) developing a novel algorithm to enhance the predictive accuracy of peak load, with a special focus on the upper tails of the peak load distributions.

Specifically, Chapter 2 leverages the state-of-the-art predictive modeling to explore the role of humidity in characterizing the climate sensitivity of electricity demand at CONUS, under both present and future climate scenarios. The analysis establishes that air temperature, the most widely used metric in demand prediction, is a necessary but not sufficient variable for characterizing the climate-sensitive portion of the demand for cooling energy, and that humidity plays an equally central role. The results indicate that in the high electricity consuming states (e.g., CA and TX), demand projections that are solely based on air temperature lead to underestimating cooling demand by as high as 10% (15%) under present (future) climate scenarios.

Chapter 3 delved into one of the main climate measures for energy prediction, namely, the Cooling Degree Days (CDD). Specifically, the chapter presents a comprehensive analysis of the historical climate and energy demand data to show that the most frequently used base temperature of 65 deg F in calculating CDD leads to mischaracterizing the residents' thermal comfort across different states in the U.S. This mischaracterization causes significant miss-estimations of the cooling demand in the residential sector. Moreover, the CDD calculations is extended to also include humidity. Results demonstrate the significant cost of disregarding the role of humidity in CDD calculations under both present and future climate change scenarios.

Finally, from exploring the electricity demand data sets in the initial chapters, the last chapter of this dissertation focused on improving the prediction power of the climate-peak demand nexus, with a special focus at the higher quantiles. Developing a data-centric algorithm, Chapter 4 demonstrated that it is possible to increase prediction accuracy of the peak load, with a data augmentation and manipulation. Applying the new algorithm to the major utilities in the states of California and Texas, this chapter showed significant improvement of peak load prediction. While the data-centric method was tested for peak load prediction in CA and TX, it can be readily extended to other data sets and regions.

REFERENCES

[1] G. Bakke, *The Grid: The Fraying Wires Between Americans and Our Energy Future*, en. Bloomsbury Publishing USA, Jul. 2016, ISBN: 978-1-62040-124-8.

[2] S. Mukherjee, R. Nateghi, and M. Hastak, "A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S.," *Reliability Engineering & System Safety*, vol. 175, pp. 283–305, Jul. 2018, ISSN: 0951-8320. DOI: 10.1016/j.ress.2018.03.015.

[3] B. J. van Ruijven, E. D. Cian, and I. S. Wing, "Amplification of future energy demand growth due to climate change," En, *Nature Communications*, vol. 10, no. 1, p. 2762, Jun. 2019, ISSN: 2041-1723. DOI: 10.1038/s41467-019-10399-3.

[4] T. M. Gür, "Review of electrical energy storage technologies, materials and systems: Challenges and prospects for large-scale grid storage," en, *Energy & Environmental Science*, vol. 11, no. 10, pp. 2696–2767, Oct. 2018, ISSN: 1754-5706. DOI: 10.1039/C8EE01419A.

[5] Y. Zhao, C. Zhang, Y. Zhang, Z. Wang, and J. Li, "A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis," en, *Energy and Built Environment*, vol. 1, no. 2, pp. 149–164, Apr. 2020, ISSN: 2666-1233. DOI: 10.1016/j.enbenv.2019.11.003.

[6] L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, and W. Livingood, "A review of machine learning in building load prediction," en, *Applied Energy*, vol. 285, p. 116 452, Mar. 2021, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.116452.

[7] P. Alipour, S. Mukherjee, and R. Nateghi, "Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the Texas region," *Energy*, vol. 185, pp. 1143–1153, Oct. 2019, ISSN: 0360-5442. DOI: 10.1016/j.energy.2019.07.074.

[8] R. Nateghi and S. Mukherjee, "A multi-paradigm framework to assess the impacts of climate change on end-use energy demand," en, *PLOS ONE*, vol. 12, no. 11, e0188033, Nov. 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0188033.

[9] S. Mukherjee and R. Nateghi, "Climate sensitivity of end-use electricity consumption in the built environment: An application to the state of Florida, United States," *Energy*, vol. 128, pp. 688–700, Jun. 2017, ISSN: 0360-5442. DOI: 10.1016/j.energy.2017.04.034.

[10] S. Mukhopadhyay and R. Nateghi, "Estimating climate - Demand Nexus to support longterm adequacy planning in the energy sector," in *IEEE Power and Energy Society General Meeting*, vol. 2018-January, 2018, pp. 1–5. DOI: 10.1109/PESGM.2017.8274648.

[11] G. Notton, M.-L. Nivet, C. Voyant, C. Paoli, C. Darras, F. Motte, and A. Fouilloy, "Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting," en, *Renewable and Sustainable Energy Reviews*, vol. 87, pp. 96–105, May 2018, ISSN: 1364-0321. DOI: 10.1016/j.rser.2018.02.007.

[12] M. A. Delucchi and M. Z. Jacobson, "Providing all global energy with wind, water, and solar power, Part II: Reliability, system and transmission costs, and policies," en, *Energy Policy*, vol. 39, no. 3, pp. 1170–1190, Mar. 2011, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2010.11.045.

[13] Q. Schiermeier, "Germany enlists machine learning to boost renewables revolution," en, *Nature News*, vol. 535, no. 7611, p. 212, Jul. 2016. DOI: 10.1038/535212a.

[14] P. A. Owusu and S. Asumadu-Sarkodie, "A review of renewable energy sources, sustainability issues and climate change mitigation," *Cogent Engineering*, vol. 3, no. 1, S. Dubey, Ed., p. 1 167 990, Dec. 2016, ISSN: null. DOI: 10.1080/23311916.2016.1167990.

[15] S. A. Sarkodie, S. Adams, and T. Leirvik, "Foreign direct investment and renewable energy in climate change mitigation: Does governance matter?" en, *Journal of Cleaner Production*, vol. 263, p. 121 262, Aug. 2020, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2020.121262.

[16] M. D. Bartos and M. V. Chester, "Impacts of climate change on electric power supply in the Western United States," en, *Nature Climate Change*, vol. 5, no. 8, pp. 748–752, Aug. 2015, ISSN: 1758-6798. DOI: 10.1038/nclimate2648.

[17] J. C. Semenza, C. H. Rubin, K. H. Falter, J. D. Selanikio, W. D. Flanders, H. L. Howe, and J. L. Wilhelm, "Heat-related deaths during the July 1995 heat wave in Chicago," English, *The New England Journal of Medicine; Boston*, vol. 335, no. 2, pp. 84–90, Jul. 1996, ISSN: 00284793.

[18] E. Choice, *Electric Choice, 9 of the worst power outages in United States history*, en, https://secure.livechatinc.com/, Apr. 2016.

[19] E.-S. Im, J. S. Pal, and E. A. B. Eltahir, "Deadly heat waves projected in the densely populated agricultural regions of South Asia," *Science Advances*, vol. 3, no. 8, Aug. 2017, ISSN: 2375-2548. DOI: 10.1126/sciadv.1603322.

[20] *Texas Heat Wave, August 2011: Nature and Effects of an Electricity Supply Shortage - Today in Energy - U.S. Energy Information Administration (EIA)*, https://www.eia.gov/todayinenergy/detail.php?id=3010.

[21] J. Creswell and R. Gebeloff, "Traders Profit as Power Grid Is Overworked," en-US, *The New York Times*, Jan. 2018, ISSN: 0362-4331.

[22] O. US EPA, *Climate Change Indicators: Heat-Related Deaths*, en, https://www.epa.gov/climate-indicators/climate-change-indicators-heat-related-deaths, Reports and Assessments, Jul. 2016.

[23] A. B. Smith, *U.S. Billion-dollar Weather and Climate Disasters, 1980 - present (NCEI Accession 0209268)*, 2020. DOI: 10.25921/STKW-7W73.

[24] D. Coumou and S. Rahmstorf, "A decade of weather extremes," en, *Nature Climate Change*, vol. 2, no. 7, pp. 491–496, Jul. 2012, ISSN: 1758-6798. DOI: 10.1038/nclimate1452.

[25] G. A. Meehl and C. Tebaldi, "More intense, more frequent, and longer lasting heat waves in the 21st century," eng, *Science (New York, N.Y.)*, vol. 305, no. 5686, pp. 994–997, Aug. 2004, ISSN: 1095-9203. DOI: 10.1126/science.1098704.

[26] USGCRP, "The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment," en, U.S. Global Change Research Program, Washington, DC, Tech. Rep., Apr. 2016, pp. 1–312.

[27] S. L. Cutter, "Community resilience, natural hazards, and climate change: Is the present a prologue to the future?" *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography*, vol. 74, no. 3, pp. 200–208, May 2020, ISSN: 0029-1951. DOI: 10.1080/00291951.2019.1692066.

[28] D. Barriopedro, E. M. Fischer, J. Luterbacher, R. M. Trigo, and R. García-Herrera, "The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe," en, *Science*, vol. 332, no. 6026, pp. 220–224, Apr. 2011, ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1201224.

[29] *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation — IPCC*.

[30] B. Tonn and J. Eisenberg, "The aging US population and residential energy demand," en, *Energy Policy*, vol. 35, no. 1, pp. 743–745, Jan. 2007, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2005.12.011.

[31] *Why renewables aren't to blame for California's blackouts*, en, https://www.nationalgeographic.com/science/2020/08/why-renewables-arent-reason-california-blackouts/, Aug. 2020.

[32] *California Heatwave Fits a Trend*, en, https://earthobservatory.nasa.gov/images/147256/california-heatwave-fits-a-trend, Text.Article, Sep. 2020.

[33] P. Stott, "How climate change affects extreme weather events," en, *Science*, vol. 352, no. 6293, pp. 1517–1518, Jun. 2016, ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aaf7271.

[34] M. Panteli and P. Mancarella, "Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies," en, *Electric Power Systems Research*, vol. 127, pp. 259–270, Oct. 2015, ISSN: 0378-7796. DOI: 10.1016/j.epsr.2015.06.012.

[35] L. T. Biardeau, L. W. Davis, P. Gertler, and C. Wolfram, "Heat exposure and global air conditioning," en, *Nature Sustainability*, vol. 3, no. 1, pp. 25–28, Jan. 2020, ISSN: 2398-9629. DOI: 10.1038/s41893-019-0441-9.

[36] D. Maia-Silva, R. Kumar, and R. Nateghi, "The critical role of humidity in modeling summer electricity demand across the United States," en, *Nature Communications*, vol. 11, no. 1, p. 1686, Apr. 2020, ISSN: 2041-1723. DOI: 10.1038/s41467-020-15393-8.

[37] X. Liu, M. Shahidehpour, Z. Li, X. Liu, Y. Cao, and Z. Bie, "Microgrids for Enhancing the Power Grid Resilience in Extreme Conditions," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 589–597, Mar. 2017, ISSN: 1949-3061. DOI: 10.1109/TSG.2016.2579999.

[38] R. Arghandeh, M. Brown, A. D. Rosso, G. Ghatikar, E. Stewart, A. Vojdani, and A. von Meier, "The Local Team: Leveraging Distributed Resources to Improve Resilience," *IEEE Power and Energy Magazine*, vol. 12, no. 5, pp. 76–83, Sep. 2014, ISSN: 1558-4216. DOI: 10.1109/MPE.2014.2331902.

[39] *PJM - Load Forecast Development Process*, https://www.pjm.com/planning/resource-adequacy-planning/load-forecast-dev-process.aspx.

[40] S. C. Sherwood, M. Huber, and K. A. Emanuel, "An adaptability limit to climate change due to heat stress," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 21, pp. 9552–9555, 2010, ISSN: 0027-8424.

[41] C. Mora, B. Dousset, I. R. Caldwell, F. E. Powell, R. C. Geronimo, C. R. Bielecki, C. W. W. Counsell, B. S. Dietrich, E. T. Johnston, L. V. Louis, M. P. Lucas, M. M. McKenzie, A. G. Shea, H. Tseng, T. W. Giambelluca, L. R. Leon, E. Hawkins, and C. Trauernicht, "Global risk of deadly heat," en, *Nature Climate Change*, vol. 7, no. 7, pp. 501–506, Jul. 2017, ISSN: 1758-6798. DOI: 10.1038/nclimate3322.

[42] Y. Kwak, D. Seo, C. Jang, and J.-H. Huh, "Feasibility study on a novel methodology for short-term real-time energy demand prediction using weather forecasting data," *Energy and Buildings*, vol. 57, pp. 250–260, Feb. 2013, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2012.10.041.

[43] K. Amarasinghe, D. L. Marino, and M. Manic, “Deep neural networks for energy load forecasting,” in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, Jun. 2017, pp. 1483–1488. DOI: 10.1109/ISIE.2017.8001465.

[44] P. Tüfekci, “Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods,” en, *International Journal of Electrical Power & Energy Systems*, vol. 60, pp. 126–140, Sep. 2014, ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2014.02.027.

[45] H. Lim and Z. J. Zhai, “Review on stochastic modeling methods for building stock energy prediction,” en, *Building Simulation*, vol. 10, no. 5, pp. 607–624, Oct. 2017, ISSN: 1996-8744. DOI: 10.1007/s12273-017-0383-y.

[46] R. Obringer, S. Mukherjee, and R. Nateghi, “Evaluating the climate sensitivity of coupled electricity-natural gas demand using a multivariate framework,” en, *Applied Energy*, vol. 262, p. 114 419, Mar. 2020, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.114419.

[47] M. Auffhammer, P. Baylis, and C. H. Hausman, “Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 8, pp. 1886–1891, Feb. 2017, ISSN: 0027-8424. DOI: 10.1073/pnas.1613193114.

[48] D. J. Sailor and J. R. Muñoz, “Sensitivity of electricity and natural gas consumption to climate in the U.S.A. – Methodology and results for eight states,” *Energy*, vol. 22, no. 10, pp. 987–998, Oct. 1997, ISSN: 0360-5442. DOI: 10.1016/S0360-5442(97)00034-0.

[49] R. Khosla, N. D. Miranda, P. A. Trotter, A. Mazzone, R. Renaldi, C. McElroy, F. Cohen, A. Jani, R. Perera-Salazar, and M. McCulloch, “Cooling for sustainable development,” en, *Nature Sustainability*, pp. 1–8, Oct. 2020, ISSN: 2398-9629. DOI: 10.1038/s41893-020-00627-w.

[50] M. Isaac and D. P. van Vuuren, “Modeling global residential sector energy demand for heating and air conditioning in the context of climate change,” en, *Energy Policy*, vol. 37, no. 2, pp. 507–521, Feb. 2009, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2008.09.051.

[51] D. J. Sailor, “Relating residential and commercial sector electricity loads to climate—evaluating state level sensitivities and vulnerabilities,” en, *Energy*, vol. 26, no. 7, pp. 645–657, Jul. 2001, ISSN: 0360-5442. DOI: 10.1016/S0360-5442(01)00023-8.

[52] L. W. Davis and P. J. Gertler, “Contribution of air conditioning adoption to future energy use under global warming,” en, *Proceedings of the National Academy of Sciences*, vol. 112, no. 19, pp. 5962–5967, May 2015, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1423558112.

[53] L. Xu, S. Wang, and R. Tang, "Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load," en, *Applied Energy*, vol. 237, pp. 180–195, Mar. 2019, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.01.022.

[54] W. S. Jaglom, J. R. McFarland, M. F. Colley, C. B. Mack, B. Venkatesh, R. L. Miller, J. Haydel, P. A. Schultz, B. Perkins, J. H. Casola, J. A. Martinich, P. Cross, M. J. Kolian, and S. Kayin, "Assessment of projected temperature impacts from climate change on the U.S. electric power sector using the Integrated Planning Model.," *Energy Policy*, vol. 73, pp. 524–539, Oct. 2014, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2014.04.032.

[55] J. L. Reyna and M. V. Chester, "Energy efficiency to reduce residential electricity and natural gas use under climate change," en, *Nature Communications*, vol. 8, no. 1, pp. 1–12, May 2017, ISSN: 2041-1723. DOI: 10.1038/ncomms14916.

[56] S. Mukhopadhyay and R. Nateghi, "Estimating climate — Demand Nexus to support longterm adequacy planning in the energy sector," in *2017 IEEE Power Energy Society General Meeting*, Jul. 2017, pp. 1–5. DOI: 10.1109/PESGM.2017.8274648.

[57] R. T. Jerome Friedman Trevor Hastie., *The Elements of Statistical Learning*. Dordrecht, NETHERLANDS, THE: Springer, 2009, ISBN: 978-1-282-82726-4.

[58] A. Padilha-Feltrin, D. A. Q. Rodezno, and J. R. S. Mantovani, "Volt-VAR Multiobjective Optimization to Peak-Load Relief and Energy Efficiency in Distribution Networks," *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 618–626, Apr. 2015, ISSN: 1937-4208. DOI: 10.1109/TPWRD.2014.2336598.

[59] R. Kumar, B. Rachunok, D. Maia-Silva, and R. Nateghi, "Asymmetrical response of California electricity demand to summer-time temperature variation," en, *Scientific Reports*, vol. 10, no. 1, p. 10 904, Jul. 2020, ISSN: 2045-2322. DOI: 10.1038/s41598-020-67695-y.

[60] U.S. Department of Energy, "U. S. Energy sector vulnerabilities to climate change and extreme weather," en, Tech. Rep., 2013, p. 84.

[61] S. Mishra, K. Anderson, B. Miller, K. Boyer, and A. Warren, "Microgrid resilience: A holistic approach for assessing threats, identifying vulnerabilities, and designing corresponding mitigation strategies," en, *Applied Energy*, vol. 264, p. 114 726, Apr. 2020, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2020.114726.

[62] M. A. Brown, M. Cox, B. Staver, and P. Baer, "Climate Change and Energy Demand in Buildings," en, p. 13,

[63] M. A. Brown, M. Cox, B. Staver, and P. Baer, "Modeling climate-driven changes in U.S. buildings energy demand," en, *Climatic Change*, vol. 134, no. 1, pp. 29–44, Jan. 2016, ISSN: 1573-1480. DOI: 10.1007/s10584-015-1527-7.

[64] R. A. Pielke, C. Davey, and J. Morgan, "Assessing "global warming" with surface heat content," en, *Eos, Transactions American Geophysical Union*, vol. 85, no. 21, pp. 210–211, 2004, ISSN: 2324-9250. DOI: 10.1029/2004EO210004.

[65] J. Xie, Y. Chen, T. Hong, and T. D. Laing, "Relative Humidity for Load Forecasting Models," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 191–198, Jan. 2018, ISSN: 1949-3061. DOI: 10.1109/TSG.2016.2547964.

[66] E. D. Coffel, R. M. Horton, and A. de Sherbinin, "Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century," en, *Environmental Research Letters*, vol. 13, no. 1, p. 014 001, Dec. 2017, ISSN: 1748-9326. DOI: 10.1088/1748-9326/aaa00e.

[67] J. R. Buzan, K. Oleson, and M. Huber, "Implementation and comparison of a suite of heat stress metrics within the Community Land Model version 4.5," English, *Geoscientific Model Development; Katlenburg-Lindau*, vol. 8, no. 2, p. 151, 2015, ISSN: 1991962X. DOI: http://dx.doi.org/10.5194/gmd-8-151-2015.

[68] M. Waite, E. Cohen, H. Torbey, M. Piccirilli, Y. Tian, and V. Modi, "Global trends in urban electricity demands for cooling and heating," en, *Energy*, vol. 127, pp. 786–802, May 2017, ISSN: 0360-5442. DOI: 10.1016/j.energy.2017.03.095.

[69] M. Sivak, "Potential energy demand for cooling in the 50 largest metropolitan areas of the world: Implications for developing countries," en, *Energy Policy*, vol. 37, no. 4, pp. 1382–1384, Apr. 2009, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2008.11.031.

[70] Y. Petri and K. Caldeira, "Impacts of global warming on residential heating and cooling degree-days in the United States," en, *Scientific Reports*, vol. 5, no. 1, p. 12 427, Aug. 2015, ISSN: 2045-2322. DOI: 10.1038/srep12427.

[71] B. Goldstein, D. Gounaridis, and J. P. Newell, "The carbon footprint of household energy use in the United States," en, *Proceedings of the National Academy of Sciences*, vol. 117, no. 32, pp. 19 122–19 130, Aug. 2020, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1922205117.

[72] S. Mukherjee, R. Nateghi, and M. Hastak, "A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S.," *Reliability Engineering & System Safety*, vol. 175, pp. 283–305, 2018. DOI: 10.1016/j.ress.2018.03.015.

[73] B. J. v. Ruijven, E. D. Cian, and I. S. Wing, "Amplification of future energy demand growth due to climate change," *Nature Communications*, vol. 10, no. 1, p. 2762, 2019. DOI: 10.1038/s41467-019-10399-3.

[74] D. Coumou and S. Rahmstorf, "A decade of weather extremes," *Nature Climate Change*, vol. 2, no. 7, pp. 491–496, 2012. DOI: 10.1038/nclimate1452.

[75] G. A. Meehl and C. Tebaldi, "More intense, more frequent, and longer lasting heat waves in the 21st century," *Science*, vol. 305, no. 5686, pp. 994–997, 2004. DOI: 10.1126/science.1098704.

[76] P. Alipour, S. Mukherjee, and R. Nateghi, "Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the Texas region," *Energy*, vol. 185, pp. 1143–1153, 2019. DOI: 10.1016/j.energy.2019.07.074.

[77] U.S. Department of Energy, *U.S. Energy sector vulnerabilities to climate change and extreme weather*. [Online]. Available: https://apps.dtic.mil/dtic/tr/fulltext/u2/a583709.pdf.

[78] J. C. Semenza, C. H. Rubin, K. H. Falter, J. D. Selanikio, W. D. Flanders, H. L. Howe, and J. L. Wilhelm, "Heat-related deaths during the July 1995 heat wave in Chicago," *The New England Journal of Medicine; Boston*, vol. 335, no. 2, pp. 84–90, 1996.

[79] *Electric Choice, 9 of the worst power outages in United States history*. [Online]. Available: https://www.electricchoice.com/blog/worst-power-outages-in-united-states-history/.

[80] C. Mora, B. Dousset, I. R. Caldwell, F. E. Powell, R. C. Geronimo, C. R. Bielecki, C. W. W. Counsell, B. S. Dietrich, E. T. Johnston, L. V. Louis, M. P. Lucas, M. M. McKenzie, A. G. Shea, H. Tseng, T. W. Giambelluca, L. R. Leon, E. Hawkins, and C. Trauernicht, "Global risk of deadly heat," *Nature Climate Change*, vol. 7, no. 7, pp. 501–506, 2017. DOI: 10.1038/nclimate3322.

[81] E.-S. Im, J. S. Pal, and E. A. B. Eltahir, "Deadly heat waves projected in the densely populated agricultural regions of South Asia," *Science Advances*, vol. 3, no. 8, 2017. DOI: 10.1126/sciadv.1603322.

[82] U.S. Energy Information Administration, *Texas heat wave, august 2011: Nature and effects of an electricity supply*. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=3010.

[83] J. Creswell and R. Gebeloff, "Traders Profit as Power Grid Is Overworked," *The New York Times*, 2018. [Online]. Available: https://www.nytimes.com/2014/08/15/business/energy-environment/traders-profit-as-power-grid-is-overworked.html.

[84] D. J. Sailor, A. Baniassadi, C. R. O'Lenick, and O. V. Wilhelmi, "The growing threat of heat disasters," *Environmental Research Letters*, vol. 14, no. 5, p. 054 006, 2019. DOI: 10.1088/1748-9326/ab0bb9.

[85] S. Mukhopadhyay and R. Nateghi, "Estimating climate - Demand Nexus to support longterm adequacy planning in the energy sector," vol. 29, pp. 1–5, 2018. DOI: 10.1109/PESGM.2017.8274648.

[86] R. Nateghi and S. Mukherjee, "A multi-paradigm framework to assess the impacts of climate change on end-use energy demand," *PLoS ONE*, vol. 12(11), no. 11, e0188033, 2017. DOI: 10.1371/journal.pone.0188033.

[87] S. Mukherjee, C. R. Vineeth, and R. Nateghi, "Evaluating regional climate-electricity demand nexus: A composite Bayesian predictive framework," *Applied Energy*, vol. 235, pp. 1561–1582, 2019. DOI: 10.1016/j.apenergy.2018.10.119.

[88] E. De Cian and I. Sue Wing, "Global Energy Consumption in a Warming Climate," *Environmental and Resource Economics*, vol. 72, no. 2, pp. 365–410, 2019.

[89] S. C. Sherwood, M. Huber, and K. A. Emanuel, "An adaptability limit to climate change due to heat stress," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 21, pp. 9552–9555, 2010.

[90] E. D. Coffel, R. M. Horton, and A. d. Sherbinin, "Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century," *Environmental Research Letters*, vol. 13, no. 1, p. 014 001, 2017. DOI: 10.1088/1748-9326/aaa00e.

[91] Y. Kwak, D. Seo, C. Jang, and J.-H. Huh, "Feasibility study on a novel methodology for short-term real-time energy demand prediction using weather forecasting data," *Energy and Buildings*, vol. 57, pp. 250–260, 2013. DOI: 10.1016/j.enbuild.2012.10.041.

[92] *PJM - Load forecast development process*. [Online]. Available: https://www.pjm.com/planning/resource-adequacy-planning/load-forecast-dev-process.aspx.

[93] F. Apadula, A. Bassini, A. Elli, and S. Scapin, "Relationships between meteorological variables and monthly electricity demand," *Applied Energy*, vol. 98, pp. 346–356, 2012.

[94] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, D. P. Lalas, M. Moschovits, F. Karagiannis, and D. Papakonstantinou, "Models for mid-term electricity demand forecasting incorporating weather influences," *Energy*, vol. 31, no. 2, pp. 208–227, 2006.

[95] R. A. Pielke, C. Davey, and J. Morgan, "Assessing global warming with surface heat content," *Eos, Transactions American Geophysical Union*, vol. 85, no. 21, pp. 210–211, 2004. DOI: 10.1029/2004EO210004.

[96] M. Auffhammer, P. Baylis, and C. H. Hausman, "Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 8, pp. 1886–1891, 2017. DOI: 10.1073/pnas.1613193114.

[97] W. S. Jaglom, J. R. McFarland, M. F. Colley, C. B. Mack, B. Venkatesh, R. L. Miller, J. Haydel, P. A. Schultz, B. Perkins, J. H. Casola, J. A. Martinich, P. Cross, M. J. Kolian, and S. Kayin, "Assessment of projected temperature impacts from climate change on the U.S. electric power sector using the Integrated Planning Model.," *Energy Policy*, vol. 73, pp. 524–539, 2014. DOI: 10.1016/j.enpol.2014.04.032.

[98] J. L. Reyna and M. V. Chester, "Energy efficiency to reduce residential electricity and natural gas use under climate change," *Nature Communications*, vol. 8, no. 1, pp. 1–12, 2017.

[99] J. R. Buzan, K. Oleson, and M. Huber, "Implementation and comparison of a suite of heat stress metrics within the Community Land Model version 4.5," *Geoscientific Model Development; Katlenburg-Lindau*, vol. 8, no. 2, p. 151, 2015. DOI: http://dx.doi.org/10.5194/gmd-8-151-2015.

[100] R. Obringer, R. Kumar, and R. Nateghi, "Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States," *Applied Energy*, vol. 252, p. 113 466, 2019. DOI: 10.1016/j.apenergy.2019.113466.

[101] S. Mukherjee and R. Nateghi, "A Data-Driven Approach to Assessing Supply Inadequacy Risks Due to Climate-Induced Shifts in Electricity Demand," *Risk Analysis*, vol. 39, no. 0, pp. 673–694, 2018. DOI: 10.1111/risa.13192.

[102] L. Raymond, D. Gotham, W. McClain, S. Mukherjee, R. Nateghi, P. V. Preckel, P. Schubert, S. Singh, and E. Wachs, "Projected climate change impacts on Indiana's energy demand and supply," *Climatic Change*, 2019. DOI: 10.1007/s10584-018-2299-7.

[103] O. Deschênes and M. Greenstone, "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US," *American Economic Journal. Applied Economics*, vol. 3, no. 4, pp. 152–185, 2011. DOI: http://dx.doi.org/10.1257/app.3.4.152.

[104] R. Obringer, R. Kumar, and R. Nateghi, “Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States,” en, *Applied Energy*, vol. 252, p. 113 466, Oct. 2019, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.113466.

[105] S. W. Hadley, D. J. Erickson, J. L. Hernandez, C. T. Broniak, and T. J. Blasing, “Responses of energy use to climate change: A climate modeling study,” *Geophysical Research Letters*, vol. 33, no. 17, 2006. DOI: 10.1029/2006GL026652.

[106] F. Mesinger, G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, and W. Shi, “North american regional reanalysis,” *Bulletin of the American Meteorological Society*, vol. 87, no. 3, pp. 343–360, 2006. DOI: 10.1175/BAMS-87-3-343.

[107] D. J. Sailor and J. R. Muñoz, “Sensitivity of electricity and natural gas consumption to climate in the U.S.A. – Methodology and results for eight states,” *Energy*, vol. 22, no. 10, pp. 987–998, 1997. DOI: 10.1016/S0360-5442(97)00034-0.

[108] H. A. Chipman, E. I. George, and R. E. McCulloch, “BART: Bayesian additive regression trees,” *The Annals of Applied Statistics*, vol. 4, no. 1, pp. 266–298, 2010. DOI: 10.1214/09-AOAS285.

[109] S. Mukherjee and R. Nateghi, “Climate sensitivity of end-use electricity consumption in the built environment: An application to the state of Florida, United States,” *Energy*, vol. 128, pp. 688–700, 2017. DOI: 10.1016/j.energy.2017.04.034.

[110] *Electric power sales, revenue, and energy efficiency Form EIA-861 detailed data files*. [Online]. Available: https://www.eia.gov/electricity/data/eia861/.

[111] *Population and Housing Unit Estimates Datasets*. [Online]. Available: https://www.census.gov/programs-surveys/popest/data/data-sets.html.

[112] M. Bessec and J. Fouquau, “The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach,” *Energy Economics*, vol. 30, no. 5, pp. 2705–2721, 2008.

[113] *NCEP North American Regional Reanalysis (NARR)*. [Online]. Available: https://www.esrl.noaa.gov/psd/data/gridded/data.narr.html.

[114] *Center for International Earth Science Information Network*. [Online]. Available: http://www.ciesin.org/data.html.

[115] M. G. Lawrence, "The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications," *Bulletin of the American Meteorological Society*, vol. 86, no. 2, pp. 225–233, 2005.

[116] R. Davies-Jones, "An Efficient and Accurate Method for Computing the Wet-Bulb Temperature along Pseudoadiabats," *Monthly Weather Review; Washington*, vol. 136, no. 7, pp. 2764–2785, 2008.

[117] E. C. Thom, "The Discomfort Index," *Weatherwise*, vol. 12, no. 2, pp. 57–61, 1959. DOI: 10.1080/00431672.1959.9926960.

[118] Y. Epstein and D. S. Moran, "Thermal Comfort and the Heat Stress Indices," *Industrial Health*, vol. 44, no. 3, pp. 388–398, 2006. DOI: 10.2486/indhealth.44.388.

[119] *Heat Stress Control and Heat Casualty Management*, ser. Army TB ; MED 507. Washington, DC: Headquarters, Dept. of the Army and Air Force, 2003.

[120] J. M. Masterton and F. A. Richardson, *Humidex ; A Method of Quantifying Human Discomfort Due to Excessive Heat and Humidity*. Downsview, Ont.: Environment Canada, Atmospheric Environment, 1979.

[121] L. P. Rothfusz, "The heat index equation," *National Weather Service Technical Attachment (SR 90–23)*, 1990.

[122] *Fifth Assessment Report – IPCC*. [Online]. Available: https://www.ipcc.ch/assessment-report/ar5/.

[123] L. Warszawski, K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe, "The Inter-Sectoral Impact Model Intercomparison Project (ISI – MIP): Project framework," *Proceedings of the National Academy of Sciences*, vol. 111, no. 9, pp. 3228–3232, 2014. DOI: 10.1073/pnas.1312330110.

[124] S. Hempel, K. Frieler, L. Warszawski, J. Schewe, and F. Piontek, "A trend-preserving bias correction – The ISI-MIP approach," *Earth System Dynamics*, vol. 4, no. 2, pp. 219–236, 2013. DOI: 10.5194/esd-4-219-2013.

[125] R. T. Jerome Friedman Trevor Hastie., *The Elements of Statistical Learning*. Dordrecht, The Netherlands: Springer, 2009, ISBN: 978-1-282-82726-4.

[126] A. Kapelner and J. Bleich, "bartMachine: Machine Learning with Bayesian Additive Regression Trees," *Journal of Statistical Software*, 2013.

[127] Energy Information Administration, *Residential Energy Consumption Survey (RECS)*. [Online]. Available: https://www.eia.gov/consumption/residential/index.php.

[128] Energy Department, "Residential Average Monthly kWh and Bills: Residential Average Monthly kWh and Bills," 3KB, 2016. DOI: 10.26000/040.000001. [Online]. Available: https://data.austintexas.gov/d/d9pb-3vh7.

[129] *U.S. Census Bureau QuickFacts: Austin city, Texas*. [Online]. Available: https://www.census.gov/quickfacts/fact/table/austincitytexas/PST045218.

[130] *San Francisco, CA Electricity Rates*. [Online]. Available: https://www.electricitylocal.com/states/california/san-francisco/.

[131] M. R. Allen and W. J. Ingram, "Constraints on future changes in climate and the hydrologic cycle," *Nature*, vol. 419, no. 6903, pp. 228–232, 2002. DOI: 10.1038/nature01092.

[132] A. Levesque, R. C. Pietzcker, L. Baumstark, S. De Stercke, A. Grübler, and G. Luderer, "How much energy will buildings consume in 2100? A global perspective within a scenario framework," *Energy*, vol. 148, pp. 514–527, 2018. DOI: 10.1016/j.energy.2018.01.139.

[133] American Society of Civil Engineer (ASCE), *ASCE 2017 Infrastructure Report Card*. [Online]. Available: https://www.infrastructurereportcard.org/cat-item/energy/.

[134] Energy Information Administration, *EIA - Annual Energy Outlook 2019*. [Online]. Available: https://www.eia.gov/outlooks/aeo/.

[135] *Global energy transformation: A roadmap to 2050 (2019 edition)*. [Online]. Available: /publications/2019/Apr/Global-energy-transformation-A-roadmap-to-2050-2019Edition.

[136] M. D. Bartos and M. V. Chester, "Impacts of climate change on electric power supply in the Western United States," *Nature Climate Change*, vol. 5, no. 8, pp. 748–752, 2015. DOI: 10.1038/nclimate2648.

[137] J. E. Mu, J. M. Antle, and J. T. Abatzoglou, "Representative agricultural pathways, climate change, and agricultural land uses: An application to the Pacific Northwest of the USA," *Mitigation and Adaptation Strategies for Global Change*, vol. 24, no. 5, pp. 819–837, 2019. DOI: 10.1007/s11027-018-9834-8.

[138] R. Khosla, N. D. Miranda, P. A. Trotter, A. Mazzone, R. Renaldi, C. McElroy, F. Cohen, A. Jani, R. Perera-Salazar, and M. McCulloch, "Cooling for sustainable development," *Nature Sustainability*, pp. 1–8, 2020.

[139] G. R. McGregor, P. Bessmoulin, K. Ebi, and B. Menne, *Heatwaves and health: guidance on warning-system development*. WMOP, 2015.

[140] F. Birol, “The future of cooling: Opportunities for energy-efficient air conditioning,” *International Energy Agency*, 2018.

[141] M. W. Berger, *Cool comfort: America’s romance with air-conditioning*, 2004.

[142] S. Mukherjee and R. Nateghi, “A Data-Driven Approach to Assessing Supply Inadequacy Risks Due to Climate-Induced Shifts in Electricity Demand,” en, *Risk Analysis*, vol. 0, no. 0, 2017, ISSN: 1539-6924. DOI: 10.1111/risa.13192.

[143] IEA, *The Future of Cooling – Analysis*, en-GB, https://www.iea.org/reports/the-future-of-cooling, 2008.

[144] R. Obringer, R. Kumar, and R. Nateghi, “Analyzing the climate sensitivity of the coupled water-electricity demand nexus in the Midwestern United States,” *Applied Energy*, vol. 252, p. 113 466, Oct. 2019, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.113466.

[145] R. Obringer, S. Mukherjee, and R. Nateghi, “Evaluating the climate sensitivity of coupled electricity-natural gas demand using a multivariate framework,” en, *Applied Energy*, vol. 262, p. 114 419, Mar. 2020, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.114419.

[146] T. Day, *Degree-days : theory and application - TM41: 2006*, eng. 2006, ISBN: 1-903287-76-6.

[147] Z. Khan, G. Iyer, P. Patel, S. Kim, M. Hejazi, C. Burleyson, and M. Wise, “Impacts of long-term temperature change and variability on electricity investments,” en, *Nature Communications*, vol. 12, no. 1, p. 1643, Mar. 2021, ISSN: 2041-1723. DOI: 10.1038/s41467-021-21785-1.

[148] M. Shin and S. L. Do, “Prediction of cooling energy use in buildings using an enthalpy-based cooling degree days method in a hot and humid climate,” en, *Energy and Buildings*, vol. 110, pp. 57–70, Jan. 2016, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2015.10.035.

[149] D. Li, J. Yuan, and R. E. Kopp, “Escalating global exposure to compound heat-humidity extremes with warming,” en, *Environmental Research Letters*, vol. 15, no. 6, p. 064 003, May 2020, ISSN: 1748-9326. DOI: 10.1088/1748-9326/ab7d04.

[150] C. Raymond, T. Matthews, and R. M. Horton, “The emergence of heat and humidity too severe for human tolerance,” en, *Science Advances*, vol. 6, no. 19, eaaw1838, May 2020, ISSN: 2375-2548. DOI: 10.1126/sciadv.aaw1838.

[151] G. C. Hulley, B. Dousset, and B. H. Kahn, “Rising Trends in Heatwave Metrics Across Southern California,” en, *Earth’s Future*, vol. 8, no. 7, e2020EF001480, 2020, ISSN: 2328-4277. DOI: 10.1029/2020EF001480.

[152] Anderson G. Brooke, Bell Michelle L., and Peng Roger D., "Methods to Calculate the Heat Index as an Exposure Metric in Environmental Health Research," *Environmental Health Perspectives*, vol. 121, no. 10, pp. 1111–1119, Oct. 2013. DOI: 10.1289/ehp.1206273.

[153] K. M. Willett and S. Sherwood, "Exceedance of heat index thresholds for 15 regions under a warming climate using the wet-bulb globe temperature," en, *International Journal of Climatology*, vol. 32, no. 2, pp. 161–177, 2012, ISSN: 1097-0088. DOI: 10.1002/joc.2257.

[154] W. Schlenker and M. J. Roberts, "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change," *Proceedings of the National Academy of Sciences*, vol. 106, no. 37, pp. 15 594–15 598, Sep. 15, 2009, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0906865106. pmid: 19717432. [Online]. Available: https://www.pnas.org/content/106/37/15594.

[155] R. Obringer, R. Kumar, and R. Nateghi, "Managing the water–electricity demand nexus in a warming climate," *Climatic Change*, vol. 159, no. 2, pp. 233–252, 2020.

[156] EIA, *Annual Energy Outlook 2020*, https://www.eia.gov/outlooks/aeo/, 2020.

[157] IPCC, *Fifth Assessment Report – IPCC*, 2014.

[158] EIA, *Florida - State Energy Profile Overview - U.S. Energy Information Administration (EIA)*, https://www.eia.gov/state/?sid=FL, 2020.

[159] M. Khoshbakht, Z. Gou, and K. Dupre, "Energy use characteristics and benchmarking for higher education buildings," en, *Energy and Buildings*, vol. 164, pp. 61–76, Apr. 2018, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2018.01.001.

[160] D. M. Santágata, P. Castesana, C. E. Rössler, and D. R. Gómez, "Extreme temperature events affecting the electricity distribution system of the metropolitan area of Buenos Aires (1971–2013)," en, *Energy Policy*, vol. 106, pp. 404–414, Jul. 2017, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2017.04.006.

[161] S. C. Parkinson and N. Djilali, "Robust response to hydro-climatic change in electricity generation planning," en, *Climatic Change*, vol. 130, no. 4, pp. 475–489, Jun. 2015, ISSN: 1573-1480. DOI: 10.1007/s10584-015-1359-5.

[162] P. Alipour, S. Mukherjee, and R. Nateghi, "Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the texas region," *Energy*, vol. 185, pp. 1143–1153, 2019.

[163] F. R. Jacobsen, "Energy signature and energy monitoring in building energy management systems," in *Proceeding of CLIMA 2000 world congress*, vol. 3, 1985, pp. 25–31.

[164] M. Bhatnagar, J. Mathur, and V. Garg, "Determining base temperature for heating and cooling degree-days for India," en, *Journal of Building Engineering*, vol. 18, pp. 270–280, Jul. 2018, ISSN: 2352-7102. DOI: 10.1016/j.jobe.2018.03.020.

[165] K. Lee, H.-J. Baek, and C. Cho, "The Estimation of Base Temperature for Heating and Cooling Degree-Days for South Korea," *Journal of Applied Meteorology and Climatology*, vol. 53, no. 2, pp. 300–309, Sep. 2013, ISSN: 1558-8424. DOI: 10.1175/JAMC-D-13-0220.1.

[166] L. P. Rothfusz, "The heat index equation," *National Weather Service Technical Attachment (SR 90–23)*, 1990.

[167] H. S. Laine, J. Salpakari, E. E. Looney, H. Savin, I. Marius Peters, and T. Buonassisi, "Meeting global cooling demand with photovoltaics during the 21st century," en, *Energy & Environmental Science*, vol. 12, no. 9, pp. 2706–2716, 2019. DOI: 10.1039/C9EE00002J .

[168] M. Jakubcionis and J. Carlsson, "Estimation of European Union residential sector space cooling potential," en, *Energy Policy*, vol. 101, pp. 225–235, Feb. 2017, ISSN: 0301-4215. DOI: 10.1016/j.enpol.2016.11.047.

[169] F. Jacobsen, "Energy signature and energy monitoring in building energy management systems," *Proceeding of CLIMA 2000 world congress*, vol. 3, 1985.

[170] R. Zmeureanu and G. Renaud, "Estimation of potential impact of climate change on the heating energy use of existing houses," *Energy Policy*, vol. 36, no. 1, pp. 303–310, 2008, ISSN: 0301-4215.

[171] EIA, *Electric power sales, revenue, and energy efficiency Form EIA-861 detailed data files*, https://www.eia.gov/electricity/data/eia861/, 2020.

[172] NARR, *ESRL : PSD : NCEP North American Regional Reanalysis (NARR)*, https://www.esrl.noaa.gov ded/data.narr.html, 2020.

[173] Y. Petri and K. Caldeira, "Impacts of global warming on residential heating and cooling degree-days in the United States," en, *Scientific Reports*, vol. 5, no. 1, p. 12 427, Aug. 2015, ISSN: 2045-2322. DOI: 10.1038/srep12427.

[174] A. F. Hamlet, S.-Y. Lee, K. E. B. Mickelson, and M. M. Elsner, "Effects of projected climate change on energy supply and demand in the Pacific Northwest and Washington State," en, *Climatic Change*, vol. 102, no. 1, pp. 103–128, Sep. 2010, ISSN: 1573-1480. DOI: 10.1007/s10584-010-9857-y.

[175] U. C. Bureau, *State Population Totals: 2010-2019*, EN-US, https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html, 2020.

[176] Z. Mehrabi, S. Donner, P. Rios, D. Guha-Sapir, P. Rowhani, M. Kandlikar, and N. Ramankutty, "Can we sustain success in reducing deaths to extreme weather in a hotter world?" en, *World Development Perspectives*, vol. 14, p. 100 107, Jun. 2019, ISSN: 2452-2929. DOI: 10.1016/j.wdp.2019.02.018.

[177] F. Creutzig, J. Roy, W. F. Lamb, I. M. L. Azevedo, W. Bruine de Bruin, H. Dalkmann, O. Y. Edelenbosch, F. W. Geels, A. Grubler, C. Hepburn, E. G. Hertwich, R. Khosla, L. Mattauch, J. C. Minx, A. Ramakrishnan, N. D. Rao, J. K. Steinberger, M. Tavoni, D. Ürge-Vorsatz, and E. U. Weber, "Towards demand-side solutions for mitigating climate change," en, *Nature Climate Change*, vol. 8, no. 4, pp. 260–263, Apr. 2018, ISSN: 1758-6798. DOI: 10.1038/s41558-018-0121-1.

[178] S. C. Fu, X. L. Zhong, Y. Zhang, T. W. Lai, K. C. Chan, K. Y. Lee, and C. Y. H. Chao, "Bio-inspired cooling technologies and the applications in buildings," en, *Energy and Buildings*, vol. 225, p. 110 313, Oct. 2020, ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2020.110313.

[179] A. De Angelis, O. Saro, and M. Truant, "Evaporative cooling systems to improve internal comfort in industrial buildings," en, *Energy Procedia*, ATI 2017 - 72nd Conference of the Italian Thermal Machines Engineering Association, vol. 126, pp. 313–320, Sep. 2017, ISSN: 1876-6102. DOI: 10.1016/j.egypro.2017.08.245.

[180] X. Nie, Y. Yoo, H. Hewakuruppu, J. Sullivan, A. Krishna, and J. Lee, "Cool White Polymer Coatings based on Glass Bubbles for Buildings," en, *Scientific Reports*, vol. 10, no. 1, p. 6661, Apr. 2020, ISSN: 2045-2322. DOI: 10.1038/s41598-020-63027-2.

[181] L. Bounoua, P. Zhang, G. Mostovoy, K. Thome, J. Masek, M. Imhoff, M. Shepherd, D. Quattrochi, J. Santanello, J. Silva, R. Wolfe, and A. M. Toure, "Impact of urbanization on US surface climate," en, *Environmental Research Letters*, vol. 10, no. 8, p. 084 010, Aug. 2015, ISSN: 1748-9326. DOI: 10.1088/1748-9326/10/8/084010.

[182] T. Susca, S. R. Gaffin, and G. R. Dell'Osso, "Positive effects of vegetation: Urban heat island and green roofs," en, *Environmental Pollution*, Selected Papers from the Conference Urban Environmental Pollution: Overcoming Obstacles to Sustainability and Quality of Life (UEP2010), 20-23 June 2010, Boston, USA, vol. 159, no. 8, pp. 2119–2126, Aug. 2011, ISSN: 0269-7491. DOI: 10.1016/j.envpol.2011.03.007.

[183] E. K. Melaas, J. A. Wang, D. L. Miller, and M. A. Friedl, "Interactions between urban vegetation and surface urban heat islands: A case study in the Boston metropolitan region," en, *Environmental Research Letters*, vol. 11, no. 5, p. 054 020, May 2016, ISSN: 1748-9326. DOI: 10.1088/1748-9326/11/5/054020.

[184] G. Erdal, H. Erdal, and K. Esengün, "The causality between energy consumption and economic growth in Turkey," en, *Energy Policy*, vol. 36, no. 10, pp. 3838–3842, 2008.

[185] D. Popp, "The effect of new technology on energy consumption," *Resource and Energy Economics*, vol. 23, pp. 215–239, Feb. 2001. DOI: 10.1016/S0928-7655(00)00045-2.

[186] D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 163–169, Feb. 2000, ISSN: 1558-2256. DOI: 10.1109/5.823996.

[187] G. R. McGregor, P. Bessemoulin, K. L. Ebi, B. Menne, World Meteorological Organization, and World Health Organization, *Heatwaves and Health: Guidance on Warning-System Development*, en. 2015.

[188] *California - State Energy Profile Analysis - U.S. Energy Information Administration (EIA)*, https://www.eia.gov/state/analysis.php?sid=CA.

[189] E. Cian and I. S. Wing, "Global Energy Consumption in a Warming Climate," en, *Environmental & Resource Economics*, vol. 72, no. 2, pp. 365–410, 2019.

[190] L. Wenz, A. Levermann, and M. Auffhammer, "North-south polarization of European electricity consumption under future warming," eng, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 38, E7910–E7918, Sep. 2017, ISSN: 1091-6490. DOI: 10.1073/pnas.1704339114.

[191] M. Auffhammer and A. Aroonruengsawat, "Simulating the impacts of climate change, prices and population on California's residential electricity consumption," *Climatic Change*, vol. 109, no. 1, pp. 191–210, 2011, ISSN: 0165-0009.

[192] *Homepage - U.S. Energy Information Administration (EIA)*, https://www.eia.gov/index.php.

[193] J. Nagi, K. S. Yap, F. Nagi, S. K. Tiong, and S. K. Ahmed, "A computational intelligence scheme for the prediction of the daily peak load," en, *Applied Soft Computing*, vol. 11, no. 8, pp. 4773–4788, Dec. 2011, ISSN: 1568-4946. DOI: 10.1016/j.asoc.2011.07.005.

[194] C. Fan, F. Xiao, and Y. Zhao, "A short-term building cooling load prediction method using deep learning algorithms," en, *Applied Energy*, vol. 195, pp. 222–233, Jun. 2017, ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2017.03.064.

[195] R. J. Hyndman and S. Fan, "Density Forecasting for Long-Term Peak Electricity Demand," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142–1153, May 2010, ISSN: 1558-0679. DOI: 10.1109/TPWRS.2009.2036017.

[196] *Electricity Data - U.S. Energy Information Administration (EIA)*, https://www.eia.gov/electricity/data.php.

[197] H. A. Chipman, E. I. George, and R. E. McCulloch, "BART: Bayesian additive regression trees," EN, *The Annals of Applied Statistics*, vol. 4, no. 1, pp. 266–298, Mar. 2010, ISSN: 1932-6157, 1941-7330. DOI: 10.1214/09-AOAS285.

[198] A. Kapelner and J. Bleich, "bartMachine: Machine Learning with Bayesian Additive Regression Trees," *arXiv:1312.2171 [cs, stat]*, Dec. 2013. arXiv: 1312.2171 `[cs, stat]`.

[199] *Center for International Earth Science Information Network*, http://www.ciesin.org/data.html.

[200] M. G. Lawrence, "The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications," *Bulletin of the American Meteorological Society*, vol. 86, no. 2, pp. 225–233, 2005, ISSN: 0003-0007.

[201] R. Davies-Jones, "An Efficient and Accurate Method for Computing the Wet-Bulb Temperature along Pseudoadiabats," English, *Monthly Weather Review; Washington*, vol. 136, no. 7, pp. 2764–2785, Jul. 2008, ISSN: 00270644.

[202] E. C. Thom, "The Discomfort Index," *Weatherwise*, vol. 12, no. 2, pp. 57–61, Apr. 1959, ISSN: 0043-1672. DOI: 10.1080/00431672.1959.9926960.

[203] Y. Epstein and D. S. Moran, "Thermal Comfort and the Heat Stress Indices," *Industrial Health*, vol. 44, no. 3, pp. 388–398, 2006. DOI: 10.2486/indhealth.44.388.

[204] J. M. Masterton and F. A. Richardson, *Humidex ; a Method of Quantifying Human Discomfort Due to Excessive Heat and Humidity*, English. Downsview, Ont.: Environment Canada, Atmospheric Environment, 1979.

[205] S. Gelper, R. Fried, and C. Croux, "Robust forecasting with exponential and Holt–Winters smoothing," en, *Journal of Forecasting*, vol. 29, no. 3, pp. 285–300, 2010, ISSN: 1099-131X. DOI: 10.1002/for.1125.

[206] Y. Mohammed, "Prediction intervals for the Holt-Winters forecasting procedure," en, *International Journal of Forecasting*, vol. 6, no. 1, pp. 127–137, Jan. 1990, ISSN: 0169-2070. DOI: 10.1016/0169-2070(90)90103-I.

[207] L. Ferbar Tratar and E. Strmčnik, "The comparison of Holt–Winters method and Multiple regression method: A case study," en, *Energy*, vol. 109, pp. 266–276, Aug. 2016, ISSN: 0360-5442. DOI: 10.1016/j.energy.2016.04.115.

[208] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," en, *International Journal of Forecasting*, vol. 20, no. 1, pp. 5–10, Jan. 2004, ISSN: 01692070. DOI: 10.1016/j.ijforecast.2003.09.015.

[209] P. R. Winters, "Forecasting Sales by Exponentially Weighted Moving Averages," en, *Management Science*, vol. 6, no. 3, pp. 324–342, Apr. 1960, ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.6.3.324.

[210] *Maps*, http://www.ercot.com/news/mediakit/maps.

[211] E. Papadis and G. Tsatsaronis, "Challenges in the decarbonization of the energy sector," en, *Energy*, vol. 205, p. 118 025, Aug. 2020, ISSN: 0360-5442. DOI: 10.1016/j.energy.2020.118025.

[212] K. Surana and S. M. Jordaan, "The climate mitigation opportunity behind global power transmission and distribution," en, *Nature Climate Change*, vol. 9, no. 9, pp. 660–665, Sep. 2019, ISSN: 1758-6798. DOI: 10.1038/s41558-019-0544-3.

[213] A. M. van Valkengoed and L. Steg, "Meta-analyses of factors motivating climate change adaptation behaviour," en, *Nature Climate Change*, vol. 9, no. 2, pp. 158–163, Feb. 2019, ISSN: 1758-6798. DOI: 10.1038/s41558-018-0371-y.

[214] R. S. Wilson, A. Herziger, M. Hamilton, and J. S. Brooks, "From incremental to transformative adaptation in individual responses to climate-exacerbated hazards," en, *Nature Climate Change*, vol. 10, no. 3, pp. 200–208, Mar. 2020, ISSN: 1758-6798. DOI: 10.1038/s41558-020-0691-6.

[215] *PG&E, Pacific Gas and Electric - Gas and power company for California*, https://www.pge.com/.

[216] *Oldest Texas electricity co-op goes bust after getting hit with $2 billion bill*, en-US, https://www.cbsnews.com/news/brazos-electric-power-cooperative-bankruptcy-texas/.

[217] A. B. C. News, *Texan files $1 billion class-action lawsuit after receiving $9,000 electric bill*, en, https://abcnews.go.com/US/texan-files-billion-class-action-lawsuit-receiving-9000/story?id=7

[218] *ROSE-package: ROSE: Random Over-Sampling Examples in ROSE: ROSE: Random Over-Sampling Examples*, en, https://rdrr.io/cran/ROSE/man/ROSE-package.html.

[219] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," en, *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, ISSN: 1076-9757. DOI: 10.1613/jair.953.

[220] K. Bogner, F. Pappenberger, and M. Zappa, "Machine Learning Techniques for Predicting the Energy Consumption/Production and Its Uncertainties Driven by Meteorological Observations and Forecasts," en, *Sustainability*, vol. 11, no. 12, p. 3328, Jan. 2019. DOI: 10.3390/su11123328.

A. SUPPLEMENTARY INFORMATION FOR THE SUBMITTED PAPER IN CHAPTER 2

A.1 SUPPLEMENTARY FIGURES

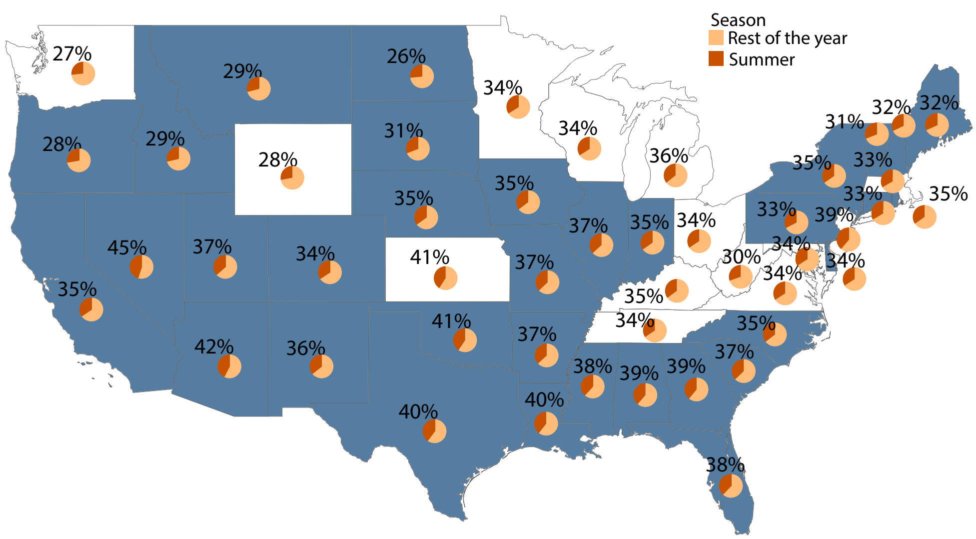
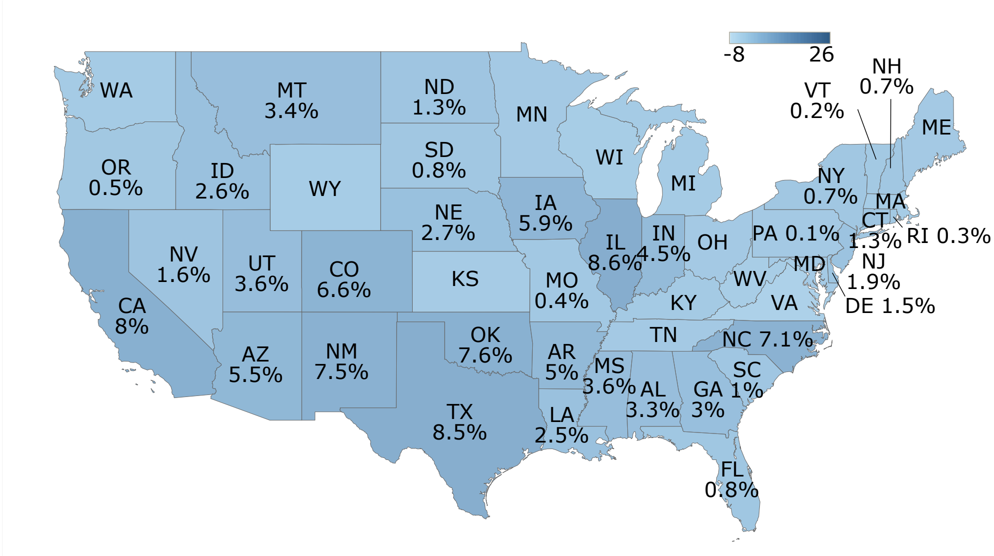
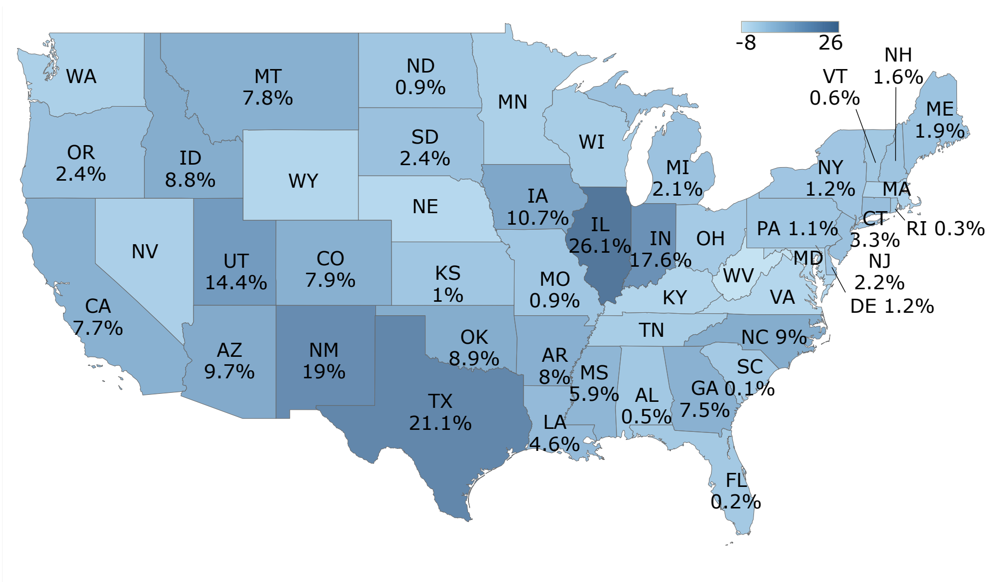


Figure A.0. SUPPLEMENTARY FIGURE 1: Highlighted states represent areas where the selected-features model outperform the air-temperature-only models in terms of predictive accuracy.



(a)



(b)

Figure A.1. SUPPLEMENTARY FIGURE 2 AND 3: Scale bars go from -8 to 26%. The relative improvement of the selected-features models over air-temperature-only models in terms of predictive accuracy—based on out-of-sample RMSE values—for the 50th quantile a, and 90th quantile predictions b. The RMSE values have not been included in areas where the state-level selected-features models did not out-perform the air-temperature-only models.

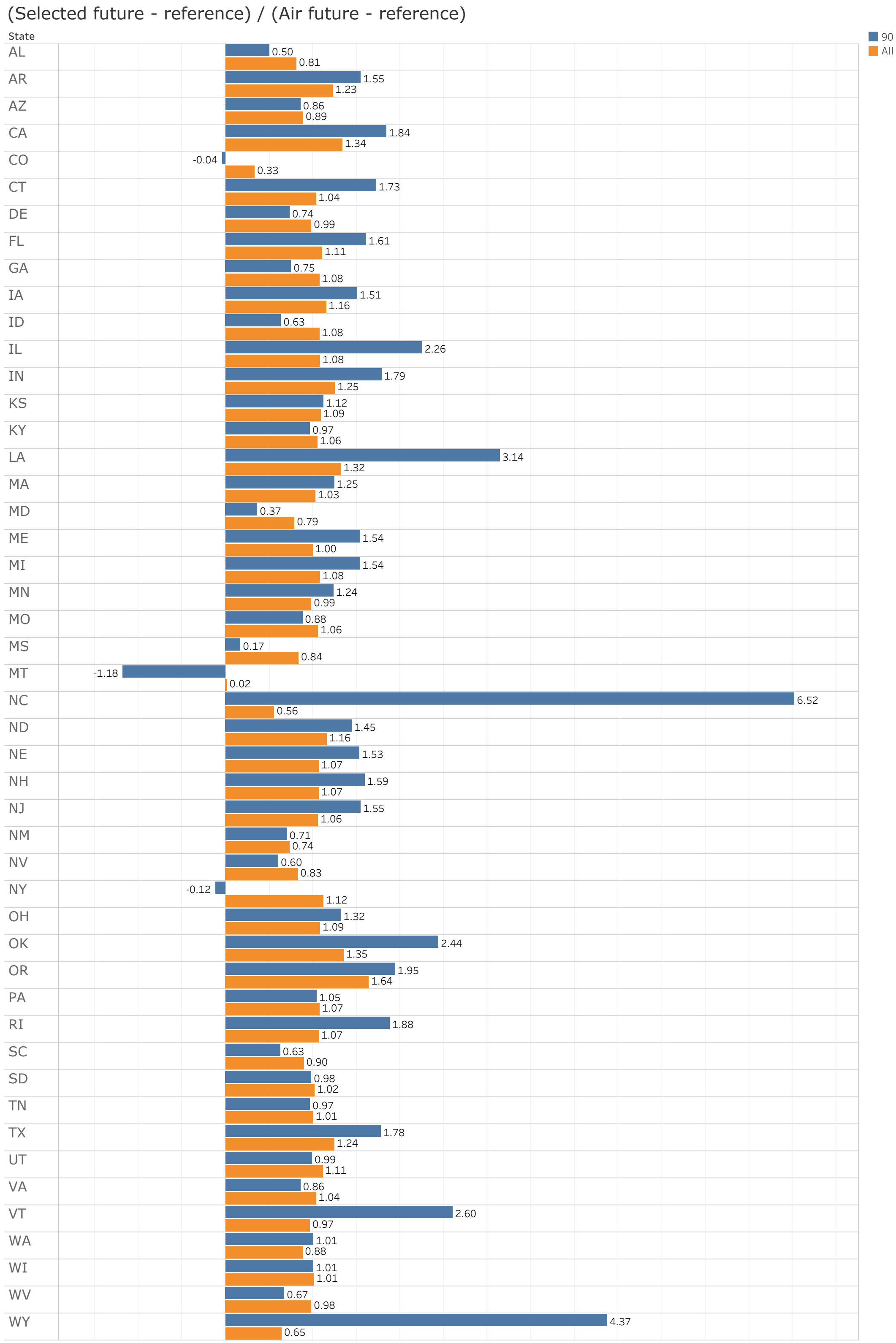


Figure A.1. SUPPLEMENTARY FIGURE 4 The relative increase of projected demand by the selected-features model over the air-temperature-only model for average consumption levels and 90th quantile predictions for all states.

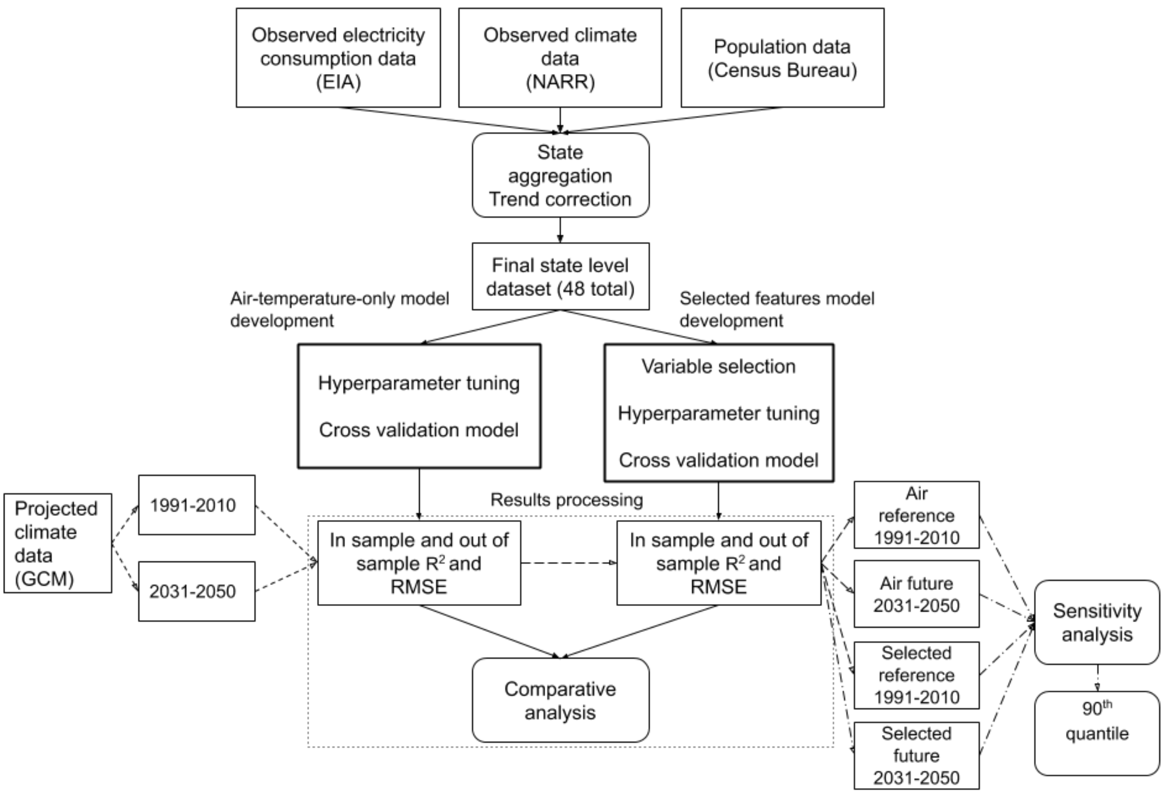
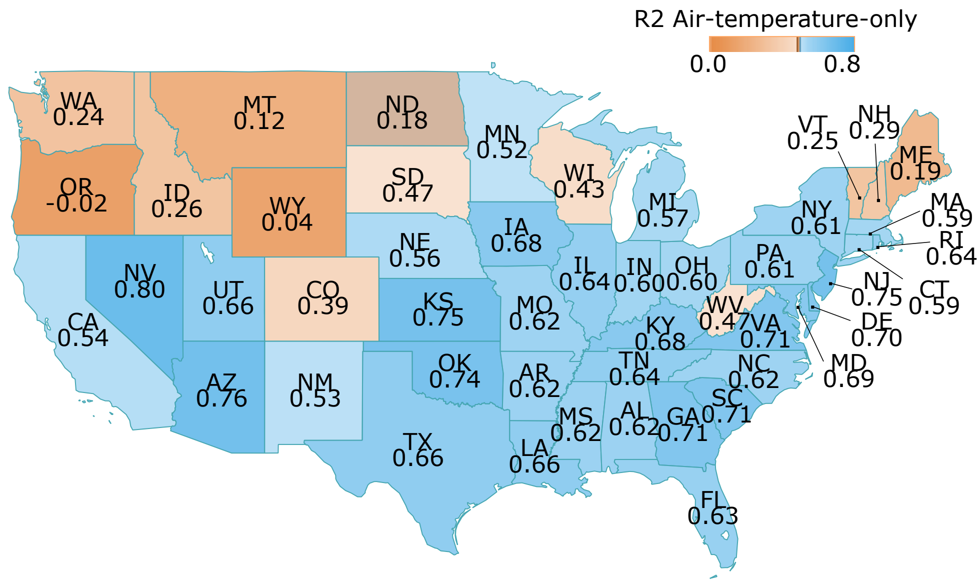
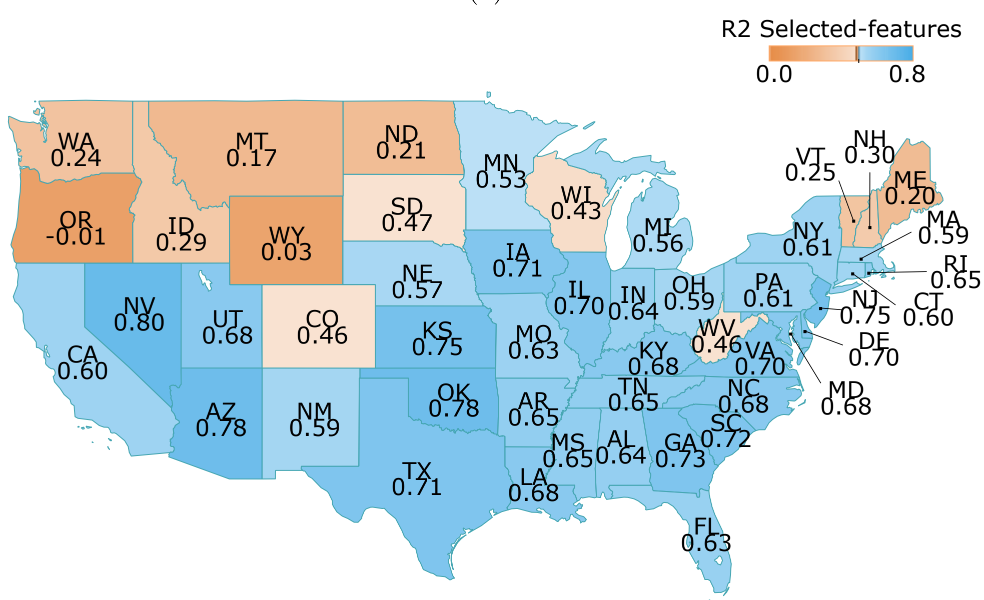


Figure A.1. SUPPLEMENTARY FIGURE 5: Flow chart with the implemented methodology, complementing Methods section Projected climate data. The two stages in model development (modeling with observed data and sensitivity analysis with projected data) are represented by the vertical direction of the flow chart (from the observed data collection until results processing, continuous arrows) and the horizontal direction of the flow chart (from projected climate data until the 90th quantile sensitivity analysis, dashed arrows).



(a)



(b)

Figure A.2. SUPPLEMENTARY FIGURE 6 and 7: (a) Out of sample R^2 air-temperature-only models. Red tones represent values negative or close to zero. Values in green are greater than 0.5; same color scheme for both R^2 plots.(b) Out of sample R^2 Selected features. As in the air models, the northwest region did not present a good fit and the southern region presents good fit with clear improvements, like in TX, from 0.66 to 0.71.

A.1.1 SUPPLEMENTARY NOTE 1

Energy consumption during summer represents a high percentage of total consumption in the U.S. The observed data from 1990 to 2016 reveal that the summer months (i.e., May, June, July and August) account for an average of 35% of total energy consumption, and even exceeding 40% in some states (Supplementary Figure 1). States which consumption is equal greater than 40% represent a highly meaningful peak.

The highlighted states in Supplementary Figure 1 signify areas where accounting for humidity in addition to surface temperature leads to an improved accuracy in characterizing the climate sensitivity of electricity demand. In other words, the selected-features models outperform the air-temperature-only model, in terms of predictive accuracy, in the highlighted states.

B. SUPPLEMENTARY INFORMATION FOR THE SUBMITTED PAPER IN CHAPTER 3

B.0.1 SUPPLEMENTARY TABLE 1

Table B.1. Calculated CDD values one-sample tests with mean = 18.3, and sample mean of states where 18.3 was within the 95% CI.

State	Test	Test p-value	sample mean within CI	CDD
AL	t-test	p-value $<$ 2.2e-16	-	21.1
AR	t-test	p-value = 1.629e-09	-	17.1
AZ	Wilcoxon signed rank test	p-value = 7.79e-10	-	20.2
CA	t-test	p-value = 0.9209	sample estimates: mean of x 18.30825	18.2
CO	t-test	p-value $<$ 2.2e-16	-	15.0
CT	t-test	p-value = 1.697e-15	-	17.6
DE	Wilcoxon signed rank test	p-value = 7.495e-06	-	17.8
FL	t-test	p-value $<$ 2.2e-16	-	23.3
GA	t-test	p-value $<$ 2.2e-16	-	20.9
IA	t-test	p-value = 0.005519	-	18.5
ID	Wilcoxon signed rank test	p-value = 0.001762	-	17.7
IL	t-test	p-value $<$ 2.2e-16	-	16.5
IN	t-test	p-value = 0.3353	sample estimates: mean of x 18.25233	18.2

continued on next page

Table B.1. *continued*

State	Test performed	Test p-value	sample mean within CI	CDD
KS	t-test	p-value = 0.006166	sample estimates: mean of x 18.36699	18.4
KY	t-test	p-value < 2.2e-16	-	19.5
LA	t-test	p-value < 2.2e-16	-	21.9
MA	t-test	p-value < 2.2e-16	-	16.6
MD	t-testt	p-value = 0.3415	-	18.4
ME	t-test	p-value = 1.847e-14	-	15.9
MI	t-test	p-value < 2.2e-16	-	17.1
MN	t-test	p-value = 4.066e-14	-	17.4
MO	t-test	p-value = 1.563e-11	-	19.0
MS	t-test	p-value < 2.2e-16	-	20.3
MT	Wilcoxon signed rank test	p-value = 7.79e-10	-	15.4
NC	t-test	p-value = 5.625e-06	-	18.7
ND	Wilcoxon signed rank test	p-value = 0.01311	-	19.0
NE	t-test	p-value = 4.04e-09	-	17.5
NH	Wilcoxon signed rank test	p-value = 1.2e-06	-	14.7
NJ	t-test	p-value < 2.2e-16	-	17.7
NM	Wilcoxon signed rank test	p-value = 0.0002946	-	15.8
NV	Wilcoxon signed rank test	p-value = 8.279e-10	-	16.0

continued on next page

Table B.1. *continued*

State	Test performed	Test p-value	sample mean within CI	CDD
NY	t-test	p-value < 2.2e-16	-	16.6
OH	t-test	p-value = 0.09624	sample estimates: mean of x 18.21356	18.3
OK	t-test	p-value = 0.9754	sample estimates: mean of x 18.29542	18.0
OR	Wilcoxon signed rank test	p-value = 2.467e-06	-	16.6
PA	t-test	p-value = 0.8601	sample estimates: mean of x 18.28939	18.2
RI	t-test	p-value = 3.665e-14	-	17.7
SC	Wilcoxon signed rank test	p-value = 0.0001871	-	18.6
SD	Wilcoxon signed rank test	p-value = 0.0001871	-	16.7
TN	Wilcoxon signed rank test	p-value = 0.3154	-	18.5
TX	t-test	p-value < 2.2e-16	-	19.8
UT	t-test	p-value < 2.2e-16	-	16.7
VA	t-test	p-value = 1.385e-05	-	18.7
VT	t-test	p-value < 2.2e-16	-	16.2
WI	Wilcoxon signed rank test	p-value = 0.0001801	-	17.5
WV	t-test	p-value = 1.163e-09	-	18.0

continued on next page

Table B.1. *continued*

State	Test performed	Test p-value	sample mean within CI	CDD
WY	Wilcoxon signed rank test	p-value = 0.03706	-	18.6

B.0.2 SUPPLEMENTARY NOTE 1

Considering the uncertainty of the Energy Signature method, in Table B.1 we see significance tests' p-values for each for the updated CDD set-points. Highlighted are states where 18.3 is within the 95% confidence interval of the 50 bootstrap sample.

B.0.3 SUPPLEMENTARY FIGURE 1

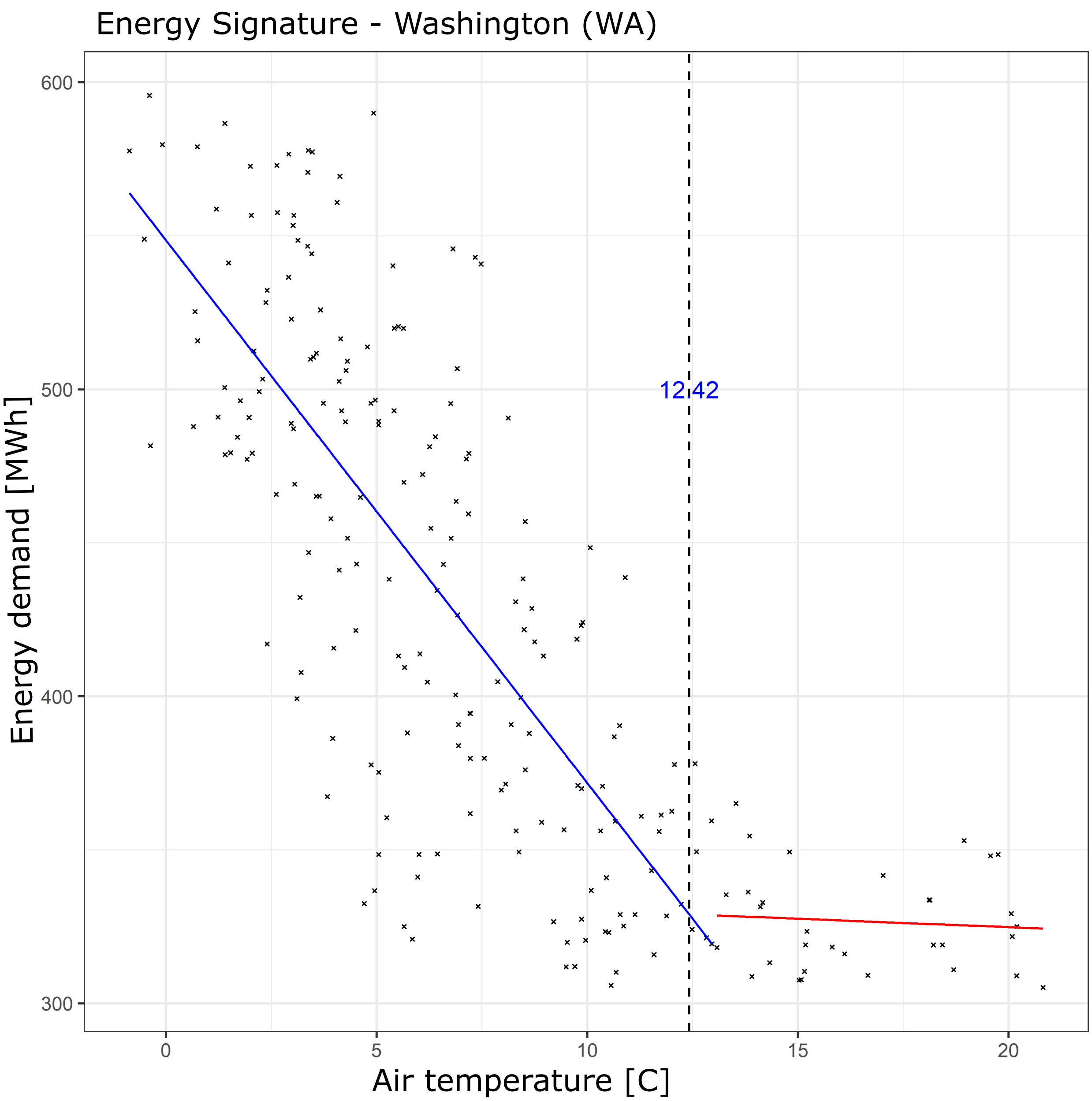


Figure B.1. The Energy Signature method for the state of Washington (WA). Even though there is a energy response for the heating demand, there is no visible response for the cooling demand. Hence, we did not add WA results for the derived air temperature results depicted in Fig. 1.

B.0.4 SUPPLEMENTARY FIGURE 2

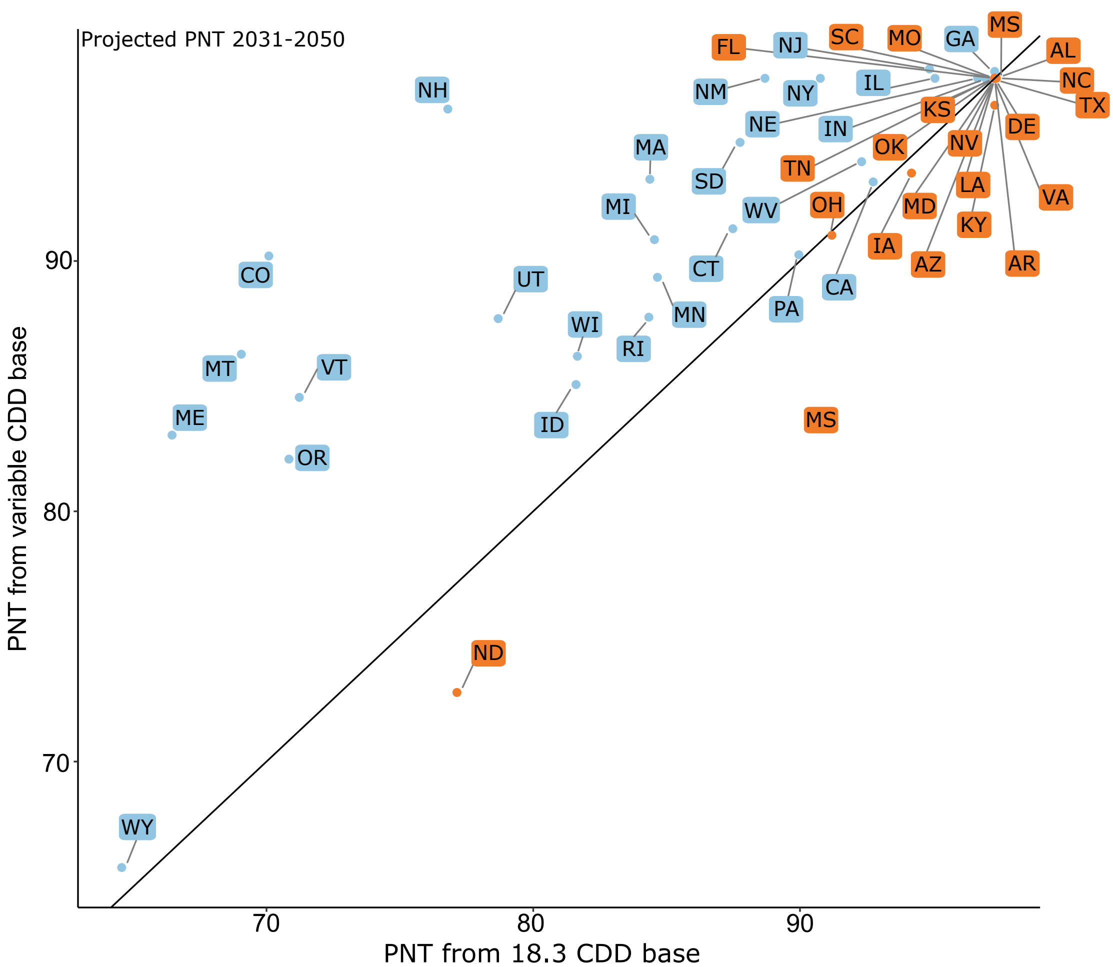


Figure B.2. Projected values (2031-2050) PNT from variable CDD versus PNT from the 65°F (18.3°C) base value.

B.0.5 SUPPLEMENTARY FIGURE 3

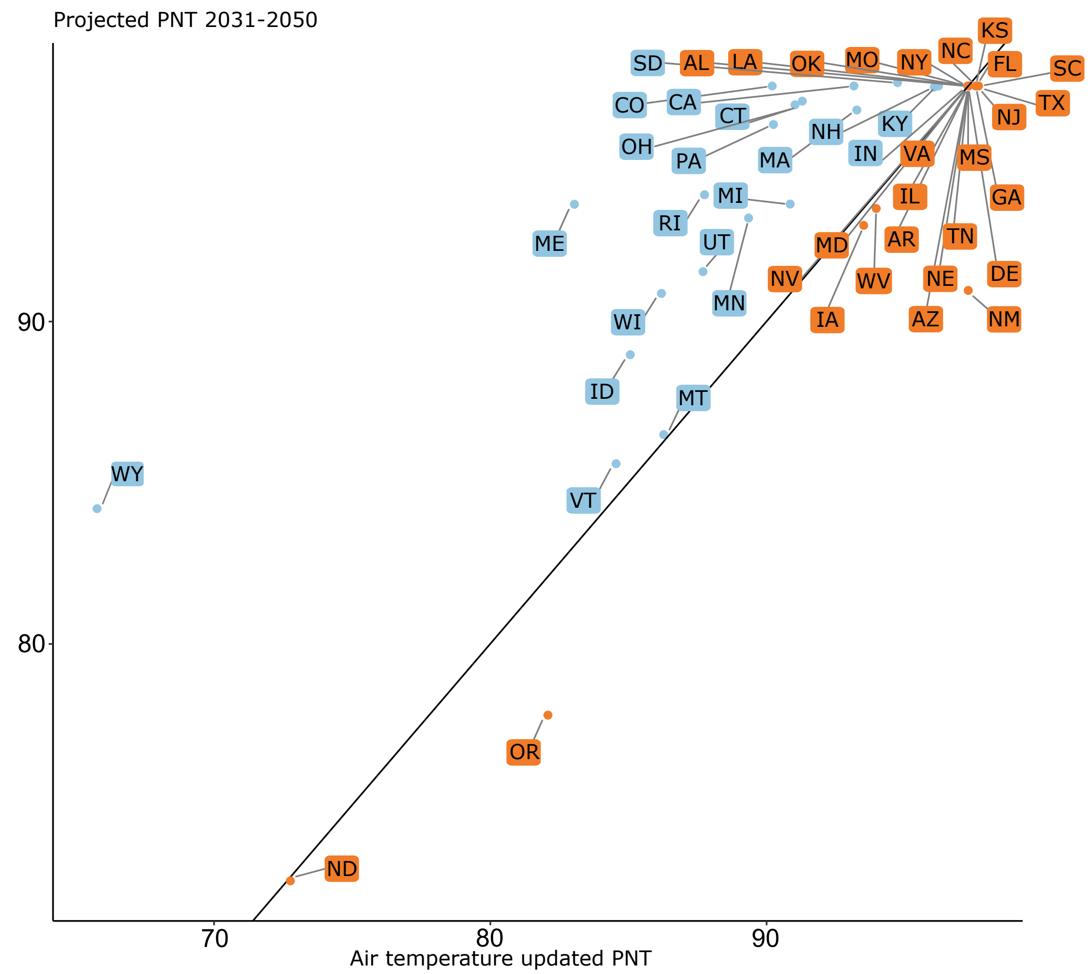


Figure B.3. Projected values (2031-2050) PNT from heat index CDD versus PNT from variable CDD.

B.0.6 SUPPLEMENTARY FIGURE 4

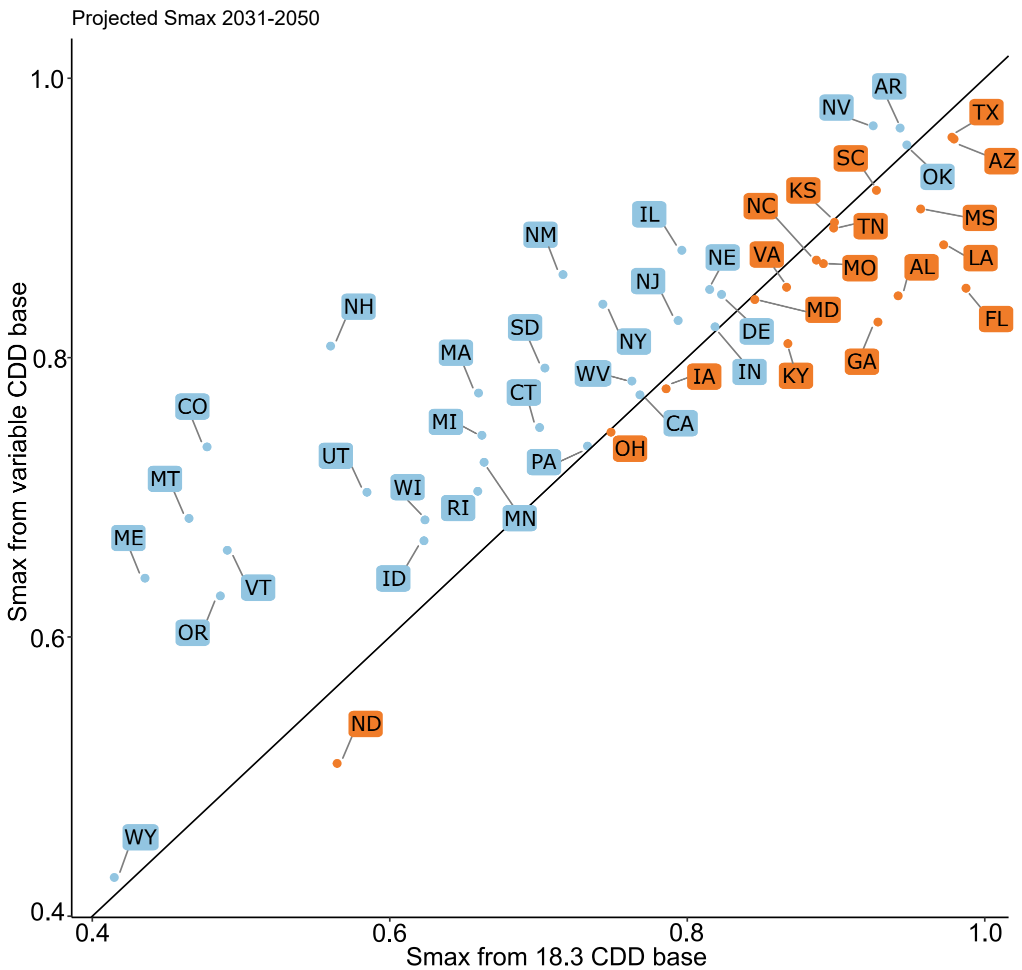


Figure B.4. Projected values (2031-2050) Smax from variable CDD versus Smax from the 65°F (18.3°C) base value.

B.0.7 SUPPLEMENTARY FIGURE 5

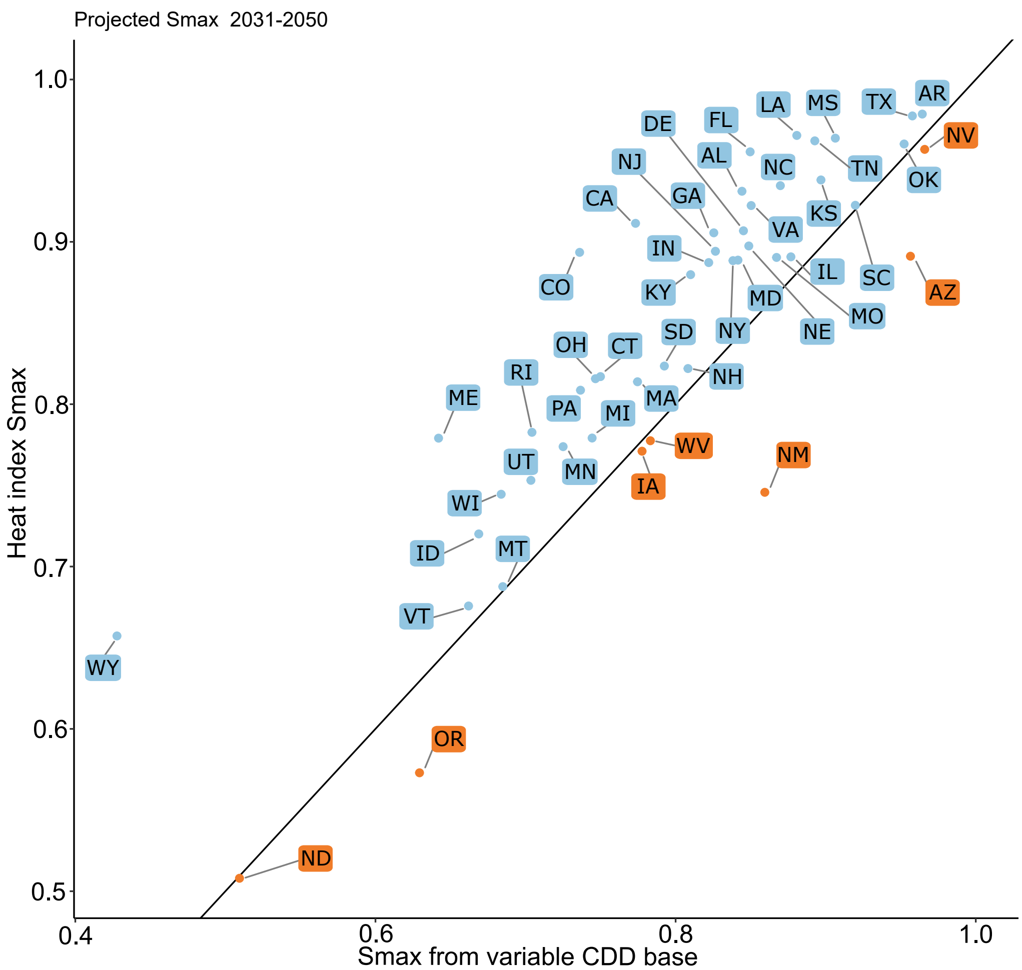


Figure B.5. Projected values (2031-2050) Smax from heat index CDD versus Smax from variable CDD.

B.0.8 SUPPLEMENTARY FIGURE 6

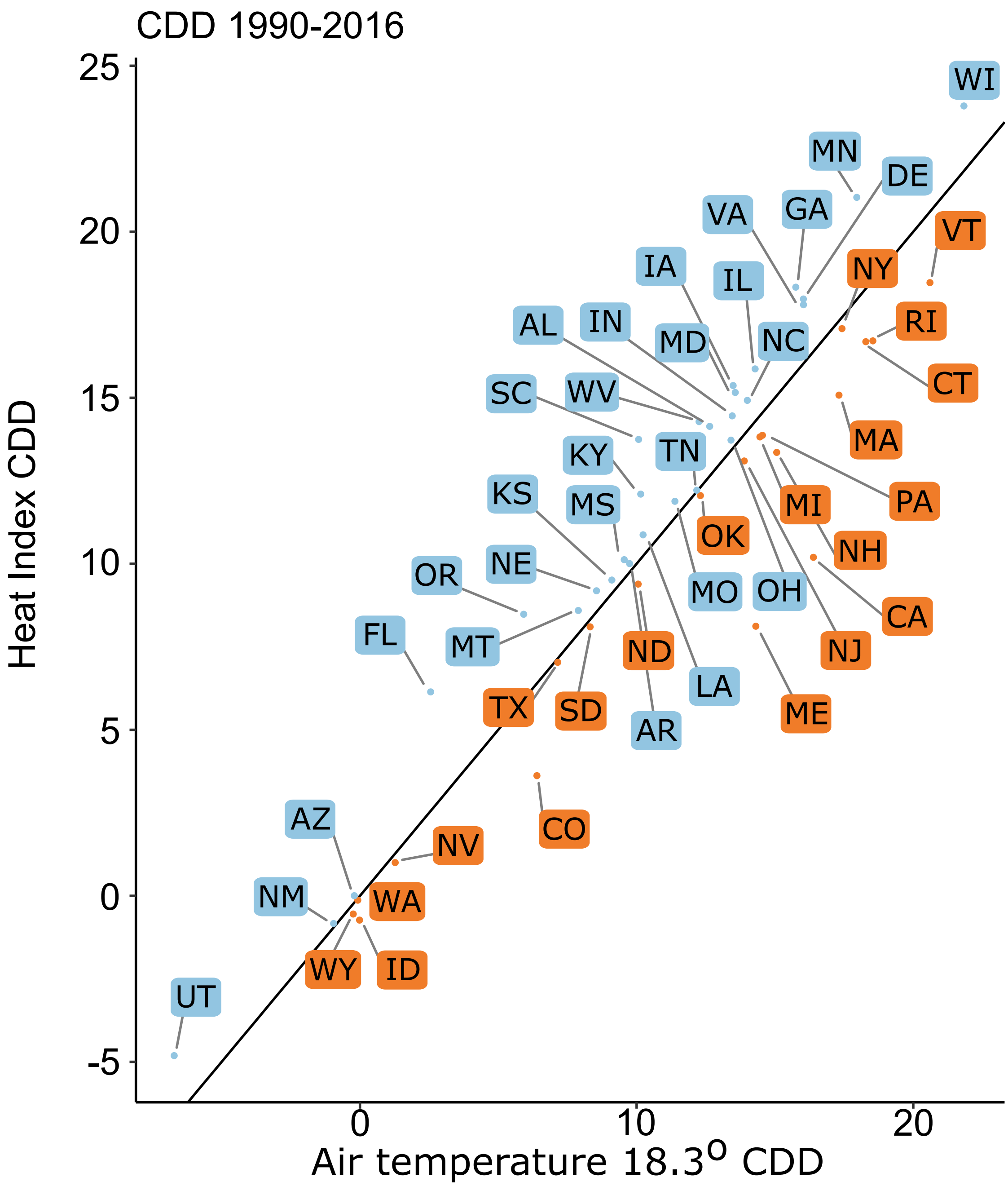


Figure B.6. Heat index CDD versus air temperature CDD for the 65°F (18.3°C) base value.

VITA

Dr. Debora Maia-Silva

B.1 Education

- Purdue University, West Lafayette, IN, USA Direct Ph.D in Environmental and Ecological Engineering, **Ross Fellow**, Aug. 2017 – May 2021
- Campinas University (UNICAMP), Campinas, Brazil Bachelor of Civil Eng., Specialization in Energy Resources; Minor in Computer Eng., 2010 – 2016

B.2 Work Experience

- Capital One , McLean, VA, USA Ph.D Data Science Intern Program, June 2020 - August 2020
- Reserach Assistant , West Lafayette, IN, USA Ph.D Reserach Assistant to Prof. Nateghi, August 2017 - May 2021
- Metodo Potencial, Sao Paulo, Brazil Engineering Intern, Feb. 2016 - May 2016
- Cortes Consulting, Campinas, Brazil Engineering Intern, 2014 - 2015

B.3 Publications

- Obringer et al., ”Anthropogenic warming intensifies household air conditioning demand across the United States”, *under Review*
- Renee Obringer, Benjamin Rachunok, **D. Maia-Silva**, Maryam Arbabzadehd, Roshanak Nateghic, Kaveh Madanief. ”The overlooked environmental footprint of increasing Internet use.” *Resources, Conservation and Recycling*, 2021, https://doi.org/10.1016/j.resconrec.2020.105389

- **D. Maia-Silva**, R. Kumar, R. Nateghi. The critical role of humidity in modeling summer electricity demand across the United States. *Nature Communications*, 2020, https://doi.org/10.1038/s41467-020-15393-8

- R. Kumar, B. Rachunok, **D. Maia-Silva**, R. Nateghi. Asymmetrical response of California electricity demand to summer-time temperature variation. *Scientific Reports*, 2020 https://doi.org/10.1038/s41598-020-67695-y

- S. Andrade, A. Pires, T. Santos, A. Beck, D. Matias, **D. Maia-Silva**, L.C.M. Vieira Junior. Impact of changing wind speeds on the reliability of steel frames: a Brazilian case study. (Under review,*ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems*, 2020)