3-2021

Climate-informed Prediction and Forecast Modeling of Installation Total Energy Consumption

Scott C. Weiss

CLIMATE-INFORMED PREDICTION AND FORECAST MODELING OF INSTALLATION TOTAL ENERGY CONSUMPTION

THESIS

Scott C. Weiss, First Lieutenant, USAF

AFIT-ENV-MS-21-281

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-21-281

CLIMATE-INFORMED PREDICTION AND FORECAST MODELING OF INSTALLATION TOTAL ENERGY CONSUMPTION

THESIS

Presented to the Faculty

Department of Graduate Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Engineering Management

Scott C. Weiss, BS

First Lieutenant, USAF

March 2021

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-21-281

UNBIASED, CLIMATE-INFORMED MODELING OF INSTALLATION TOTAL ENERGY CONSUMPTION

Scott C. Weiss, BS

First Lieutenant, USAF

Committee Membership:

Maj Justin D. Delorit, PhD, P.E.
Chair

Lt Col Steven J. Schuldt, PhD, P.E.
Member

Dr. Christopher Chini, PhD
Member

AFIT-ENV-MS-21-281

Abstract

Climate variability is an external and stochastic factor that causes energy demand uncertainty. Energy managers can use climate-based models to understand future trends of energy demand and to adjust operations, policy, and budgets accordingly. This research focuses on 1) identifying how climate attributes impact energy use, 2) creating a historically informed statistical modeling framework to skillfully predict energy use, and 3) forecasting future changes to energy use and costs, using CMIP5 temperature projections, at the campus level. After synthesizing the existing breadth of research on climate-informed energy modeling, a skillful, unbiased, climate-informed total energy consumption prediction model is developed for Wright-Patterson AFB (WPAFB) that is particularly skillful at predicting energy use during high and low use periods: the periods where impactful energy policy decisions are made (r^2 = 73%, MAPE = 6.15, RPSS = 0.59). CMIP5 projections of temperature inform the model to generate energy use forecasts, which reveal significant changes to energy use within the next decade and increases in annual energy use costs by $7.3-7.9M by the end of the century. Overall, energy use predictions and forecasts can pinpoint the impact of climate factors, inform how and when to mitigate changes, and justify intervention timing and financial decisions.

Acknowledgments

First and foremost, I thank God for the gifts, abilities, and opportunities He has given me throughout my life. I am specifically appreciative of my wife, whose dedication to her profession and unwavering support is a constant source of inspiration. I dedicate this work to my daughter who is the joy of my life and who drives me to better myself every day. I am also appreciative of my advisor, Major Justin Delorit; thank you for choosing to take me on as an advisee. I have gained more knowledge through the thesis process and your courses over the past eighteen months than I thought I could achieve thanks to your commitment and skill at teaching your expertise. Finally, I want to acknowledge the GEM Class of 21M, whose fellowship and dedication to excellence through a national pandemic is truly inspirational. I am grateful to have met each one of you and hope that we cross paths again in the future.

Scott C. Weiss

Table of Contents

List of Figures

Page

List of Tables

UNBIASED, CLIMATE-INFORMED MODELING OF INSTALLATION TOTAL ENERGY CONSUMPTION

I. Introduction

1.1 Background

The impacts of climate change are a growing concern for facility and infrastructure asset managers across the private and public sectors. The performance and lifecycle of infrastructure systems, whose design and operations are influenced by climate trends and factors, could become less predictable in the future. Asset managers use data to drive operational decisions and decide where and when policy and structural interventions may be necessary to stabilize systems. The management of public infrastructure is a particularly important focus because impacts to these systems are a matter of public safety and national security, and this infrastructure is publicly funded. However, stability can be promoted by understanding how climate attributes impact facility and infrastructural systems, and by being able to predict how performance and use will change with changing climate.

The U.S. Congress tasked the Department of Defense (DoD) to provide a report of potential climate change impacts to military installations to understand how defense operations may be impacted by climate change. The resulting DoD report identified climate change associated weather events (e.g., drought, wild-fires, tornados, hurricanes, flooding), and identified installations that are susceptible those events. In response to the DoD's findings, the Government Accountability Office (GAO) published a report highlighting the DoD's installation robustness and resilience shortcomings, as related to extreme events intensification driven by climate change [1]. While, the GAO 2019 report recommends the incorporation of climate projections and extreme event trends for installation planning, applications in trend-based analyses were largely overlooked [1].

Energy is a pivotal, no-fail, infrastructure system category at any organizational scale. Research exists examining climate-informed modeling of energy use from facility to multinational scales. However, due to data accessibility issues, campus-scale analyses are not common. Military installations provide a unique opportunity to conduct analyses at this scale. DoD installations are the U.S. government's largest energy user [2], [3]. Therefore, energy reliability and accurate budgeting of this recurring cost should be a priority, to ensure installation operations are uninterrupted and taxpayer dollars are used responsibly. However, the amount of energy consumed can vary based on operational use patterns, temporality, and climate conditions. Through statistical modeling, it is possible to use climate factors to predict and forecast energy consumption accurately, so energy managers can proactively plan policy and budgets for short- and long-term energy consumption.

This thesis investigates the development of statistical, climate-informed, energy consumption prediction models to understand the future impacts of climate change on installation energy consumption. This exploration requires an understanding of the impact of climate on energy consumption, modeling types, and whether skillful models can be created using climate data. With a skillful model, not only can the most influential inputs be identified, but the model can be applied in a forecast mode using climate projections to understand long term trends in energy consumption.

1.2 Problem Statement

Within the Air Force, installations are required to submit annual energy budgets to the Air Force Installation Management and Support Center (AFIMSC) which, in turn, combines and submits an enterprise estimate as part of its comprehensive operational budget to Congress for approval. Installations typically use the prior-year energy use as the estimate for the next year's energy consumption, adjusting only for utility rate inflation and mission changes, such as new mission beddown, if necessary. If actual consumption outpaces predictions at the enterprise-scale, the deficit is taken from the Air Force's Facility Sustainment, Restoration, and Modernization (FSRM) budget. Due to chronic overestimation, AFIMSC is considering holding installations accountable for energy use overestimates (Personal communications, AFIMSC/RMAO). Accountability could take the form of direct deductions from installation-level FSRM funds if a positive difference between consumption and predictions exceeds a pre-determined error threshold. Many installations are ill-prepared to produce skillful predictions and forecasts of energy use, as energy managers typically lack the ability to account for stochastic and pseudo-stochastic variability in their estimates [1].

1.3 Research Objectives

Given the intent of this thesis, which is to provide an understanding of the impact of future climate change to military installation energy consumption, the following research objectives are as follows:

1. Review the existing literature to identify the components and methods of developing climate-informed energy consumption prediction models.

2. Determine the influence of climate inputs on the military installation energy consumption prediction modeling and whether skillful models are achievable using climate inputs.
3. Determine forecasted trends for total energy consumption, and its implications, for WPAFB using climate projections generated through the 21st century.

1.4 Thesis Organization

This thesis follows a scholarly format in which chapters 3 and 4 have been prepared as stand-alone journal publications. In Chapter 2, “A Literature Review of Approaches for Facility to Enterprise-Level Energy Consumption Prediction and Forecast Modeling Under Non-Stationary Climate and at Multiple Temporal Scales,” an overview of the existing literature on energy consumption modeling with climate factors is conducted. This chapter highlights the variation of approaches and inputs in the modeling of energy consumption along with gaps in the existing body of knowledge.

Chapter 3, “An Unbiased Climate-Informed Tool for Campus Energy Policy Development and Budget Predictions,” fills several gaps in the existing body of knowledge by developing an energy prediction model for a campus-sized community (WPAFB). Additionally, bias correction and multicollinearity reduction techniques are incorporated in the model development framework alongside an extensive list of open-source climate data. This study aims to determine if skillful energy prediction models can be created solely using open-source climate data and uncovers how climate variables impact the skill of the energy prediction models. The paper then discusses the model’s significance to energy managers and introduces the tradeoff of model skill and complexity. At the time this thesis is submitted, this manuscript is under review with *Applied Energy*.

Chapter 4, "Climate-Informed Energy Consumption Projections for Campus Policy Development and Budget Decisions," applies the findings of Chapter 3 and climate projections to forecast energy consumption for WPAFB through the year 2100. The associated changes in cost due to changing climate are determined and discussed, along with how these forecasts can be used to inform management decisions. At the time this thesis is submitted, this manuscript is a working paper.

Finally, Chapter 5 outlines conclusions and suggests follow-on research to expand upon the research.

II. A Literature Review of Approaches for Facility to Enterprise-Level Energy Consumption Prediction and Forecast Modeling Under Non-Stationary Climate and at Multiple Temporal Scales

2.1 Introduction

Due to the impact of greenhouse gas emissions on climate, the international academic community has allocated time and resources to the identification and quantification of potential risks to businesses, governments, and communities, that stem from extreme events or gradual trends in climate. One key area of research involves gradual impacts on energy. Energy consumption is one of the highest, "must pay" costs at any organizational size. In 2017, the United States of America (U.S.) used 97.6 trillion BTUs of energy totaling $1.1T, or 5.8% of Gross Domestic Product. More specifically, residential, commercial, and industrial sectors spent $224B, $201B, and $359B in energy costs in 2017, respectively [4]. As such, accurate predictions of future energy costs are important in informing organizational and facility operating budgets, as are policies and technologies to mitigate potential energy cost variability. Overestimating energy demands wastes resources and may lead to unnecessary energy infrastructure expansion. Also, underestimation can result in failures and shortages [5]. Additionally, energy prediction accuracy may provide facility managers confidence when planning other budget items.

Researchers have developed climate-driven prediction models, of varying skill, to determine future energy costs [5]–[7]. The underlying connection between facility energy and climate factors stems from electrical energy used for heating and cooling: since heating and cooling systems adapt facility conditions from the outdoor environment. The most recent Energy Consumption Survey released by the U.S. Energy Information Association (EIA) identifies that 61% and 47% of energy consumption for commercial and residential facilities, respectively, is electrical energy. Additionally, heating and cooling costs account for 32% and 51% of energy consumption for commercial and residential sector facilities, respectively [8].

Empirical models are widely used to predict energy use at multiple spatiotemporal scales [5], [7], [9]–[14] and are generally informed with the following aspects: forecasted climate data, historical energy data, which are climate models conditioned on likely greenhouse gas emissions scenarios enable forecasting of natural and human systems [15]. A variety of climate projection models exist, including the Global Change Assessment Model, which is developed by the Joint Global Change Research Institute and recognized by the Intergovernmental Panel on Climate Change (IPCC) [16]. Once statistically significant regressive relationship between historical climate variables and energy consumption has been established in a hindcast mode, the projected climate data is inserted into this model to determine likely future energy use. In the case of RCP, or other ensemble predictions, multiple forecasted energy use scenarios are produced. Studies using forecasts of future climate to predict end-of-century energy use have found similar trends. Predictions spanning the 21st century, show that peaks in demand will grow in winter and summer seasons. Also, there will be a period of reduced energy consumption approaching the middle of the century as energy consumption decreases in the winter season, faster than energy usage increases in the summer. However, by the end of the century energy consumption levels will have increased past early 21st century levels [13], [14].

The bulk of the effort required to predict energy use with climate variables is devoted to (1) determining whether supply and/or demand will be analyzed, (2) selecting the appropriate spatial and temporal level of analysis, (3) collecting data, and (4) selecting a regression technique. These aspects are what create a wide range of research in this field [10], [14], [17]–[20]. As such, this paper focuses on a contemporary review of these aspects, and specifically as they relate to the relationship between climate and energy consumption.

2.2 Discussion

2.2.1 Energy Supply v. Demand

When performing analyses on the effect of climate on energy use, researchers can focus on the demand or supply side, or both. The demand side analyses involve estimating energy consumption and associated costs; whereas, the supply side takes into account production and distribution costs. The bulk of recent research has focused on the demand side, because demand side is impacted more by climate [21] and, intuitively, the primary way energy users control total energy cost is through demand interventions. Also, suppliers tend to adjust their supply based on past demand quantities. With the likely increase in summer peak energy demand, additional infrastructure investments may be necessary to ensure that the energy generation and distribution systems can meet the demand [22].

Peak temperature across the year is expected to increase 6.3-9% by 2050, driving a 10-20% expansion of electrical generation capacity and require billions of dollars in infrastructure investments [23]. These increases can be apportioned to climate's effect on transmission line losses, power plant heating effects, substation heating affects, and energy demand. An example of climate change on power plant heating effects includes fluctuations in water availability and temperature, which could make power generation via fossil fuel or nuclear energy less efficient since water is the primary method to cool generators [22]. These are just a few of the areas where the energy supply impacts of gradual changes to climate have been analyzed. The level of analysis section identifies specific temporal and spatial scales that could be beneficial in determining energy supply and generation predictions. Further explanation of the impacts of changes in climate are discussed.

2.2.2 Level of Analysis

2.2.2.1 Temporal

The level of analysis refers to the temporal and spatial scales or resolutions at which climate-driven energy prediction research is conducted. Temporal analyses are grouped in four categories: very short-term (VST), few minutes ahead to a few hours ahead; short-term (ST), one day to two weeks ahead; medium-term (MT), two weeks to three years ahead; and long-term (LT), three to fifty years ahead [24].

Each temporal level holds value for different aspects of utility services [10]. Improved forecasting enables power system operators to make better informed decisions concerning "supply planning, generation reserves, system security, dispatching scheduling, demand-side management, and financial planning," to name a few [25]. VST forecasts are particularly useful for real-time scheduling of electricity generation, load frequency control, and demand response. These forecasts are crucial to business operations of retailers, power marketers, and trading firms [26]. VST provides energy security to consumers by highlighting sub-hourly periods where supplemental power may be required [27]. Moral-Carcedo and Pérez-García [28] utilized the VST level of analysis to analyze the sensitivity of electricity load to the "rest" and "active" hours of the day [28]. The study finds that at lower temperatures, electricity demand increases significantly more per degree Celsius drop during "rest" periods than "active" periods. This relationship is similar, but less significant, for higher temperatures; however, electricity demand decreases more per degree Celsius drop in “rest” periods than in “active” periods. This could allow power plant managers to more efficiently allocate resources and manpower during rest and active periods of the day based solely on temperature forecasts.

ST forecasts are particularly essential for the operations of the power market as a whole; inaccurate ST forecasts result in large financial losses throughout the market [25]. Xie and Hong [29] utilized week-ahead projections of energy consumption in the New England region of the U.S. to determine the impact of wind speed meteorological variables on a model's predictive ability. It was found that wind speed, when coupled with temperature-only models, adds skill in a hindcast mode. The New England energy market may be able to increase the accuracy of energy consumption expectation by using models that include wind speed. Higher accuracy could translate into less wasted energy production and fewer monetary losses.

MT forecasts primarily aid fuel supply scheduling, utility maintenance operations, and contract negotiations with users, as this level of analysis encompasses month, season, and year(s)-ahead predictions [10]. For instance, De Felice et al. [11] reported that data obtained before and during May for Central and Southern Italy produces significant and skillful predictions for that year's summer electrical energy consumption. This method of prediction could provide more surety to the Central and Southern Italian energy supply utilities in supply planning, maintenance, and contract negotiations.

Finally, LT forecasts are particularly beneficial for informing capacity expansion, capital investment, revenue analysis, and budgeting [30]. Wenz et al. [31] used climate models adapted from the RCPs to conduct a LT analysis of the expected change in daily peak energy load across Europe for the remainder of the 21st century. The study found that a polarization effect occurs where northern countries are likely to experience a decrease in daily peak load, while southern countries may experience an increase by the end of the century. Research using LT forecasts may allow countries on both sides of the polarization, and their respective power generation utilities, to begin planning for potential demand changes.

2.2.2.2 Spatial

Existing studies evaluate model skill at various spatial scales: building, state/regional, and national/multinational. The determination of spatial scale is primarily driven by data accessibility or the specific focus of researchers, for example, impact of climate variables, new method development, the sensitivity of models to location, etc. It is well documented that the effects of climate change vary spatially [6] and therefore, models are calibrated to a specific climate zone, and then tested across climate zones to determine model exportability.

One approach used at the building-level involves the sensitivity of energy to changes in location. De Rosa et al. [18] modeled a standard residential building in Rome and Milan Italy and exported the model to multiple climate regions throughout Europe. Wang et al. [32] performed a similar analysis for a medium-sized commercial office building within different climate zones in the U.S. These studies utilize building-level analysis to more simply test the skill of their models for a single facility type, knowing that the effects of changing climate will differ across each climate region. Another approach is to obtain information about multiple building types. Dirks et al. [12] utilized data from the U.S. Eastern Interconnection (EIC) power grid, along with building energy modeling software containing approximately 26,000 facility types, to model the entire EIC region: spanning multiple climate regions within the U.S. This data was then utilized to inform facility-level energy mitigation and savings techniques.

The next spatial scale is the regional or state level. It encompasses multiple towns or cities potentially across multiple climate zones. Analysis at this level is common since data is readily available from energy providers, as they operate regionally and collect extensive statistics. Cities and states use these statistics to drive policy decisions. Zhou et al. [14] utilizes disaggregated state-level energy data, disaggregated to business sectors, to better capture the spatial heterogeneity of building energy use within each state. Mukherjee and Nateghi [7] focused on energy consumption within Florida. Later, in Mukherjee et al. [19], models developed by Mukherjee and Nateghi [7] were utilized across multiple states to test the model under a diverse array of climate conditions.

The final level of analysis is studies performed at the national or multinational scale. One of the main areas of value for studies at this scale is the data collection. In many cases, a nation will have multiple power grids and generation facilities: both privately and publicly operated. Municipalities may receive energy from multiple providers which may make data collection more [33]. Collecting energy data across multiple countries is even more difficult. It is also possible that multiple climate zones exist within a single nation, like the United States, or multiple districts of a country, which must be accounted for. However, on a multinational scale, geographical location can be accounted for as long as countries observed are either small enough to encompass one or few climate regions, or the range of countries encompasses a large enough area across the globe. Wenz et al. [31] gathered electrical energy consumption data from across Europe to develop their wide-reaching study. It was effective in accounting for different climate regions because the study includes countries close to and farther from the equator.

2.2.3 Data Collection

2.2.3.1 Energy Data

As mentioned in the previous section, accessibility to data is important and, at times, drives decisions within studies, for example, the temporal or spatial scale that a study can be performed. In some cases, energy data is readily accessible, such as first world countries, where asset management and "big data" collection are widely utilized. Chandromawli and Felder [6] identified that there is a lack of research in energy consumption predictions for developing countries [6]. The probable cause is a lack in the amount and quality of historical data.

2.2.3.2 Climate Data

There are a variety of climate variables to consider for energy consumption prediction models. Most modelers seek to balance forecast interpretability and skill, though they initially take an exhaustive approach to identifying potential climate drivers, particularly for statistical models. As with energy data, past analyses have been restricted based on accessibility. In cases of limited access to data, variables are selected based on intuition or expertise in a specific area [5]. In cases of limited computational capacity, the ability to perform exhaustive analyses may be limited. However, existing literature has highlighted specific key climate variables, which may help researchers limit their search space. The following section provides additional detail on past research that has identified notable climate drivers including: temperature, relative humidity, cloud cover, precipitation, wind speed, and irradiation.

Temperature is the most intuitive and widely utilized climate variable for predicting energy consumption [11], [29]. One of the most important aspects of ensuring comfort in a facility is the control of temperature. This climate variable has been utilized in a variety of ways throughout energy prediction literature. This is because temperature is measured in a variety of ways: including, dry bulb temperature (most common), wet bulb temperature, and dew point temperature. Dry bulb temperature is the ambient air temperature when not subject to air moisture. Wet bulb temperature is the ambient air temperature that takes into account the cooling effect of moisture evaporation. Dew point temperature is the temperature where air vapor begins to condense out of the air (a good indicator of relative humidity). With the implementation of more advanced regression techniques, studies have recently recorded accurate predictions that incorporate stand-alone temperature variables [7], [19], [34].

However, temperature has also been modified using degree day methodology (DDM) to also be incorporated into energy predictions. The DDM finds the difference between the daily temperature and a predetermined facility “bliss” point (65° F is the accepted value) [9]. If the actual air temperature is above the "bliss" point, the difference is the amount of cooling degree days; thus, modeling the need for cooling. If the actual air temperature is below the "bliss" point, the difference is the amount of heating degree days; thus, modeling the need for heating. The primary argument against the DDM is the setting of a "bliss" point. In reality, not every facility has a temperature set at 65° F and the setpoint may change throughout the day, thus adding uncertainty to the model from conception [9]. Additionally, the DDM does not account for the ventilation requirement of a facility, which could account for 2-3% of a facility's total energy use [12].

Temperature has also been used to normalize climate variables. Researchers have used index variables to convert other climate variables to temperature [10], [29]. For example, Apadula et al. [10] accounted for relative humidity and windspeed using heat and wind-chill index variables, respectively.

Relative humidity is the measure (percentage) of water vapor saturation in the air. This variable has appeared in multiple recent studies [10], [12]. The need for incorporating those variables that affect ventilation is partially accounted for using relative humidity, as the human production of water vapor with breath necessitates new conditioned air while expelling moist air from the HVAC system. There is an apparent similarity between relative humidity and the measures of wet bulb and dew point temperature; relative humidity is specifically moisture centric. However, this relationship hints at potential multicollinearity when used together in models; principal component, higher order, or non-parametric analyses would be required to account for this [35].

Cloud cover, though not widely utilized, is another variable used in energy consumption prediction [10]. This variable is usually measured with cloud fractions, where 0/8 is equivalent to clear skies and 8/8 is equivalent to completely overcast skies. Apadula et al. [10] utilized cloud cover as it relates to facility lighting, where, the more cloud cover there is, the more facilities will need to use their lights (interior and exterior). Cloud cover could potentially have an impact on the effect of solar radiation and the heating of a facility's envelope.

Precipitation has shown varying impact across studies. There are a few studies that have found precipitation to be insignificant in energy prediction models [36], [37]. However, more recent studies have identified precipitation as key to the predictive ability of models [7], [19]. As such, the effect of precipitation requires further exploration. Intuitively this climate variable would provide a similar effect to how moisture makes wet bulb temperature read lower than dry bulb temperature. In other words, precipitation provides a cooling effect for facilities.

Wind Speed has been highlighted in a variety of studies as being a significant climate variable in predicting energy consumption. Like precipitation, Mukherjee and Nateghi [7] touted wind speed as being a key predictor of energy consumption. In some cases, wind speed is the second most influential climate variable [19]. Further, Xie and Hong [29] finds that models that incorporate wind speed with temperature outperform other benchmarked models.

Solar irradiance is a measure of the Sun's radiant energy. Irradiance impacts energy usage by striking and ultimately heating a facility's envelope. De Rosa et al. [18] find that irradiance begins to have an impact on energy consumption as the number of cooling degree days decreases. To account for the impact irradiance, [18] adapted DDM calculations. However, only one form of irradiation is utilized by [18]. Al-Bayaty et al. [17] incorporated seven different measures of irradiation into a machine learning model, including global horizontal, direct normal, diffuse, total surface, direct surface, and diffuse surface irradiation. This variety of measures along with the projected increase of irradiant magnitude over time makes irradiation's inclusion in energy prediction calculations important moving forward [38].

2.2.4 Regression Modeling

The next piece needed for generating energy use predictions is regression modeling. Chandramowli and Felder [6] presented a review of energy consumption methods, where four commonly used regression modeling methods were identified: multiple linear regression, bottom-up energy accounting, fuzzy regression, and artificial neural networks. These and three more recently utilized methods are reviewed in this section.

Multiple linear regression is an extension of simple linear regression; however, multiple predictor variables are linearly related to the response variable. In Vu et al. [39], backward elimination regression analysis was used to determine the climate and energy relationship. Backward elimination involves inserting all possible predictor variables into the multiple linear regression model, and systematically eliminating variables that are not significant. A similar approach can be used for higher-order terms if they have not already been incorporated. The unavoidable occurrence of multicollinearity in climate data was accounted for by using variance inflation factor (VIF) and the consequent elimination of redundant variables.

Bottom-up energy accounting collects energy demand data from building equipment and appliances across different types of facilities. This data is then aggregated to the building, and potentially city, level to determine building level climate change impacts. This type of modeling is beneficial when extrapolating across certain building types across climate zones to determine the differing spatial impacts of climate change [40]. Here, the relationship between climate variables and energy consumption is captured in the appliance and equipment energy system readings.

Fuzzy regression analysis is a widely utilized derivative of linear regression that relaxes the assumptions and requirements of linear regression. As aptly stated by [41], "the complexity of real-life problems often makes the underlying (linear) models inadequate, since information is frequently imprecise in many ways." The ambiguity of a complex system is expressed using "fuzzy parameters" or probability distributions, which correspond to the "fuzziness of the system" [42]. Other methods of bypassing the parametric requirements of multiple linear regression have been conceived and are discussed in the remaining models.

The purpose of an artificial neural network (ANN) is to mimic the inner workings of the human brain. The value of ANNs come from their speed and simplicity in extracting non-linear relationships to develop the desired outputs. ANNs can simultaneously test many combinations of a variety of inputs to find these outputs [20].

There are also ways for the network to reincorporate output results back within the network. Ismail and Abdullah [34] utilized a back-propagation ANN by inputting 13 socioeconomic and weather type factors to predict electricity demand in Malaysia. "Back-propagation" refers to the ANN feeding the output error back through the network for updates to be made until the desired adjusted output is achieved [20]. Ismail and Abdullah [34] found that hybridized with principal component analysis (which corrects for multicollinearity) this model outperformed other common model types. Al-Bayaty et al. [17] analyzed the performance of multiple machine learning models in predicting ST energy demand using meteorological data and found that the ANN model was one of the two top performers.

The other top performer identified by [17] was automated decision tree analysis. A type of decision tree analysis was validated by [7] in comparing the usage of degree days and dew point temperature in predicting power consumption for the state of Florida. In this study, a Bayesian additive regression trees (BART) model was compared against and surpassed generalized linear, additive, regressive, and other decision tree-based models in predictive ability. BART is a non-linear, non-parametric statistical learning method where a data space is split into subregions and simple models are fitted to each subregion. Upon correction for covariates (via splitting), the subregions are compiled to form the final model. The value of this model lies in its freedom from parametric assumptions (granting the model better prediction capability), is robust against outliers, and is fully probabilistic (allowing it to yield full distributions of predicted response values). However, its non-parametric nature makes results more difficult to interpret [7].

This method was later utilized by [19] to show the asymmetry of sensitivity to climate for different energy demand intensity levels.

Support vector regression uses non-linear mapping to project data onto a higher dimension space, thus simplifying regression [11]. De Felice et al. [11] specifically utilized support vector regression, with the addition of principal component analysis, to explore the high-dimensionality of its input data and exploit all possible patterns for forecasting. Son and Kim [5] used a hybrid form of support vector regression to precisely forecast month-ahead electricity demand in the residential sector. Fuzzy-rough particle swarm optimization (genetic algorithm type optimization) and support vector regression were hybridized to both optimize regression model variables and account for cyclical trends. This method was shown to outperform ANN, auto-regressive integrated moving average, multiple linear regression, and other previously proposed predictive models in the following error types: mean absolute percentage error, mean absolute error, root mean squared error, and mean bias error.

The final reviewed model is the autoregressive distributed lag (ARDL) model. ARDL has been lauded as "the major workhorse in dynamic single equation regressions" particularly because of its error-correction ability [43]. In other words, it skillfully accounts for the turbulent and cyclical nature of climate and socioeconomic variables. This model has been effectively utilized in determining the ST and LT impacts of climate on energy demand in Australia by [13].

2.3 Limitations and Future Research

One limitation within the current literature is data accessibility issues. As previously referenced, global changes in climate will affect each climate region differently. As such, it is important to include not only meteorological variables that have been supported with literature, most notably some form of temperature, but also a variety of variables that capture the uniqueness of a given geographical area. Apadula et al. [10] faced this issue with meteorological data and must use only temperature (assumed to be dry bulb temperature), relative humidity, wind speed, and cloud cover because of availability. Son and Kim [5] were pointedly more exhaustive with climate variables, identifying that some variables were not analyzed because the data was not available for the entire period of their analysis. Though the variables selected in both of these studies are supported by or were chosen because they were supported by the existing body of knowledge, they are not as exhaustive as they could be.

Chandramowli and Felder [6] exposed a gap in the literature regarding analyses for under-developed and developing economies. Even today the frequency of studies in these economies is low. These countries are likely in the process of building their asset management capabilities and may not have the clean and high-quality historical data necessary to develop in-depth, accurate analyses of energy trends. However, researchers should continue seeking opportunities to provide these analyses though it may take more time for under-developed and developing economies to obtain the necessary data resources.

Another limitation is the lack of studies at the campus level of analysis. The challenge of analyses at the campus level is energy data accessibility. Some municipalities do not have a singular method of procuring energy. Therefore, multiple energy providers may serve the same city or town [33]. Analyses at this level would require more preliminary effort to collect and compile the necessary energy data, especially for smaller cities or towns. Studies that provide the closest thing to the campus level of analysis are those that either aggregate or disaggregate energy data from the facility or state/region level to the appropriate level [12], [18], [32].

Although briefly mentioned in this review, principal component regression (PCR) has not been fully utilized for its ability to counteract multicollinearity. Often meteorological variables overlap in explaining variance for response variables. For example, there are three types of temperature commonly measured (dry bulb temperature, wet bulb temperature, and dew point temperature) and regressing all of them together would result in biased results. PCR is a combination of principal component analysis and multiple linear regression. This method can extract the unique contributing aspects in the relationship of independent variables and the dependent variable and regress based on those aspects. Principal component regression is effective because it reorients regression models to highlight patterns (called components) rather than input variables. This is done by normalizing variables to capture the multidimensional trends of the error within the model. Upon transforming the trends into linearized components, the original variables are correlated to each component to evaluate which variables are, in fact, most influential to the model [44]. Unlike in Vu et al. [39], this method does not remove entire variables that appear redundant, but discards redundant components of each variable and pulls the unique qualities from even the most similar variables to better explain relationships with the response variable.

Several of these limitations provide a launching point for the thesis research herein, where a predictive model and forecast of a military installation's total energy consumption is pursued. This research will address data availability by utilizing long-established meteorological data archives, such as the National Oceanic and Atmospheric Administration, to create an extensive list of climate variables. The campus level of analysis is achieved by the compilation of energy data from Wright Patterson AFB, which functions at this level. Finally, cross-validated PCR is applied through the predictive model's development framework to account for bias and multicollinearity.

III. An Unbiased Climate-Informed Tool for Campus Energy Policy Development and Budget Predictions

3.1 Abstract

Climate variability is an external factor that creates energy demand uncertainty and makes energy consumption modeling complex. Poor predictions can lead to energy overages resulting in the need to borrow funds from fixed operations and maintenance budget areas. In order to make data-driven management decisions, energy managers require consumption prediction models that account for climate, and that can skillfully and simply anticipate energy consumption. This research uses a statistical model-based approach and open-source climate data to predict hourly energy consumption for a campus-sized community (population: 30,000). The skill of several model configurations, informed with combinations of consumption periodicity, climate, and temporal state variables, are tested. The modeling framework consists of a cross-validated principal component regression followed by post-prediction statistical bias correction. Deterministic prediction skill is measured by mean absolute percentage error (MAPE) and contingency tables. Ensemble predictions are created from the model error distribution and categorical skill is determined using ranked probability skill score (RPSS). Deterministic hindcasts explain more than 73% of variability in hourly energy consumption. The top-performing model achieves an RPSS of 59% and is skillful year-round. The cumulative results suggest that incorporating forward-looking projections could be useful for long-term utility policy development and budgetary planning when coupled with price forecast models.

3.2 Introduction

Climate variability is an external factor that causes energy demand uncertainty and makes energy consumption modeling complex. Poor predictions can lead to energy overages resulting in the need to borrow funds from fixed operations and maintenance budget areas.

Energy managers require models that can skillfully and simply predict energy consumption in order to make data-driven management decisions that inform policy and advocate for energy operating budgets. Managing energy at the campus level is particularly difficult because available resources and funding must be distributed across all facilities and infrastructure [45]. If campus energy managers can skillfully predict levels of performance, then they can bolster asset resource and funding resilience.

Due to the impact of greenhouse gas (GHG) emissions on climate, many international organizations and institutions have allocated time and resources to identify and quantify the potential risks to businesses, governments, and communities that stem from extreme events or gradual trends in climate. One key area of investigation is gradual changes in climate, and its impact on energy consumption [5], [9], [10], [12]–[14], [28]. For example, Zhou et al. [14] model the impact of changing climate across the 21st century for United States heating and cooling, finding that heating energy demand should drop gradually for all states, but the inverse will happen for cooling energy demand. Emodi et al. [13] study electrical energy consumption in Australia and finds that electrical energy consumption may slightly decrease in the middle of the 21st century before increasing and eventually surpassing current energy consumption levels due to internationally accepted climate change projections.

Energy is an important recurring and variable cost for any organization or sector, independent of size [7], [9], [10], [17], [18], [38]. For example, studies such as[9] Amato et al. [9] focus on regional energy impacts due to spatial differences in energy infrastructure, while De Rosa et al. [18] focus on facility impacts and policies designed to bring about relatively quick energy savings. Residential, commercial, and industrial sectors spent $224B, $201B, and $359B on energy in 2017, respectively [4]. In a national context, United States of America energy users consumed 97.6 trillion Btus at the cost of $1.1T, or 5.8% of gross domestic product.

Applying energy consumption modeling at the campus, or organizational, level has received little attention. As the owner of 800 military installations that operate at the campus level, Department of Defense (DoD) installations provide a unique opportunity to apply statistical modeling at such a scale. Despite broader federal action, the prospect of changing climate has driven the U.S. Congress to question the resilience of DoD installations and its operations. Consequently, Congress tasked the DoD to investigate climate-change-driven impacts on many aspects of its asset and mission portfolios [1], [2]. As the U.S. Government's largest operational and facility energy user, the DoD must consider how changing climate may affect energy demand and use patterns [3]. Accurate predictions of future energy costs are essential for informing organizational and facility operating budgets, developing use and energy management policies, and integrating technologies to mitigate potential energy cost variability and achieve energy savings.

Despite the contributions of the aforementioned research studies, there is limited reported research that: (1) utilize various bias mitigation techniques like cross-validated principal component regression (PCR) and statistical correction, and (2) evaluate energy consumption at the campus level. Accordingly, the objectives of this research aim to fill several knowledge gaps in climate-informed energy management. First, this work will determine whether a climate-informed statistical approach is skillful in predicting historical campus-level (multi-facility) energy use. Few studies exist that focus on statistical modeling of organizational-level energy use and fewer have collected historical data corresponding to all operations of the specific organization or campus. The second objective is to test different model configurations to identify the value of climate and other relevant predictors in explaining variation in energy use. Though an extensive list of input variables is initially desirable, simplified yet skillful models are those most desirable to energy managers. A case study is used to calibrate the models herein. However, the generalized framework is flexible such that any number of continuous or categorical independent variables can be regressed.

This research uses observed, campus-level, total energy consumption data and a variety of open-source climate data to create various energy consumption prediction models using combinations of differing input types to achieve these objectives. Additionally, cross-validated PCR and post-prediction statistical bias correction are adopted as methods to create deterministic predictions, and account for independent variable multicollinearity and model bias. After the initial models are developed and tested, a lean model is developed using the most statistically influential input variables from the most skillful model to determine the effects of limiting input variables. This model is a tool to aid asset managers in planning operations and energy infrastructure.

3.3 Background

Statistical, climate-driven prediction models have been used across many fields to gain insight into past and future impacts on operations and to inform policy. Models applied to the management of built and natural systems are applied broadly and produce results with varying degrees of deterministic and probabilistic skill [35], [46]–[48]. For example, Delorit et al. [35] model streamflow in a Chilean river basin and uses ranked probability and categorical hit skill scores to determine model skill; whereas, Zeng et al. [48] model the wind speed and solar irradiation in a Yangtze River estuary and uses Akaike Information Criterion, adjusted coefficient of determination, and mean squared error to determine model skill Similar models have been developed for the energy sector at a wide range of temporal, spatial, and organizational scales, and are most commonly calibrated to evaluate energy consumption with mention of climate impacts [5], [9]–[11], [13], [36].

Many climate variables have been shown to provide value in energy consumption prediction models. In cases of limited access to data, variables are selected based on intuition or expertise in a specific area [5]. In cases of limited computational capacity, the ability to perform exhaustive analyses may be limited considering that energy managers are typically not modelers or climate scientists. However, existing literature has highlighted specific key climate variables that may help researchers limit their search space, including temperature [7], [9]–[12], [14], [17], [19], [29], [34]; relative humidity [10], [12], [17]; cloud cover [10], [17]; precipitation [7], [19], [36], [37]; wind speed [7], [17], [19], [29]; and irradiation [17], [18], [38]. This study builds on these previous works and investigates a multitude of climate factors for their explanatory power in predicting energy use.

The spatial scale of energy prediction analyses is primarily driven by data accessibility or the specific focus of researchers, for example: the impact of climate variables, new method development, or the sensitivity of models to location. The effects of climate change vary spatially; therefore, many models are calibrated to a specific climate zone, and once optimized, tested across climate zones to determine model exportability [6]. With this in mind, the highest resolution energy data available and an exhaustive list of climate variables is collected for this study.

Energy prediction analyses occur anywhere between facility to multinational scales. At a facility-level specific building types have been modeled and exported to various climate zones to understand spatial impacts of a changing climate on those facility types [18], [32]. Few organization or campus-level studies exist, though bottom-up facility aggregation methods are generally used to model this level. Dirks et al. [12] utilize building energy modeling software, containing approximately 26,000 facility types, to model entire geographical regions of the U.S. and inform energy mitigation and savings techniques. These methods require robust datasets. The difficulty of campus-level analyses lies in capturing the different use regimes between facilities and accessibility to facility-level data, thus limiting these analyses when assumptions are made. Bottom-up efforts, which rely on robust facility-level datasets to model energy usage of each facility are difficult to calibrate, time-intensive to construct, and are highly stylized to the campuses for which they are developed. Additionally, some municipalities do not have a singular method of procuring energy; therefore, multiple energy providers may serve the same city or town, making the collection of the data burdensome [33].

Like facility-level models, state- or region-level models can be exported to other states and regions to test the model's stability under a diverse array of climate conditions [7], [19]. Additionally, state-level data can be disaggregated to better understand business sector energy consumption and to capture the spatial heterogeneity of building use within each state [14]. For national or multinational scale, the difficulty lies in data collection. Wenz et al. [31] gather electrical energy consumption data from across Europe to develop their wide-reaching study and provided a better understanding of Europe's predicted peak energy consumption under climate change.

Finally, Chandramowli and Felder [6] present a review of energy consumption prediction methods that found multiple linear regression to be one of the prominent techniques [6]. Other technique types include fuzzy regression [41], Bayesian additive regression trees [7], [19], support vector regression [5], [11], and artificial neural networks [17], [34]. However, no studies have leveraged principal component analysis (PCA) with regression, much less with cross-validated multiple linear regression to account for multicollinearity and bias present in climate and other predictors.

3.4 Data

3.4.1 Energy Data

Few studies have focused on creating prediction models and analyses for a municipality or campus using that campus's historical energy data. This is largely due to limited data availability and resolution, which hamper statistical significance. For this study, energy consumption data were provided by Wright Patterson Air Force Base (WPAFB), located near Dayton, OH, across four consecutive years at the half-hourly scale, given in kilowatt-hours (1 Oct 2015 - 30 Sep 2019). The scale of these data most closely resemble that of a city, manufacturing complex, or medical or university campus. WPAFB employs over 30,000 people and includes various operation types such as, industrial, commercial, community support, and residential. In all, the data include energy demand from approximately 26,500 facilities. In terms of climate conditions, Dayton, Ohio is a temperate climate with moderate rainfall throughout the year, warm to hot summers, and cool to cold winters. From the time series of the observed data, it is apparent that a dominant signal exists, which suggests that periodicity and categorical (dummy) time variables could be valuable inputs to energy consumption prediction models (Figure 1).

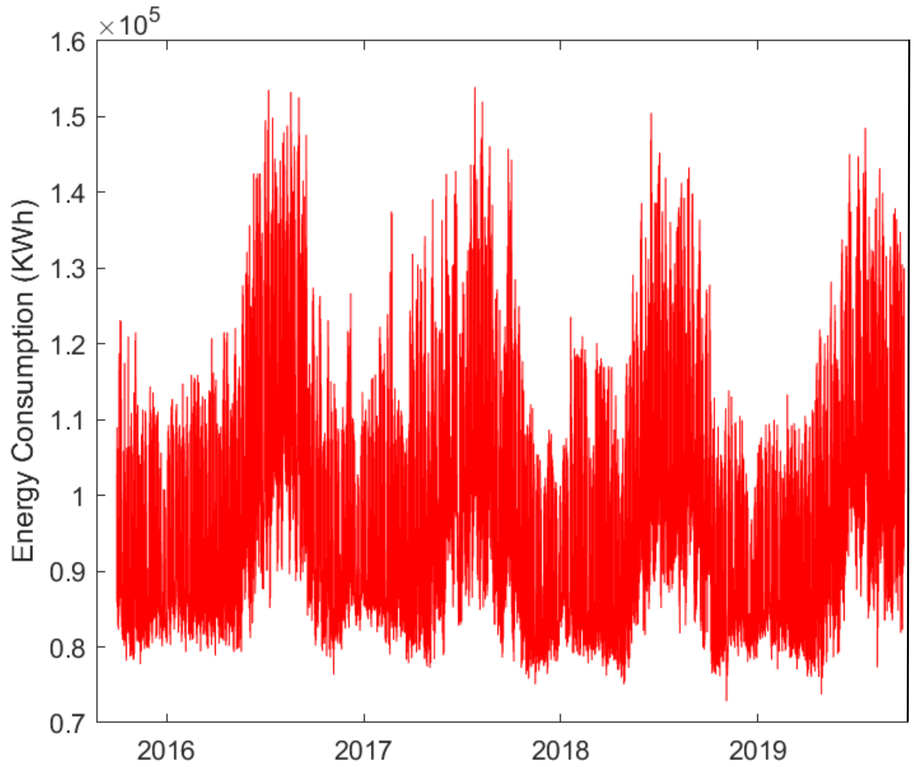

Figure 1. Observed energy consumption timeseries: Four years of half-hourly raw observed energy data for WPAFB show a relative periodicity in electricity demand with spikes in the summer months due to space conditioning. The tick marks on the horizontal axis represent the start of the specific calendar year (i.e., January 01)

3.4.2 Climate Data

A majority of the climate data used in this prediction framework was retrieved from the NOAA Local Climatological Data (LCD) database to include dry bulb temperature, wet bulb temperature, dew point temperature, relative humidity, station pressure, sea pressure, wind speed, precipitation, and cloud fraction. From the existing literature, solar irradiation was noted as potentially impactful but was not available in the NOAA LCD database [17], [18]. Therefore, irradiation, along with cloud opacity and precipitable water, were obtained from the commercial solar forecasting company Solcast

[49]. Incorporating irradiation, opacity, and precipitable water were deemed important to ensure a broad analysis of climate's contribution to energy consumption.

The energy consumption data was aggregated to an hourly scale because it was the finest temporal resolution common across all climate variables. Using a Fourier Transform, the underlying periodicity was extracted and time variables representing various daily, weekly, monthly, and yearly time-step scenarios were developed as input variables. Dummy or categorical time variables can inform energy managers of temporal levels at which data should be tracked to produce accurate energy predictions. The data was then preprocessed by consolidating input and response variables, removing time-steps with missing information, and formatting.

3.5 Methodology

Statistical modeling techniques are used in many fields for attribution and forecast analyses. This research uses combinations of three variable types (climatology, periodicity in the underlying energy consumption pattern, and time) in a cross-validated PCR to create ensembles of statistically unbiased energy consumption prediction models. This process is depicted in Figure 2, where the lighter arrows represent the process being repeated for the lean model. Input variable types are combined to produce the following models (7): periodicity only, climate only, periodicity and climate, periodicity and time, climate and time, and a collective model (combining climate, periodicity, and time variables). Testing various input combinations will ultimately determine the model that best balances performance and complexity. The skill of each model is determined using several standard validation techniques, including mean absolute percent error (MAPE), ranked probability skill score (RPSS), and contingency table exposition. After identifying the skill of the best performing deterministic model, a lean model is created by incorporating only the most influential input variables, and recompleting the production and validation sequence. The purpose of the lean model is to test the skill retention of the best performing deterministic model while reducing input variable complexity. This entire process is depicted in Figure 2.

The resulting methodology is intended to be applicable for any location; however, this research and resulting models have been calibrated for the area of Dayton, OH, where the energy consumption data was retrieved.

Figure 2. Systematic framework visualization: numbers correspond to section and subsection; black arrows represent the initial model development process, and grey arrows represent how the lean model is developed by reusing the initial process

3.5.1 Periodicity and Time Variables

To isolate the dominant signal in the energy consumption time series, a Fourier Transformation was conducted. Fourier Transformation approximates the underlying signal of time series data through the superposition of sinusoids. The underlying signal of the observed energy data contained eight sinusoidal components. As components of the signal are taken away, the signal captures fewer trends in the observed data [50]. Ultimately, six components of the overall signal were selected to represent the periodicity of the observed data, because a less complex signal could be achieved while not compromising the interannual trends visible in the observed data (Figure 3).

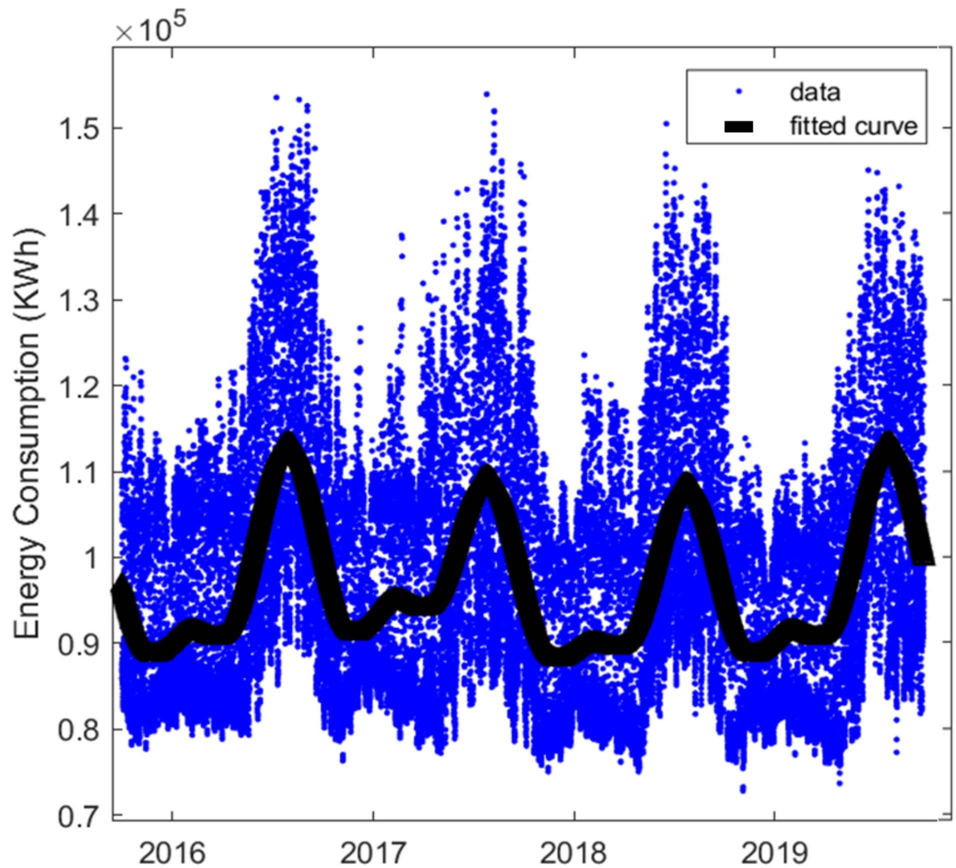

Figure 3. Fourier Transformation timeseries; the observed data's underlying signal is captured; larger peaks in energy consumption are visible during the summer periods, while smaller peaks occur during winter periods.

Through capturing the dominant signal within the observed data, it is apparent that categorical time-steps might also be skillful predictors of energy consumption. For instance, it appears that, annually, peak energy consumption in summer months reach approximately 150,000 kWh. To test this phenomenon, categorical time ("dummy") variables are introduced as input variables. Because of the fine temporal resolution of the data, a variety of time variables are included (hour of the day, day of the week, weekday vs. weekend, month, heating vs. cooling vs. no-heat-no-cool seasons, and fiscal year).

The final list of inputs includes three variable types (climate, periodicity, and time), totaling 59 specific input variables. The input types are combined in various ways to make the seven models. Only one model, termed the collective model, contains all 59 possible input variables.

3.5.2 Description of the Models

Six different combinations of input variables, that generate the six models tested herein, are analyzed to find the combination of input variables that produces the best performing and lowest complexity model. All six models are generated using the same framework (described in sections 3.5.3-3.5.5).

The periodicity only model contains the Fourier Transformation, or underlying signal of the observed energy data, as the only input variable. The climate only model contains 12 input variables, including dry bulb temperature, wet bulb temperature, dew point temperature, relative humidity, station pressure, sea pressure, wind speed, precipitation, precipitable water, cloud fraction, cloud opacity, and irradiation. The periodicity and climate model consists of 13 variables, including of all of the variables from both the periodicity only and climate only models. The periodicity and time model consists of 47 input variables, including the single variable in the periodicity only model and all of the categorical time variables (hour of the day [23], day of the week [6], weekday vs. weekend [1], month [11], heating vs. cooling vs. no-heat-no-cool seasons [2], and fiscal year [3]). The climate and time model consists of 58 input variables, including all of the variables from the climate only model and all of the categorical time variables. The last model, the collective model, consists of all 59 input variables, including all variables from the periodicity only and climate only models, and all categorical time variables.

3.5.3 Cross-validated PCR

With the final input variables established, multicollinearity is addressed through cross-validation and PCR [51]. Delorit et al. [35] explain that PCR is commonly applied in forecasting and hindcasting to reduce both variable dimensionality and multicollinearity, and result in a set of principal components (PCs) that represent the variance in a set of predictors [35]. First, PCA is conducted where input variables are broken down into their PCs. Next, a leave-one-out cross-validated hindcast is undertaken across the entire dataset to produce a less biased, deterministic prediction of expected energy consumption for WPAFB. Because this form of cross-validation removes the time-step being predicted, the percentage of variance explained by the model will generally decrease. Furthermore, Jolliffe's Rule is applied as a PC retention and dimensionality reduction technique [52]. Only the most influential PCs for the prediction model are retained. The coefficient of variation falls as fewer PCs are retained for the regression. The cumulative effect of the cross-validated PCR is an unbiased and conservative variance explained estimate.

3.5.4 Statistical Correction

Statistical bias correction, also known as quantile mapping, is prevalent in climate forecast modeling [53]–[55]. By comparing the fit of the regressed models to the observed data, statistical model bias can be identified and corrected. Statistical correction methods account for consistent bias across a model. To correct the models, the distribution type of the observed energy data is identified. Using the associated distribution parameters, the distribution of the predictions is matched to the distribution of the observations using quantile mapping. The resultant outputs are the final deterministic prediction models.

The observed energy data follows a bimodal normal distribution that necessitates a uniquely tailored statistical correction process. Normal distribution parameters from both distribution “modes” are collected to perform the statistical correction. This method requires both the observed and modeled data to be split at a calibrated point while maintaining time-step position indexing. Each “half” of the modeled data is corrected based on the corresponding “half” of the observed data, and the two “halves” of the model are reassembled to produce the statistically corrected model.

3.5.5 Validation Metrics

Deterministic model performance is illustrated using mean absolute percent error (MAPE). MAPE is commonly used in energy prediction research and established thresholds are used to determine the skill of prediction models [5], [28], [56]. When utilizing MAPE, a score below 20 signifies a prediction model of "good" quality. If the MAPE falls below 10, the forecast model is said to be of "excellent" quality [57].

Uncertainty is incorporated into the finalized models through prediction ensemble generation. Ensembles are used to calculate ranked probability skill score (RPSS), which is a metric of probabilistic, or categorical, performance and is meant to account for uncertainty in the deterministic models' outputs. First, a reference climatology is established by separating the distribution of observed data into categories based on the characteristics of the distribution. This becomes the standard against which the prediction ensembles are tested. Climatology is scored based on the percentage of observed data points that fall within each category, while the prediction model is scored based on the number of ensemble predictions that fall in the same category as the observed data. For this research, the climatology was created by partitioning the distribution into four categories based on peaks and saddle points within the distribution, as shown in Figure 4.

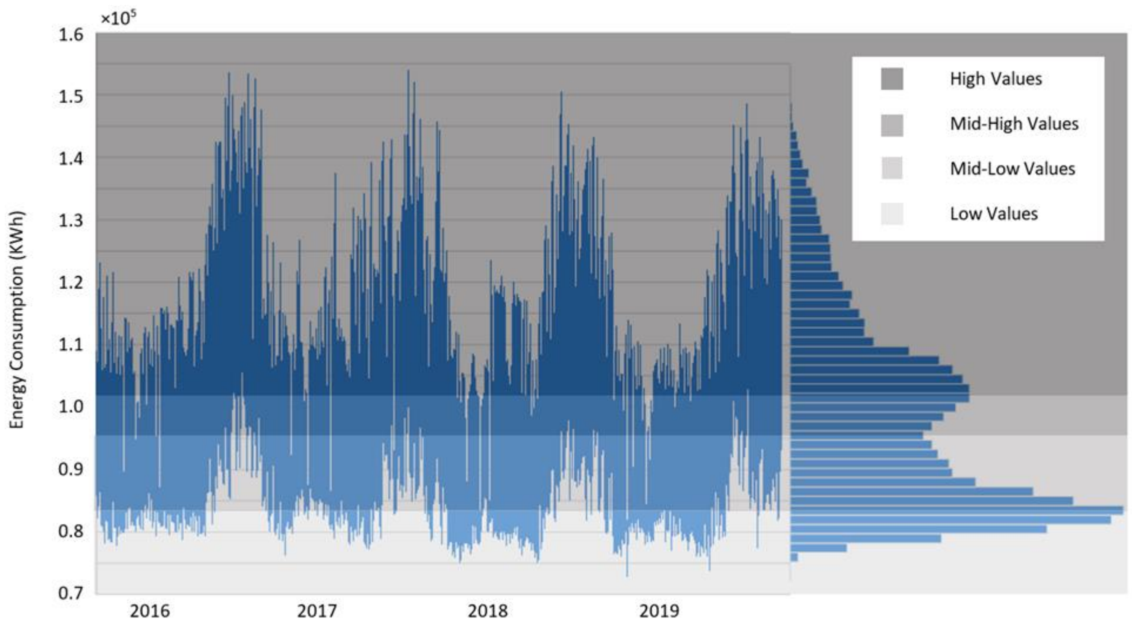

Figure 4. Observed data timeseries with climatological categories; depicts the chosen climatological categories based on the character of the accompanying histogram. Each category is not indicative of a single season; however, general seasonal associations can be drawn. A majority of the high values occur during the summer and a majority of the low values occur in the fall, winter, or spring. The intermediate categories include various portions of all seasons.

An ensemble is generated by randomly drawing a value from the deterministic model's error distribution (with replacement), and adding the error to a deterministic model outcome. This process is repeated until a randomly drawn error is added to each outcome in the deterministic time series. The result is a new ensemble time series, where the addition of the error terms is a means of applying uncertainty. The number of ensembles generated for each model is determined by the number required to achieve RPSS variance <0.01.

Ranked probability skill score (RPSS) is a categorical measure of ensemble prediction compared to a reference forecast, in this case the climatology discussed above [58]. The RPSS uses the mean ranked probability score ($\overline{RPS}$), a measure of the square differences in the cumulative probability of a multi-categorical ensemble. The RPSS ranges from -∞ to 1, where values greater than zero indicate greater skill than climatology. RPSS values less than zero indicate that predictions are inferior to climatology, and a RPSS equal to zero indicates that the model is equivalent to climatology [35]. An RPSS value is generated for each time-step of the hindcast using Equation 1, and the median value of all time-steps is reported as the RPSS for the model formulation.

$$RPSS = \frac{\overline{RPS} - \overline{RPS}_{reference}}{0 - \overline{RPS}_{reference}} = 1 - \frac{\overline{RPS}}{\overline{RPS}_{reference}}$$

Equation 1

Where:

$\overline{RPS}$ = ranked probability score

Contingency tables are leveraged to test the categorical performance of the top performing deterministic model. The contingency tables applied in this research are separated into the four categories representing climatology, as described in Figure 4. Hits are defined as hourly predictions that align with the correct observed consumption category, and misses are hourly predictions that do not fall within the same category as observed consumption. Extreme misses are defined here as those predictions which miss by more than two use categories and are unequivocally wrong.

3.5.6 Lean Model Compilation

Once the skill for each of the seven deterministic models is evaluated, a lean model is assembled using only the most dominant input variables from the most skillful model. Dominance is determined by correlating the model's retained PCs with the original input variables. The following steps were followed to identify the input variables with the greatest signal:

1) Isolate the top two-thirds of retained PCs. This decision is arbitrary, but serves to illustrate that an energy manager could down-select to the number of PCs desired based on data availability.

2) Select input variables from each PC with the absolute value of correlation coefficients greater than 0.30. This is done as 0.30 is widely regarded as "moderate" correlation.

3) Retained input variables are those that occur most often in the remaining PCs.

The new model is then redeveloped through cross-validated PCR, statistical correction, ensemble generation, and skill analysis (see Figure 1). The model's statistical performance is then compared to the initial models to determine the effect of including less but the most important information from the larger input variable set.

3.6 Results

The results are organized such that they reflect the order of the methodology. First, the effect of statistical correction on the deterministic skill of the models is addressed. Next, the predictive capabilities of the models tested are presented and compared. Lastly, the performance of a lean model, developed from the top performing model of the original six models, is compared to the performance of the original six models.

3.6.1 Bimodal Corrected Models

It was found that varying the statistical correction splitting points of both the observed and modeled data resulted in a different variance explained, and different fits to the observed data cumulative density functions (CDFs) and probability density functions (PDFs). Through manual calibration of the split points, models resulting in the best fitting CDFs were obtained. Figure 5a illustrates the bimodality of the distribution of the observed data (blue); however, the modeled data (orange) does not take this phenomenon into account. As depicted in Figure 5b, quantile mapping and calibration of the split points create a corrected model (orange) that better matches the bimodality of the observed data.

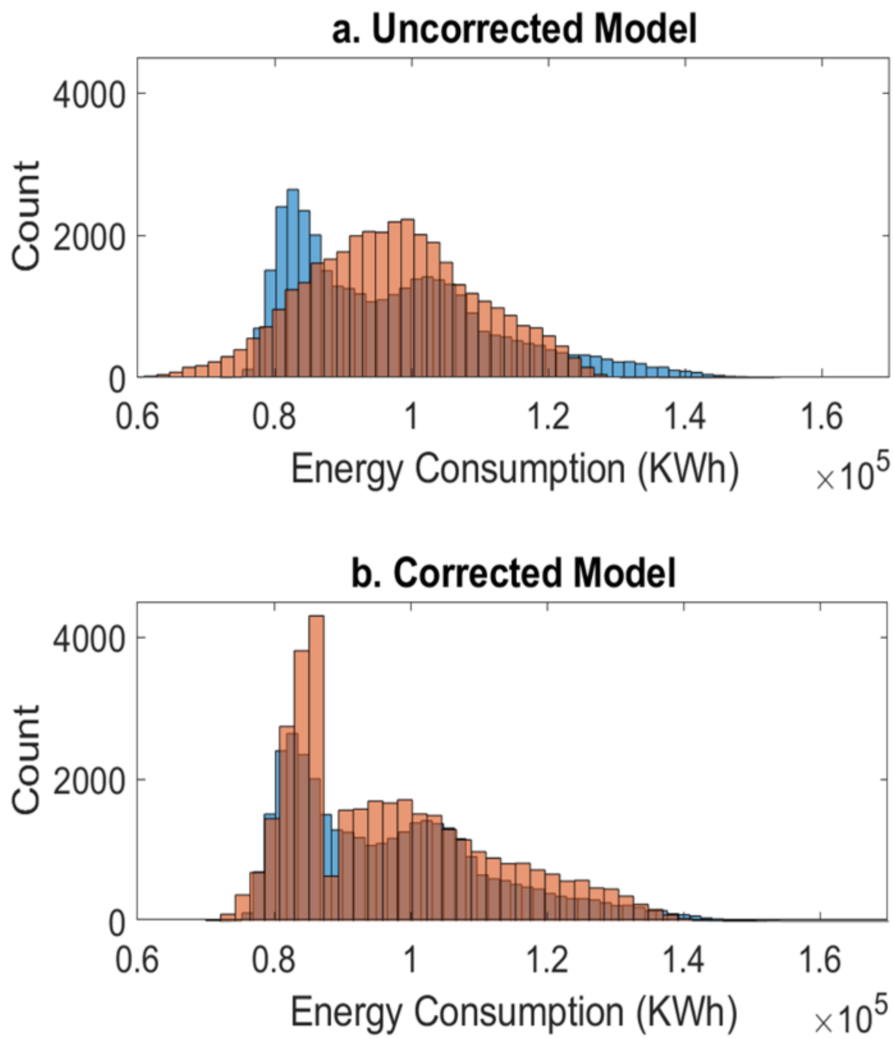

Figure 5. Impact of statistical correction on a prediction model; (a) modeled data (orange) over observed data (blue); (b) corrected model (orange) over observed data (blue); statistical correction adapts the predictive model to better capture the bimodality of observed energy consumption.

3.6.2 Predictive Capabilities

Deterministically, the top performing models are the collective and climate and time models, as each produces an explained variance of 0.73 (r^2) (Table 1). Probabilistically, the significance of incorporating statistical correction into model development is manifested as a 9% average increase in RPSS across all models, except for the periodicity only model (1% improvement), which is likely due to the fact it is extracted from the observed data.

Table 1. Model performance metrics

Models	Variance Explained (r^2)	MAPE	RPSS	Dimensionality Reduction	Dominant signals for PC1 and PC2 (input name, *Pearson's coefficient of correlation*)	
					PC1	PC2
Collective	0.73	6.25	0.57	30.5%	FourierTrans (0.87)	FourierTrans (0.94)
Climate and Periodicity	0.44	9.62	0.30	61.5%	FourierTrans (0.77), Irradiation (0.41), DewPtTemp (-0.37)	FourierTrans (0.96)
Climate and Time	0.73	6.15	0.59	26.3%	DewPtTemp (0.95)	DewPtTemp (0.94)
Periodicity and Time	0.55	8.22	0.39	27.7%	FourierTrans (0.95)	Wednesday (0.52), Weekday/Weekend (-0.40), Thursday (0.39)
Climate Only	0.43	9.83	0.29	58.3%	DewPtTemp (0.94)	DewPtTemp (0.90)
Periodicity Only	0.27	11.53	0.19	0%	FourierTrans (1.0)	

Dimensionality reduction compares the number of retained PCs to the initial number of PCs for each model. Specifically, it is the ratio of the difference between the initial number of PCs and the final number of retained PCs in a model to the final number of retained PCs in a model expressed as a percentage. Reducing dimensionality in a model is important because it reduces the complexity of the model and highlights what input variables are not necessary to produce the model skillful. In other words, it narrows the scope of input variables that energy managers must collect and input into a model.

A substantial number of PCs are retained for the models with larger input sets (collective, climate and time, and periodicity and time). Because the periodicity only model contains one variable, there is no reduction in dimensionality, and therefore, the PCR process is not useful.

Additionally, the dominant signals from PCs 1 and 2, which always explain the most variance in the underlying input data, are recorded in Table 1. The periodicity variable ("FourierTrans") and dew point temperature ("DewPtTemp") are consistently designated as key signals in the first two PCs of many of the models. Whenever periodicity is included in a model, the periodicity variable is notably the most dominant signal. When climate is included in a model, dew point temperature is either the first or second most dominant signal, which suggests that temperature is also the main feature of the periodicity variable. However, the first two PCs of the models are not entirely dominated by these two inputs. Solar Irradiation, Wednesday, Thursday, and weekday vs. weekend are other notable input variables. Several of the week related time variables that emerge as influential are consistent with trends regarding the flow of people onto the installation (personal communications, WPAFB security personnel).

The performance metrics identify that some models do show particularly encouraging skill. All models, with the exception of the periodicity only model, produce MAPE scores consistent with "very good" prediction/forecast model candidates (< 10), though the periodicity only model is considered “good.” A total of approximately 100 ensembles for each model formulation were computed to achieve RPSS convergence (<0.01 score deviation). Clearly, higher RPSSs stem from models with larger deterministic variance explained. Periodicity provides predictive power only when coupled with categorical time variables. When paired with the climate variables, periodicity provides a slight improvement from the climate only model. And, in the collective model, periodicity adds very little improvement; performance is approximately the same as that of the climate and time model. This suggests that the information provided by the periodicity variable is represented by the climate and categorical time input variables.

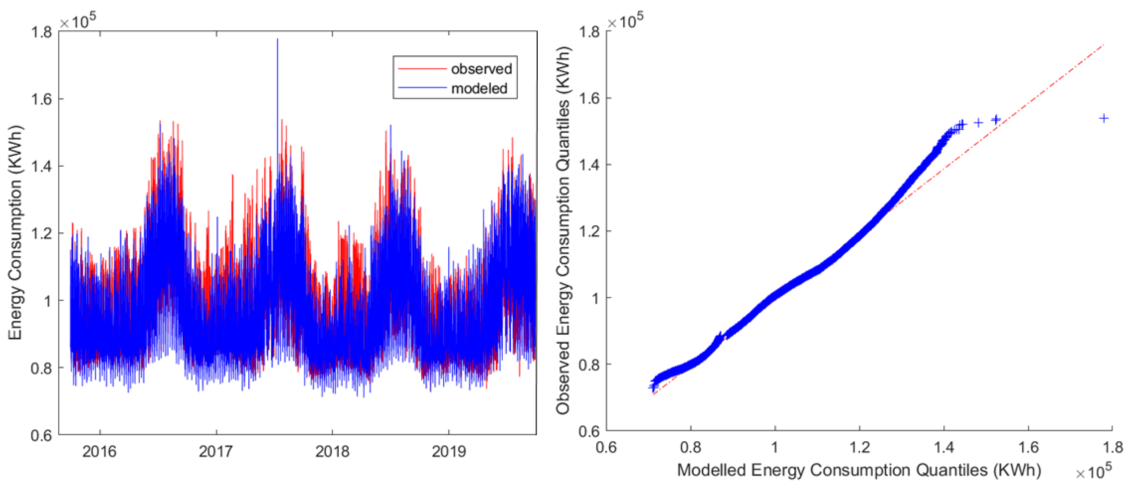

Figure 6. Final Climate and Time model versus observed energy consumption; time series and quantile-quantile plot; the single outlier occurs in the summer and is likely due to an erroneously high temperature measurement, and the response in energy consumption is due to cooling demand

Because the climate and time model was the least complex and highest performing model, it was used to create two lean models consisting of only those input variables with the most dominant signals. After cross-validated PCR, 42 of 58 PCs were retained. By correlating PCs to the specific input variables, it was determined that the inputs with the most dominant signals include the three temperature variables (dew point, dry bulb, and wet bulb) and the time variables weekday/weekend, January, February, June, Sunday, Friday, 1100 hours, 1400 hours, 1500 hours, 1600 hours, and 2300 hours.

Contingency tables are leveraged to test the categorical performance of the top performing deterministic model (climate and time). Hits and misses are expressed as a percentage of the total number of forecasts in each climatological category. In the contingency tables, hits appear along the diagonal from top left to bottom right. It follows that misses appear as a divergence from the diagonal. The hit scores align closely to model RPSS; however, the extreme miss score is new information, and represents cases when the prediction was for low energy consumption, but the actual energy consumption was high, and vice versa. The hit score of the best performing deterministic model (climate and time model) is 58.6%, and the extreme miss score is 7.9%. Additionally, the hit score for the highest and lowest use categories is 72%. While the overall hit score is unimpressive, the model's performance in the extremes is encouraging. If the energy manager's goal is to avoid extreme misses, and maximize skill in predicting extreme use times, then the model should be preferred over a climatological analog.

Table 2. Median energy consumption value contingency tables for the climate and time model (%); regions: L=low values, ML=mid-low values, MH=mid-high values, H=high values; green=hit, red=extreme miss

		Modeled			
		L	ML	MH	H
Observed	L	61.6	24.2	2.8	0.5
	ML	36.9	50.6	33.6	8.4
	MH	1.2	12.8	25.6	14.1
	H	0.3	12.4	38.0	76.9

To further analyze the model's categorical skill under forecast optimism and pessimism contingency tables were developed for the 75th and 25th percentile ensembles, respectively. The 75th and 25th percentile contingency tables support what is observed through graphical representation (Fig. 6). The deterministic model tends to predict below the lowest energy levels and above the highest energy levels. The 75th percentile ensemble more accurately accounts for variability in low energy consumption values than the median predictions (61.6% → 78.6%), and the 25th percentile ensemble is more skillful for high energy consumption values than the median predictions (76.9% → 85.8%).

Being that the High and Low categories are likely to be of greatest importance to energy managers, the contingency tables for the median predictions can also be readapted to consolidate the middle two regions (Mid-Low and Mid-High) to a single category because specificity in these regions may not be necessary or important to energy managers. The result is an increased hit score of 67.7% and a decreased extreme miss score of 0.17%. The increase can be attributed to the higher accuracy in the new "Middle" region due to the consolidation of the Mid-High and Mid-Low regions, and the decrease in extreme misses is attributed to the fewer opportunities for values to fall in extreme miss categories.

Though the median prediction values perform well categorically, improved predictions can be achieved for the extreme regions using the 75th and 25th percentiles. That being said, the results for the median prediction values show that just by knowing climate and time variables, we can predict 76.9% of the variability in peak hourly consumption can be explained. Using 25th percentile predictions, explained variability rises to 85.8%.

3.6.3 Lean Models

Two lean models are generated with different combinations of input variables to specifically analyze the effect of including categorical time variable types rather than only including specific time variables. In terms of model calibration, it would likely be just as easy to incorporate all variables in a categorical time variable type as incorporating specific categorical time variables. Lean model A consists of 44 input variables, including all of the temperature variables (dew point temperature, dry bulb temperature, and wet bulb temperature) and only the most impactful time variable types (hour of the day [23], day of the week [6], weekday vs. weekend [1], and month [11]). For example, since several specific hour-of-the-day variables are noted as being impactful, the entire set of hour-of-the-day variables were included in the model. Lean model B consists of 14 input variables, including only the specific inputs with the most dominant signals (dew point temperature, dry bulb temperature, wet bulb temperature, weekday vs. weekend, January, February, June, Sunday, Friday, 1100 hours, 1400 hours, 1500 hours, 1600 hours, and 2300 hours).

Lean model A maintains higher performance results compared to the six original models, while lean model B experiences larger drops in performance (Table 3). This result occurs because lean model A contains more total input variables than lean model B. However, lean model B still outperforms three of the six original models (periodicity only, climate only, and climate and periodicity) and performs similarly to the periodicity and time model. The scree plot reveals that considerable drops in performance should occur for lean model B because nearly 30 variables from the climate and time model are not included, and many of those variables each explain around 2% of the total explained variance in the climate and time model (Fig. 7).

Table 3. Lean Model statistical results

Models	Variance Explained (r^2)	RPSS	MAPE
Lean Model A	0.66	0.54	6.90
Lean Model B	0.49	0.35	9.02

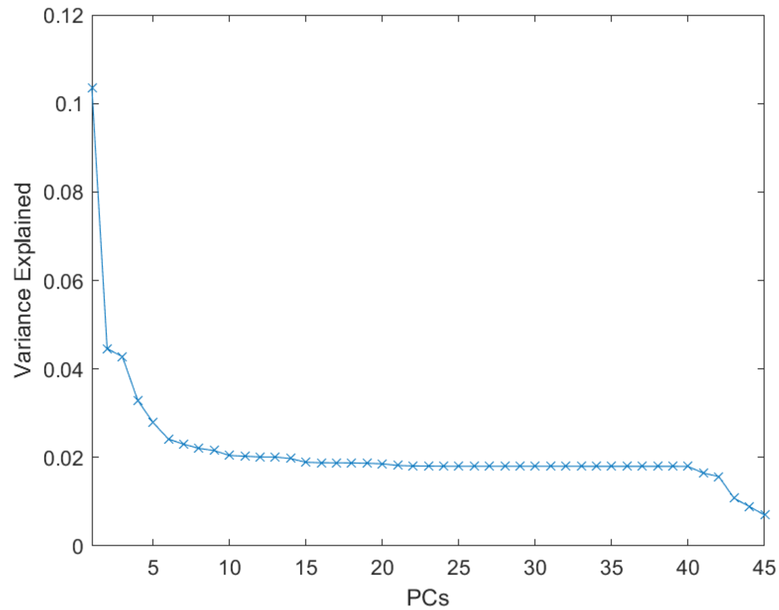

Figure 7. Climate and time model scree plot

The results tend to reflect or slightly under-perform what is found in similar studies in terms of input variable impact and model skill [5], [7], [19]. However, these studies tend to include socio-economic inputs (e.g., population increase, energy pricing, etc.), where this study purposefully does not. Additionally, in contrast to existing studies, the model developed herein analyzes energy consumption at the campus level using campus-level energy data, provides insight on model categorical performance, and proposes a framework that thoroughly accounts for bias.

3.7 Discussion

The results demonstrate that skillful predictions of hourly campus-wide energy consumption can be achieved using statistical models informed with mixtures of continuous climate and categorical time variables. Moreover, models can be created with techniques (PCR, cross-validation, and statistical correction) that minimize bias and reduce dimensionality. Furthermore, using uncertainty in deterministic predictions, a model's probabilistic skill can be determined. The skill of the proposed framework and use of open-access data suggests that energy and facility managers could be well-positioned to create their own models. The correlation between the regressed PCs illustrates that temperature and time variables are the most useful in explaining hourly energy consumption. Energy consumption patterns were used to decide which categorical time variables to include, while the temperature data was obtained from the open-access NOAA LCD database. Both of these variable types require limited effort to obtain. However, as the comparison of the Climate and Time model and Lean Model B shows, there is a significant tradeoff between reducing dimensionality and maintaining skill.

Though overall predictive strength is important, accuracy at the highest and lowest energy use periods is perhaps of greatest importance to energy managers, who must make operational decisions (e.g. load shedding), and make equipment and policy recommendations to decisionmakers. For example, predicting peak energy consumption can inform energy managers of when peaker generators, or those generators only used to compensate for peak energy periods, should be utilized or if energy infrastructure needs expansion to support increased demand. Predicting low energy periods accurately can inform seasonal decisions to override heating and cooling systems when environmental conditions are mild (e.g. spring and fall) [59].

Through categorical analysis, the best performing model appears to be skillful for these extreme energy levels. Additionally, leveraging the most skillful components of the 75th and 25th percentile predictions is a feasible solution for energy managers seeking to maximize peak and low energy use. The model also produces low percentages of extreme misses, where an extreme miss prediction could lead to poor decision making about when to load shed or when to centrally heat or cool buildings.

Energy managers are generally not modelers, and thus tools that are informed with readily available data are likely to be favored. Data accessibility, computational power, modeling ability, and time availability could be factors in model construction. Though models with climate variables tend to outperform less data-intensive constructs, managers may favor a periodicity-based model as it only requires the energy consumption data itself. Ultimately, both approaches are viable and can produce skillful models.

3.8 Conclusions

The prospect of a variable climate places accurate impact predictions at a premium. To make data-driven management decisions, energy managers require consumption prediction models that can anticipate peaks and lows in energy consumption skillfully and simply. The methodology herein provides a flexible framework that can be adapted to any number of continuous or categorical independent variables, utilizes open-source data, and extensively accounts for modeling bias. As a result, skillful campus-level energy consumption prediction models were generated. Each performs skillfully in those areas most important for decision-making. By way of contingency tables, it was identified that using 25th and 75th percentile predictions can additionally bolster the accuracy of peak and low energy consumption predictions. These results validate that, rather than paying for hourly/daily modeling capabilities, it is possible for energy managers to fairly predict energy consumption with high skill using open-source climate information.

This research is limited in that the models were calibrated to the singular location of Dayton, OH. Future research must be conducted to evaluate the skill of such models across a span of varying climate regions to validate its adaptability. This is particularly important because the aspects of climate that impact energy use are likely to vary. Therefore, exhaustive inclusion of climate input types should be favored in initial model development in order to identify which are most impactful in the PCA. However, a benefit of the modeling framework developed in this work is that it is exportable. So long as the modeler or manager possesses some amount of energy use and climate data, the role climate and categorical variables play in explaining energy use can be determined.

Another limitation is that the selection of split points in statistical correction was calibrated manually based on the fit of the modeled CDF on the observed data CDF. Based on incremental changes to the split points, an effective amount of statistical correction was still achieved through manual calibration. Be it that explained variance only changed by approximately +/- 0.01 with changes to the split point in statistical correction, applying optimization techniques could be applied to better fit CDFs while maximizing explained variance.

Future research should focus on additional complexity reduction of the models to improve uptake potential. Here, there was no precise methodology used to select inputs variables for the lean models. Although, the lean models are an approximation of the least complex input compilation, following the approach herein may not be conducive to energy managers. However, it is likely that lean models will vary by location due to the aspects of climate that influence energy use.

Finally, a large and potentially conservative number of PCs for the models with larger input sets were retained using Jolliffe's rule. Adopting other rules (e.g. Kaiser's Rule) could further narrow the retained PC count of larger input sets.

To decide which model is best suited for energy managers, the simplicity and accuracy tradeoff must be analyzed. The climate variable temperature is shown to be the most influential input data in this research; however, it does require collection and processing time. Depending on how energy managers value time and computational capacity, utilizing the periodicity and time inputs only may be a better option.

This study focuses on the calibration and validation of a predictive model. To assess how use patterns and magnitude could change with climate and the degree to which historically relevant variables remain useful, future research should run this model framework in a forecast mode with future predictions of climate variables. Furthermore, applying cost factors to such a forecast could allow managers to budget for future energy use rather than react to changes.

Ultimately, this exploration of campus-level energy consumption prediction modeling is one of the first of its kind. Campus to city level decisionmakers require energy managers to produce accurate expectations of energy use to plan and budget for their daily, and even hourly, operations. These findings demonstrate that proactive energy planning through energy consumption predictions at the campus-level can be accomplished through accessible, fair, and skillful means.

IV. Climate-Informed Energy Consumption Projections for Campus Policy Development and Budget Decisions

4.1 Abstract

Climate variability creates energy demand uncertainty and complicates long-term asset management and budget planning. Without understanding future energy demand trends related to climate intensification, changes to energy consumption could result in budget escalation. Energy demand trends can inform campus infrastructure repair and modernization plans, effective energy use reduction policies, or renewable energy resource implementation decisions, all of which aim to mitigate energy cost escalation and variability. To make these long-term management decisions, energy managers require unbiased and accurate energy use forecasts. This research uses a statistical, model-based forecast framework, calibrated retrospectively with open-source climate data, and run in a forecast mode with CMIP5 projections of temperature for RCPs 4.5 and 8.5 to predict total daily energy consumption and costs for a campus-sized community (population: 30,000) through the end of the century. The model suggests that median annual campus energy consumption, based on temperature rise alone, could increase by 4.8% with RCP4.5 and 19.3% with RCP8.5 by the end of the century, and create budget deficits of $2.5M and $7.9M, respectively. A probabilistic analysis suggests that end-of-century projections are relatively certain and will span $7.3 to 7.9M for RCP8.5 temperatures at the interquartile range. Monthly forecasts indicate that summer month energy consumption could significantly increase within the first decade (2020-2030), and nearly all months will experience significant increases by the end of the century. The

cumulative results reveal annual cost increases monetarily equivalent to building new facilities. Overall, the forecast model framework simply and efficiently provides campus-scale projections that enable energy managers to understand how and when interventions should occur and to justify intervention timing and financial decisions.

4.2 Introduction

Climate variability is an exogenous, stochastic factor that causes energy demand uncertainty and complicates energy consumption modeling. Limited understanding of future energy demand trends can leave energy managers ill-prepared to make long-term decisions, such as advocating for infrastructure modernization or expansion, making facilities more energy efficient, or considering renewable energy resources. Energy managers require forecast models to project future energy demand and inform these decisions and their budgets. Managing energy at the campus level is particularly difficult because resource allocation must be prioritized amongst many facilities [45]. With access to future campus energy demand trends, energy managers are better positioned for long-term decision-making to mitigate the impact of demand and cost fluctuations for campus-wide operations.

Many international organizations and institutions have allocated time and resources to understanding the future impacts of greenhouse gases (GHG) on climate extremes and gradual trends. Future risks to businesses, governments, and communities due to these changes are then explored through sectoral models [31], [35], [46], [48]. Gradual changes to climate and its forecasted impact on energy consumption is one key area of research [7], [9], [17], [28], [29], [38]. For example, Zhou et al. [14] forecast the impact of changing climate across the 21st century for United States heating and cooling, finding that heating energy demand will gradually drop for all states, but cooling energy demand will ultimately increase after decreasing slightly from years 2005 to 2020. Emodi et al. [13] forecast electrical energy consumption in Australia using climate projections and find that electrical energy consumption may decrease in the first few decades before increasing and surpassing current energy consumption levels by the end of the 21st century.

Climate projections are necessary to forecast future energy consumption. Globally generated climate projection models, known as general circulation models (GCMs), have been developed that span the 21st century. These projections are primarily based on GHG representative concentration pathways (RCPs), which project several global GHG emissions scenarios [15]. The World Climate Research Programme's (WCRP) Coupled Model Intercomparison Project (CMIP) is an effort that has consolidated the output of over 50 GCMs with variable resolution ranges to generate universally accepted projections [60]. CMIP5 climate projections are used in this research.

Energy prediction and forecast analyses occur anywhere between facility to multinational scales [7], [9], [17]–[19], [29], [38]. For example, De Rosa et al. [18] addresses how energy savings policy can impact energy consumption at the facility level, while Amato et al. [9] provide a regional-level analysis that focuses on how changes to energy infrastructure impact energy consumption spatially. Organizations, independent of size, must pay the recurring and variable cost of energy to operate. Furthermore, the pivotal nature of organizational energy is exhibited through the costs to market sectors. Residential, commercial, and industrial sectors spent $224B, $201B, and $359B on energy in 2017, respectively [4]. Even more broadly, U.S. energy users paid $1.1T, or 5.8% of gross domestic product, for energy in 2017.

At a facility-level, specific building types have been modeled and exported across various climate zones to understand the temporal and spatial impacts of a changing climate on facility types [18], [32]. However, in cases where facilities cannot be individually managed, organization or campus-level estimates are valuable. Few studies exist at an organization or campus-level, though aggregation or disaggregation methods are most commonly used to achieve results at this level of analysis. Dirks et al. [12] acquire facility energy modeling software that houses thousands of facility types, and energy information on an entire geographical region of the U.S. to model energy consumption and inform energy mitigation and savings techniques. The difficulty of campus-level analyses lies in capturing the different use regimes between facilities and accessing large quantities of facility-level data. Accessibility is an issue because municipality-sized communities may not have a singular method of procuring energy, and may not meter all facilities; therefore, multiple energy providers may serve the same city or town, making the collection of the data burdensome [33]. As a result, when assumptions are made, the reliability and exportability of the analyses are limited. The facilities-level studies provide useful results but are far too granular in their focus to provide value for energy managers of large campuses. At the same time, as depicted by [14] and [13], state or national trends are not actionable for campuses. The lack of reliable studies at the organizational level, coupled with future demand uncertainty and the magnitude of energy costs, clearly warrant investigation.

Despite the significant contributions of the aforementioned research, there is limited reported research that develops and analyzes energy consumption forecasts at the campus level. Accordingly, the objectives of this research aim to fill this knowledge gap in climate-informed energy management. First, using the results and model framework of Chapter 3, the most influential climate and time inputs are combined to create and calibrate a retrospective total energy consumption prediction model. Then, using CMIP5 climate projection data, total energy consumption and costs are projected through the end of the century. This research is an example of how statistical modeling tools can aid energy managers in long-term operations and energy infrastructure planning.

4.3 Data and Case Study

The U.S. Department of Defense (DoD) owns 800 military installations that operate similarly to business, medical, and university campuses, and due to data availability, provide a unique opportunity to apply statistical modeling and climate projections at such a scale. As detailed in Chapter 3, the U.S. Congress is interested in understanding climate change's impact on installation resilience and its operations. The DoD is the U.S. Government's largest operational and facilities energy user [3]. As such, the DoD must consider how projected changes to climate may impact energy demand and use patterns. The 2019 U.S. Government Accountability Office report gauged the application of climate considerations on military installation infrastructure planning and operations. It highlighted the limited use of climate projections by installations to understand future climate-driven impacts [1], [2]. Accurate projections of future energy costs are essential for informing short- and long-term decisions and budgets. For instance, in the short-term, accurate year-ahead forecasts can prevent underbudgeting, which drives the need to borrow from other facility sustainment funds. In the long-term, they inform long-range organizational operating budgets, aid the development of use and energy management policies, and justify incorporating technologies to mitigate potential energy cost variability to achieve energy savings. This research aims to apply climate projections to a military installation to inform these decision types.

Few studies have focused on creating forecast models and analyses for a municipality or campus using historical energy data. This shortage of studies is largely due to limited data availability and resolution, which hamper statistical significance. Building Chapter 3, analyses are performed for Wright-Patterson Air Force Base (WPAFB), located near Dayton, OH. WPAFB employs over 30,000 people and includes industrial, commercial, community support, and residential operation types (approximately 26,500 facilities). Dayton, Ohio, is a temperate climate that experiences warm to hot summers, cool to cold winters, and moderate rainfall throughout the year. Chapter 3 finds that temperature is the dominant climate variable for total, campus-level energy consumption predictions at WPAFB, which is consistent with other studies at temperate locations. See Section 3.1 for more details on selected input types.

WPAFB provided total energy consumption data across four consecutive years at the half-hourly scale, given in kilowatt-hours (1 Oct 2015 - 30 Sep 2019). The data's scale most closely resembles that of a city, manufacturing complex, or medical or university campus. Because the finest resolution of the climate data is at the daily scale, the energy consumption data used to calibrate the prediction model was aggregated to the daily scale (average hourly energy consumption per day). WPAFB provided energy invoices to gauge the unit cost of energy consumption for the installation. While the authors are not permitted to supply billed consumption, or rates, they may be requested via FOIA from the installation. Through personal communications with the utility provider, 5 cents per kWh is a reasonable unit cost for high energy consumers. Most often, negotiated rates for large consumers fall below household rates.

CMIP5 projections of temperature at the highest possible resolution are used as the means to forecast energy consumption. Open-source climate projections were obtained through the Lawrence Livermore National Laboratory website. All available models and ensembles for the CMIP5 bias-corrected daily climate projections (BCCAv2-CMIP5-Climate-daily) for maximum and minimum temperature were selected for both RCP 4.5 and RCP8.5. The two RCPs are used to demonstrate two potential ranges of future energy consumption values. The projection set for RCP4.5 consists of 19 models, with 42 projection ensembles, and RCP8.5 consists of 20 models, with 41 projection ensembles. The median value for all ensembles for each temperature variable and RCP was used to consolidate the ensembles into a single set of deterministic predictions. This approach is consistent with many studies where capturing the general responses to climate change is desired. WPAFB provided historical daily maximum and minimum temperatures to calibrate the statistical prediction model, matching the projection data's scale obtained through CMIP5. Because temperature is the most variable data type used to calibrate the forecast model, knowing past trends in temperature could help understand how energy consumption may change in the future.

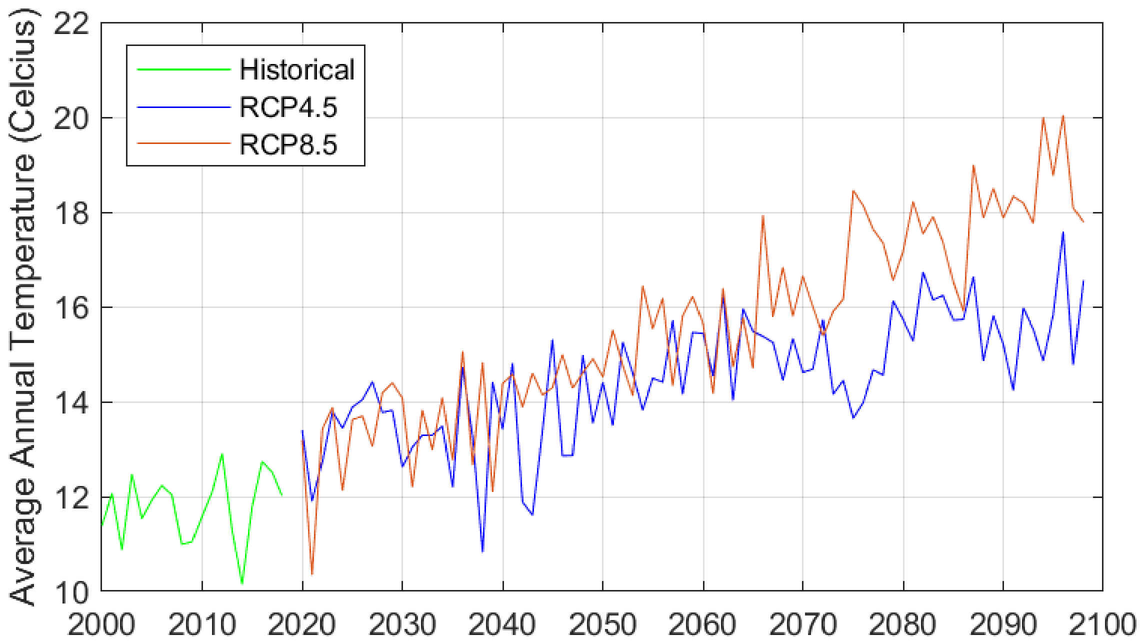

Figure 8. Time series of the historical and forecasted average annual temperatures: RCP4.5 (blue) and RC8.5 (orange). Generally, temperature will gradually increase throughout the century. There is a period of less substantial temperature increase between 2010 and 2040, where a few years may experience average annual temperatures similar to those experienced at the beginning of the century.

Historically, the period of 2016 to 2019 contains four of the top five warmest years on record. Due to the cyclical nature of yearly temperature, it appears that through 2025, average annual temperatures could drop close to the lowest average annual temperatures experienced this century. From 2020 to 2040, there appears to be a period where little temperature increase occurs. These periods of low or stable temperatures could result in low or stabilized energy consumption, which is consistent with the research of [14] and [13]. However, after 2040, substantial temperature increases are projected (Fig. 8).

4.4 Methodology

Statistical forecasting techniques are used in many fields to understand the future impacts of climate change. This research uses temperature and categorical time variables, identified as the most influential variable types in Chapter 3, to create a cross-validated multiple linear regression model framework to apply RCP 4.5 and 8.5 climate projections. With energy consumption projections established, the unit cost is applied to determine the change in campus energy consumption cost due to changing climate.

4.4.1 Forecast Model Calibration and Validation

Specific input types were selected based on the results of Chapter 3 and the CMIP5 projections available. Chapter 3 identifies the model that was least complex while retaining the most skill, and principal components were correlated to the initial inputs of the model to determine those inputs that were most influential in the model. Creating lean models, reveals the variations of temperature, primarily dew point temperature, dominate the first six principal components. However, categorical time variables day-of-the-week and month are also influential inputs as they were also selected to create the lean model. As such, these two input types (temperature and categorical time variables) are adopted as the primary inputs that create the prediction model.

A cross-validation step is added to a standard principal component multiple linear regression to develop a bias-limited prediction model that is informed with temperature and categorical time inputs (i.e., dummy variables). A leave-one-out cross-validated hindcast is produced for the entire historical dataset to obtain a set of deterministic predictions of expected energy consumption for WPAFB. The cross-validation process produces a less biased and more conservative set of predictions because it removes the time-step being predicted but eliminates the need for a standard calibration and validation set. The tradeoff is that the percentage of variance explained by the model is generally less than it would be if cross-validation were omitted. The model developed is conservative as it assumes that the community's use patterns are unchanging, i.e., the base does not grow or change with time.

Explained variance (Pearson's correlation coefficient) and mean absolute percentage error (MAPE) are generated to validate the prediction model's skill. MAPE is commonly used in energy prediction research, and established thresholds exist to communicate the skill of prediction models [5], [56]. A MAPE score below 20 designates a prediction model of "good" quality, and a MAPE score below 10 designates a model of "excellent" quality [57].

4.4.2 Forward-Looking Forecasts of Energy

Forecasts of daily energy consumption are generated by applying CMIP5 forecasted maximum and minimum temperature inputs and categorical time inputs to the coefficients generated by the cross-validated energy consumption prediction model. This research uses a standard dummy variable assignment methodology, where the number of categorical variables is one less than the total actual number of variables. For example, the number of variables for weekdays is six (6), which is one less than the number of days in a week. The results are then aggregated to total yearly energy consumption to illustrate the century-scale energy consumption trend and differences between predictions informed with RCP4.5 and RCP8.5 temperatures. The annual total consumption aggregation is computed by multiplying the resulting output from the forecast model, which is average hourly energy consumption per day, by 24 hours per day. This computation provides an estimate of total daily energy consumption. Next, each day's total energy consumption was added within the same year to obtain total annual energy consumption.

The energy forecasts are aggregated to total monthly energy consumption and placed in decadal categories (2020-2030, 2030-2040, etc.) to present a range of possible yearly energy consumption values for each year within a decade. Trends in monthly energy consumption are compared in this way across the century. One-way ANOVA tests are used to determine the significance of monthly energy consumption changes between the first decade and each subsequent decade to highlight when, during the century, monthly and seasonal trends diverge from current use behaviors.

4.4.3 Forecasted Energy Costs

A cost factor is applied by multiplying energy consumption by the per consumption unit cost of 5 cents per-kWh, discussed in Section 2.2. To understand the potential change in costs for WPAFB, annual change is determined by calculating the cost difference between a base year (2020), and each subsequent year (2021-2100). Values are reported as constant 2020 dollars. The result is the change in energy consumption costs due to changing climate, and more precisely, temperature. When using decadal categories, change in cost is determined by the difference in cost of energy consumption between decade 2020-30 and each subsequent decade. To probabilistically understand the potential change in energy consumption costs across the century, the 75th, 50th, and 25th percentile energy consumption values for mid- (2050-2060) and end-of-century (2090-2100) decades are compared to those of the first decade predicted (2020-2030). The difference in the range of energy consumption values shows how certainty in forecasted energy values changes over the century, and provides a method to develop a probabilistic range of possible cost outcomes that could be used to inform facilities and energy budget planning.

4.5 Results

The results are organized such that they reflect the order of the methodology. First, the quality of the predictive model is evaluated through its ability to explain variance in historical energy consumption and through coefficient of determination and MAPE. Next, through the application of the predictive model in a forecast mode, projected total energy consumption is analyzed for the remainder of the century. Lastly, the energy unit costs are applied to forecasted yearly and monthly energy consumption to determine the future change energy consumption cost due to climate change-induced temperature rise.

4.5.1 Prediction Model Calibration and Validation

The prediction model produced a “moderate” explained variance of (r^2 = 0.46), though it only contained the climate factor temperature: the most impactful climate factor by a large margin. The highest performing model produced in Chapter 3 resulted in an explained variance of approximately 0.70 but used more inputs of both climate and categorical time type. The model produces a MAPE of 6.95, designating the prediction model as being of “excellent” quality.

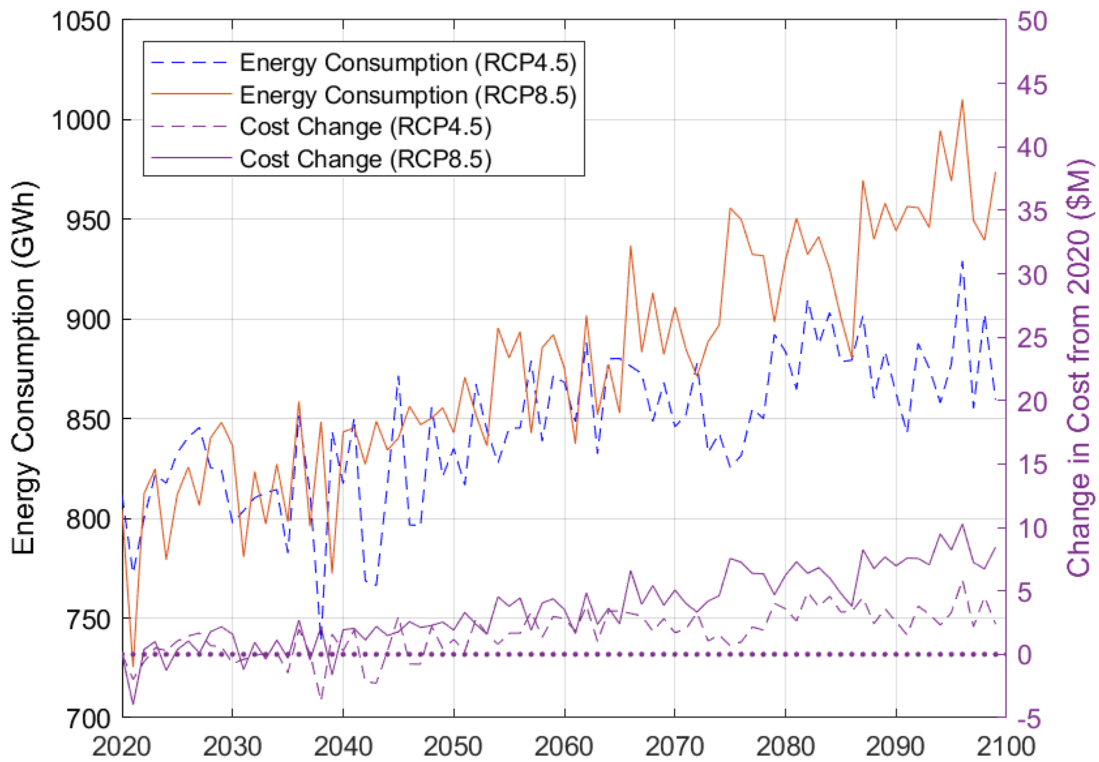

Figure 9. Time series plot of forecasted total annual energy consumption for RCP4.5 (dashed blue) and RCP8.5 (orange), and the change in cost of total energy consumption from the year 2020 for RCP4.5 (dashed purple) and RCP8.5 (purple).

4.5.2 Forward-Looking Forecasts of Energy

Applying the model in a forecast mode reveals that energy consumption will increase by the end of the century for both RCP cases. However, RCP8.5 energy consumption increases more aggressively beginning around 2065 (Fig. 9). Between 2020 and 2040, there is no substantial increase in energy consumption, nor is there a significant difference in consumption predictions for the RCPs. This result could be attributed to the recently observed decreases in temperature since 2017 and milder maximum and minimum temperatures projected for the first two decades following the year 2020. Applying a linear fit to both RCP forecasts shows that energy consumption could increase by 0.80 GWh per year for RCP4.5 and 2.14 GWh per year for RCP8.5. The resulting cost increases are $40K per year and $107K per year, respectively. Though a basic result—energy use increases with temperature rise—is expected, the onset and magnitude of consumption and cost are not.

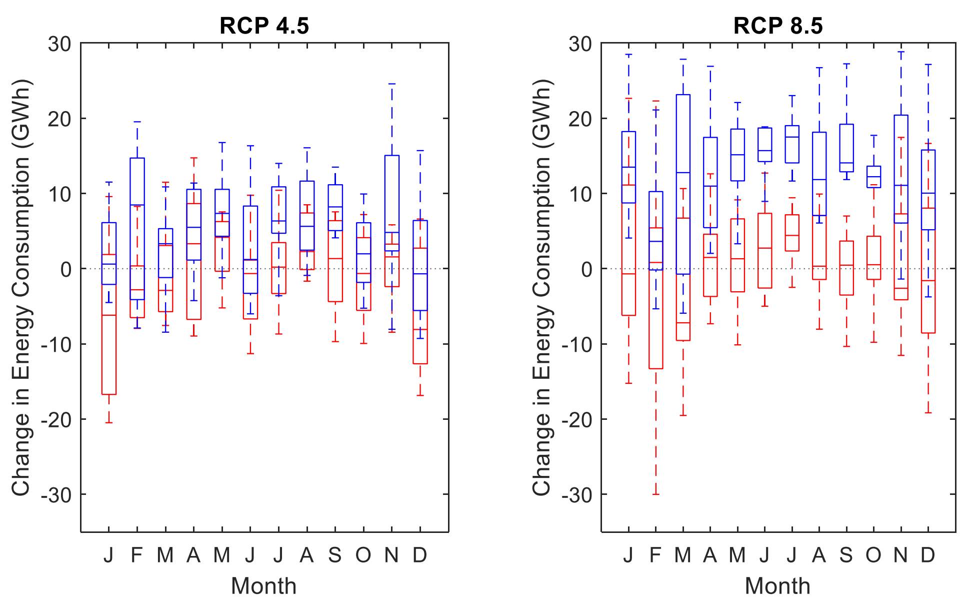

Figure 10. Boxplots of the difference in monthly total energy consumption between decades 2020-30 and 2030-40 (red) and 2020-30 and 2090-2100 (blue) for RCP4.5 and RCP8.5.

Table 4. Significance of the difference in forecasted monthly energy consumption for a) RCP4.5 and b) RCP8.5 between the decade 2020-2030 and subsequent decades

$p < 0.05$ $0.05 < p < 0.10$ $0.10 < p < 0.25$ $p > 0.25$

\+ positive change – negative change

a.

	Significance Compared to 2020-2030 Decade (*ANOVA test p-value*)						
	2030-2040	**2040-2050**	**2050-2060**	**2060-2070**	**2070-2080**	**2080-2090**	**2090-2100**
Jan	–	–	–				
Feb	–	–		+		+	+
Mar		–			+		+
Apr				+	+	+	+
May		+	+	+	+	+	+
Jun				+	+	+	+
Jul		+	+	+	+	+	+
Aug	+	+	+	+	+	+	+
Sep		+	+	+	+	+	+
Oct				+		+	
Nov				+		+	+
Dec	–	–					

b.

	Significance Compared to 2020-2030 Decade (*ANOVA test p-value*)						
	2030-2040	**2040-2050**	**2050-2060**	**2060-2070**	**2070-2080**	**2080-2090**	**2090-2100**
Jan		+	+	+	+	+	+
Feb					+		+
Mar	–		+		+	+	+
Apr		+	+	+	+	+	+
May		+	+	+	+	+	+
Jun	+	+	+	+	+	+	+
Jul	+	+	+	+	+	+	+
Aug		+	+	+	+	+	+
Sep			+	+	+	+	+
Oct		+	+	+	+	+	+
Nov			+	+	+	+	+
Dec			+		+	+	+

Both RCP scenarios indicate that from decades 2020-30 to 2030-40 there will likely be no change or a drop in total energy, primarily within the boreal winter months (Fig. 10). By the end of the century, most months will likely surpass their 2020-30 energy consumption levels. Spring, summer, and fall months achieve greater energy consumption under RCP8.5, with higher degrees of significance, much sooner than RCP4.5 (Table 4). Again, while this general result is expected, the onset of significantly elevated energy consumption values was unknown until this point. Also, RCP8.5-informed forecasts produce significant increases in winter energy consumption as early as the decade 2040-2050, while RCP4.5 results show decreases in winter energy consumption in this same period. Additionally, RCP8.5 monthly energy consumption exhibits a higher degree of inter-annual variability than RCP4.5, which could mean more uncertainty in forecasted results or less stability in annual consumption (Fig. 10). Less stability in consumption would make the task of budgeting on the part of energy managers difficult. Overall, there is high confidence that summer and adjacent seasons' energy consumption will increase earlier in the century, while the timeframe for increases in winter energy consumption may vary.

4.5.3 Forecasted Energy Cost

Because the energy billing structure WPAFB faces is likely temporally insensitive at the seasonal scale, i.e., using 5 cents per-kWh as a cost factor, the century-long predictions of energy cost change follow predicted consumption patterns. Though this basic result is perhaps uninteresting based on the billing structure, the magnitude of the project cost increases reported in this section emphasize how planning and budgeting will become increasingly important as temperature rises.

There is little change in cost between 2020 and 2040, reflecting the consistency in the energy consumption predictions. Notably, the potential for energy consumption decreases in the winter months during this period. However, annual energy consumption cost is permanently positive (increase over historical conditions) by 2040 for RCP8.5 and 2050 for RCP4.5 (Fig. 9, purple curves).

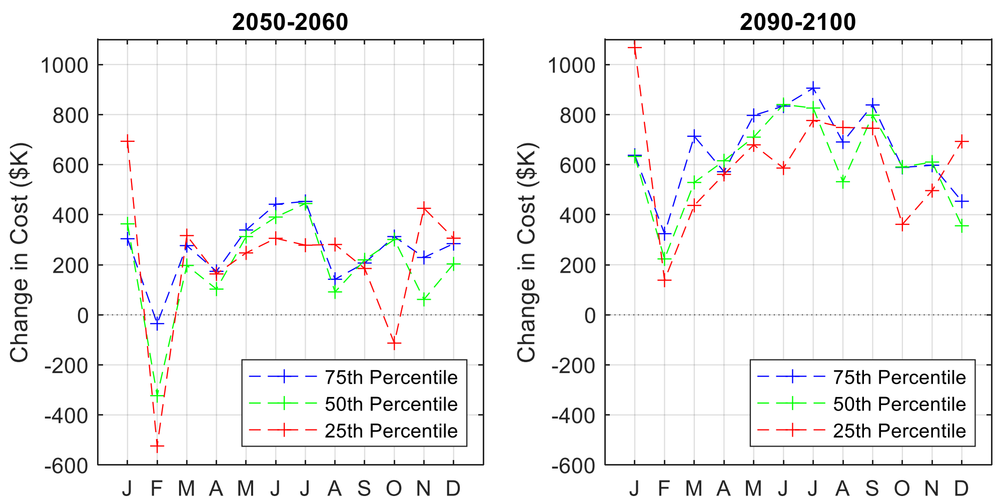

Figure 11. The monthly difference in 75th percentile (blue), 50th percentile (green), and 25th percentile (red) changes in cost from decade 2020-30 to decades 2050-60 and 2090-2100 for RCP8.5.

While deterministic predictions illustrate the general trend and magnitude of changes in consumption and costs tied to temperature, it is clear that interannual variability in climate, and therefore energy consumption, must be evaluated. To communicate the difference in the range of possible changes in cost by the middle- and end-of-century, the interquartile range (IQR) percentiles are computed as a monthly value (Fig. 11). The trend is positive for most months and almost all values in the IQR. In some cases, during winter months, the 25th percentile forecasted cost change is significantly greater than the change at the median and 75th, meaning that costs in these seasons are likely to be most similar. In other words, the cost at the 25th percentile closes the gap with the median and 75th percentile predictions. When they change together, like in September 2050-2060 or April 2090-2100, there is no substantial change in the range of values.

Overall, for RCP8.5, while many of the months will have increased costs by the middle of the century, months traditionally associated with low energy consumption may experience decreases in costs. For example, February could experience up to a $500K drop in energy costs, likely due to milder mid-winter temperatures. By the end of the century, summer monthly costs will have increased by $500-900K, and winter monthly costs will have increased by $150-650K. By summing percentile groups across months for the decade 2090-2100, it is found that the cost range of yearly energy consumption will increase between $7.3-7.9M.

These results are consistent with similar works in this field of study. The energy consumption forecasts developed in this research show a constant, and even a slight decline, in energy consumption approaching the middle of the century. The work of [14] and [13] capture this phenomenon. For example, Zhou et al. [14] find that a slight decline in heating demand and a slight decline in cooling demand occur in the first half of the 21st century for the state of Ohio. Since the primary facility energy drain related to climate is heating and cooling, [14] appear to explain a large part of what is observed with total energy consumption in this research. In contrast to existing studies, the forecasts developed in this research aid in analyzing century-long energy consumption trends at the campus level using campus-level energy data. Additionally, the inputs and predictive model development used to forecast energy consumption were informed by the bias-accounting framework developed in Chapter 3.

4.6 Discussion

This research suggests that energy managers and campus leaders must be prepared for energy consumption and costs to increase over the century. Long-term consumption and cost forecasts, consistent with the type produced here, provide valuable information with which 1) the capacity of currently owned assets or infrastructure to support future operations can be evaluated, 2) the value of proposed renewable energy projects can be determined, and 3) future budget predictions can be generated. Overall, energy use increases with temperature increase. Though a basic result, the magnitude of consumption and cost increases is likely useful for energy managers who are expected to produce accurate, year-ahead budgets. For WPAFB, median expectations suggest that costs could increase anywhere from $2.5-7.9M by the end of the century ($40-107K annually). In other words, by the end of the century, the campus will need to budget for the monetary equivalent of the cost of a new facility each year. Annually, campus energy costs are increasing by the monetary equivalent of a facility system repair-sized project. Decision-makers must determine how and when to mitigate these costs with blends of policies and infrastructure adaptations. Possible mitigation strategies could include enacting more effective operations policies that reduce energy consumption across a campus, like non-conditioned seasons [59]; using prediction modeling to inform proactive budgeting and prioritization of energy projects; implementing new energy-efficient technologies; or integrating renewable energy resources. Determining when strategies must be implemented can also influence which action is taken. This research's statistical forecast model is an effective tool for determining when strategies should be enacted.

For instance, the deterministic analysis reveals that a consistent positive change in costs will not occur for WPAFB until approximately 2040 for RCP8.5 and 2050 for RCP4.5. These results suggest that energy managers have time to plan for or wait to make future decisions on how to mitigate energy consumption and cost increases. Even those strategies with the most time-intensive planning and funding processes (e.g., construction projects and integration of renewable resources) could wait 10 to 15 years to be enacted. With the possibility of a $2.5-7.9M increase in yearly energy consumption, the implementation of renewable energy resources, like geothermal or solar energy, could become a feasible option. The high initial investment involved with most renewable energy alternatives, because of large infrastructure additions, makes these options economically infeasible to energy users. Investment in renewable energy resources may become more justifiable knowing that $2.5-7.9M in cost increases could be mitigated through its less costly addition.

Sub-annual analyses can inform decision-makers of what energy consumption types to target for mitigation strategies. From 2020 to 2040, energy consumption and costs may remain consistent or drop during the winter months. However, summer months, such as July and August, may experience significant increases starting within the first decade. As such, decision-makers may initially target strategies that would mitigate summer energy loads. For example, higher facility temperature set points and load shedding could decrease facility cooling load.

Considering the fixed nature of organizational budgets, and particularly those of the public sector, energy managers require tools, like this forecast model, to thoroughly justify project cost. These results suggest that the savings realized by larger investments, such as infrastructure modernization or renewable energy resources, could surpass the cost of the investment, given the forecasted energy cost increases. This justification is exacerbated by the fact that the estimates in this research are conservative.

Mission and energy market changes were not factored into this research so that the focus would be the impact of climate change, specifically temperature, on energy consumption and cost. By not accounting for the future growth of WPAFB, or energy cost forecasts, the analyses here are likely conservative: especially considering that mission sets and the installation's daily population will likely increase, campuses are more likely to expand capabilities, and energy prices increase as resources become scarcer. Specifically, for military installations, new mission beddowns are expected, which will increase daily energy use as people are brought in to staff new facilities. This expansion at WPAFB is synonymous with growth in any campus environment. However, development changes, on the century-scale, are difficult to predict. Because growth and cost forecasts were not the focus of this effort, these complications were not added. This allows the focus to remain squarely on the increase in energy use and cost growth due to climate.

This research is also limited in that the specific temperature variables identified as most influential in Chapter 3 were not available in the used climate projections. Other, less influential climate factors are also not included in this research. Though temperature type variables were used in this study, the hindcast skill is expected to be substantially less than when using the exact temperature inputs and other climate factors. The model's framework also assumes that the impact of inputs, such as temperature, on energy consumption does not change over the century. Studies, like [31], that extrapolate temporal changes in climate factor impacts, must be conducted to understand the impacts of climate change holistically across the century.

Regarding CMIP5 projections, the median of all available temperature ensembles was used to create the forecasts. Future research should test each ensemble as the temperature input for this forecast model to create a range of temperature-informed energy consumption predictions, thus better communicating model uncertainty. Finally, energy data was aggregated from an hourly scale to yearly and monthly scales, which introduces more uncertainty into the final aggregated energy values.

Future research should incorporate less influential climate factors, ideally from CMIP6 and other projections, and possible mission and energy market changes as inputs to determine their impact on the model and forecasts. The prediction and forecast methodologies must also be applied across climate regions to test the framework's exportability and adaptability. This research and that of Chapter 3 highlight the distinct connection between temperature and energy consumption. This research is conducted in a moderate climate region. If a $7M increase in annual energy consumption costs is possible in a moderate climate region, it would stand to reason that area that already experience high temperatures could see more drastic increases in energy consumption with the forecasted increases in temperature.

On the other hand, for colder climate regions, increased temperatures could reduce energy consumption because of a reduced need for heating [59]. As Chapter 3 emphasized, these analyses must be applied across climate regions due to the spatial variation of impacts that climate change, and specifically global temperature rise, can create. As such, not all energy managers should have the same long-term energy asset management plans but customize them to fit their specific goals and environmental, monetary, and resource availability situations.

4.7 Conclusions

The prospect of a variable climate places accessibility of forecasts at a premium. To make long-term, data-driven management decisions, energy managers require consumption forecasts to inform how and when they should take significant mitigation efforts to alleviate energy budget increases. This research uses a two-phase statistical modeling approach to predict future energy consumption for WPAFB (Dayton, Ohio) through the end of the century. The first phase uses a cross-validated, principal component multiple linear regression, informed with historical observations of temperature and time dummy variables to produce parameter estimates. The second phase uses the parameter estimates in a forecast mode, using end-of-century temperature predictions from CMIP5 climate projections to forecast campus-level energy consumption.

Results show that WPAFB could face meaningful increases in energy consumption: 19.3% for RCP8.5 and 4.8% for RCP4.5 by the end of the century for median predictions. The associated annual cost increases, for one military installation, are monetarily equivalent to building a new facility. The forecasts also reveal that significant changes in monthly energy consumption may occur within the next decade. Viewed at the DoD level, this cost drain per installation is substantial. The information provided by these forecasts enable campus energy managers to understand the magnitude and timeframe where energy consumption and costs could significantly escalate and justify how and when interventions are necessary.

However, the forecast model framework also provides an accessible and spatiotemporally flexible pathway for energy managers to evaluate energy consumption. Thus, as climate forecasts become more skillful and climate data becomes more available, the model can be easily re-generated to understand how the influence of variables and future energy trends change.

Ultimately, this exploration of campus-level forecasts of energy consumption is one of the first of its kind. Campus to city-level decision-makers require methods to determine future energy consumption trends and justify and advocate for adaptation investments. Such information can prepare energy managers to anticipate location-specific changes to energy consumption and adapt long-term asset management plans.

V. Conclusions and Recommendations

5.1 Conclusions of Research

Aligning with the first objective of this research, the literature review (Ch. 2) revealed the components of an empirical energy consumption prediction modeling, including 1) the selection of a spatial and temporal scale for predictions, 2) collection of historical energy consumption and climate data, and 3) selection of a regression technique. By researching each of these areas, the specific niche of this research was determined. First, it was found that few studies perform energy modeling analyses at an organization- or campus-level, using campus scale energy consumption data. Also, a variety of climate factors have been analyzed in energy consumption prediction modeling. Due to the spatial heterogeneity of climate impacts, it was determined that each location performing energy consumption modeling analyses should be exhaustive in the climate factors incorporated. Finally, though many regression techniques of varying complexity are used in energy prediction modeling, few utilize PCR. To achieve the second objective of this research, a campus-level energy prediction model developed using PCR and an exhaustive list of open-source climate factors is compiled.

To test the impact of climate factors on energy consumption and to determine whether skillful energy consumption can be achieved using climate inputs, a prediction model was generated in Chapter 3. All combinations of inputs (6) tested to develop the prediction model proved to be skillful, producing RPSSs greater than zero. However, a model consisting of climate and categorical ("dummy") time inputs proved to be the most skillful alongside a model consisting of all variable types (climate, periodicity, and time), with an explained variance 0.73, MAPE less than 10, and RPSS of 0.59. This result, the comparisons between models' performance, and correlations between the initial inputs and PCs reveal that the climate variables are the most impactful inputs, and especially variations of temperature (dew point, dry bulb, and wet bulb). Other notable inputs include time variables day-of-the-week and month, and the climate factor for solar irradiation. For further energy analyses at the specific location of WPAFB, these variables will be important to include. When the categorical performance was tested, it was found that the climate and time model performed best in the high and low energy categories. These energy categories are where most key decisions regarding energy occur or will occur. Additionally, when using 75th and 25th percentile predictions, performance in the low and high energy consumption categories, respectively, increase. The model is not as skillful in intermediate energy consumption categories; however, fewer policy decisions are necessary for ordinary energy use. Overall, though the most skillful model is not as equipped to predict intermediate levels of energy consumption, it does perform well across all key skill metrics (explained variance, MAPE, and RPSS) and in the categories where most important decisions are made regarding energy consumption. It was determined that the prediction model performs adequately meet the second research objective and to be applied accurately in a forecast mode.

Forecasts of energy consumption can inform energy managers of how and when short- and long-term interventions could be necessary, along with providing a tool to economically justify selected interventions. The generated forecasts revealed a consistency in deterministic annual energy consumption and costs from years 2020 to 2040 for RCP4.5 and RCP8.5, after which steady increases ensue through the remainder of the century. These results communicate that energy managers have time to wait to make decision or can begin planning for changes to energy consumption without immediately implementing interventions. At a monthly scale, statistically significant changes to both RCP4.5 and RCP8.5 summer month energy consumption could occur as early as the first decade after the year 2020, which usher energy managers toward targeting summer energy drains, like facility cooling. The probabilistic and deterministic results both indicate that annual energy consumption could change by as much as $7.3-7.9M, in 2020 dollars, by the end of the century. Though campus leaders have the option of bearing the financial burden, many could resort or be forced to implement mitigative strategies. For smaller forecasted increases, implementation of effective energy consumption policy could be most feasible, while larger increases could warrant infrastructural upgrades. With end of century energy cost increases on the monetary scale of new facility construction, integration of expensive infrastructure modernization or renewable energy resources could become viable as potential long-term savings outmatch initial investment. In alignment with research objective three, forecasts were successfully generated that gave insight into future budgeting decisions for energy managers, while providing an effective tool to justify intervention timing and financial decisions.

5.2 Significance of Research

In a universal context, this research is one of the first of its kind because of its focus on campus-level energy consumption analyses using campus-level, historical energy data. In the past, researchers have resorted to aggregating facility data or disaggregate regional or state data to perform analyses for multi-facility to city-sized communities. Access to military installation historical energy consumption data has been invaluable and established the accurate starting point necessary to develop meaningful energy predictions and prevent assumption-based limitations. Also, the combination of statistical bias correction and principal component regression is infrequently applied for energy consumption prediction research. This combination creates a prediction model framework that thoroughly accounts for statistical modelling bias and the multicollinearity inherent across model inputs. Additionally, with the ability to accurately predict and forecasts energy consumption at the campus level, the need to meter energy at facilities, and the associated costs, may become unnecessary for campus energy managers. This is particularly the case for installations or campuses where facility-level decision-making or control is not possible.

In the context of the United States' Air Force and military at large, this research acts as a response to Congress's call for understanding climate change impacts on DoD operations. Using both historical and projected data, a prediction model was developed that determines the impact of individual climate factors on WPAFB energy consumption, but also forecasts future changes to energy consumption due to climate change using the most impactful climate factors. It was established that all models generated by this framework produced more skillful predictions than using historically driven predictions, thus making statistical modeling a viable alternative to prior-year energy consumption information. For the temperate location of WPAFB, forecasts depict end-of-century cost increases that could warrant implementation of renewable energy resources, thus implying justification for renewable energy resources in less temperate environments. Furthermore, this work provided discussions on the applications of both the prediction model and forecasted information for energy consumption policy and budgeting decisions and justification.

5.3 Recommendations for Future Research

There are several areas in this research regarding input selection and compilation that should receive future research emphasis. These analyses focus on how climate factors and changing climate, via temperature, impact energy consumption. As such, not all explained variance is accounted for in the prediction model, and the forecasts are generally more conservative. Future research could explore how non-climate factors, such as energy market, mission, and population changes, impact energy consumption prediction and forecast model performance. Climate factors other than temperature should be included in future energy consumption forecast models and analyses to produce more accurate forecasts. Additionally, incorporating the potential changes in influence of climate factors across the century is not considered in this research and should be explored to more holistically understand the impacts of climate change on energy consumption. When compiling CMIP5 temperature projections, the median of all ensemble projections at each timestep were used to consolidate the ensembles. Future research should test each ensemble as the temperature input for this forecast model to create a range of temperature-informed predictions of energy consumption, thus better communicating model uncertainty.

Further experimentation is warranted for several of the technical aspects of the prediction model development process in this research. For instance, one principal component retention ruleset—Joliffe's Rule—was used to generate the final prediction model. Other rules, such as Kaiser's Rule, could be used to test the range of skill resulting from more conservative or liberal rulesets. Also, the split points in the bimodal statistical bias correction process were manually optimized to find the setting resulting in the tightest fit to the observed data cumulative distribution function. Work could be done to computerize the optimization of the statistical correction split point selection to maximize prediction model skill.

Future research is also required in the testing and application of models developed in this research. Several input combinations were tested in Chapter 3 to determine which produced the most skillful models. Each model performed better than climatology. However, for energy managers, deciding which model to use is more complicated than selecting which is most skillful. Depending on constraints to time, computational capacity, and expertise of energy managers, even the least skillful model may be the most viable. Further consideration should analyze the model complexity versus accuracy tradeoff and survey what energy managers find most valuable in prediction models.

The performance of energy consumption predictions and forecasts is promising. However, performance of the model framework must be analyzed across climate regions to determine its adaptability and exportability. As such, future research should apply the model framework to other campuses, or military installations, across different climate regions. Stemming from the literature review, few studies have analyzed energy consumption or generated energy prediction models for developing countries. However, as data quality and quantity improve, these countries will provide a unique opportunity to analyze prediction models and climate change impacts on energy consumption in a variety spatial and situational circumstances. Additionally, if conducted for allying countries, it could be beneficial for foreign affairs by fortifying national infrastructure and resilience for these countries.

Bibliography

[1] "DOD Needs to Assess Risk and Provide Guidance on Use of Climate Projections in Installation Master Plans and Facilities Designs," United States Government Accountability Office, Jun. 2019.

[2] C. Pasion, T. Wagner, C. Koschnick, S. Schuldt, J. Williams, and K. Hallinan, "Machine Learning Modeling of Horizontal Photovoltaics Using Weather and Location Data," *Energies*, vol. 13, no. 10, p. 2570, Jan. 2020, doi: 10.3390/en13102570.

[3] "Department of Defense Annual Energy Management and Resilience Report (AEMMR) Fiscal Year 2018," Office of the Assistant Secretary of Defense for Sustainment, Jun. 2019.

[4] "State Energy Data System (SEDS) 2017 Data," *U.S. Energy Information Administration: Independent Statistics & Analysis*, 2017. https://www.eia.gov/electricity/annual/html/epa_01_01.html (accessed Jan. 22, 2020).

[5] H. Son and C. Kim, "Short-term forecasting of electricity demand for the residential sector using weather and social variables," *Resources, Conservation and Recycling*, vol. 123, pp. 200–207, Aug. 2017, doi: 10.1016/j.resconrec.2016.01.016.

[6] S. N. Chandramowli and F. A. Felder, "Impact of climate change on electricity systems and markets – A review of models and forecasts," *Sustainable Energy Technologies and Assessments*, vol. 5, pp. 62–74, Mar. 2014, doi: 10.1016/j.seta.2013.11.003.

[7] 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, Apr. 2017, doi: 10.1016/j.energy.2017.04.034.

[8] "2012 Commercial Buildings Energy Consumption Survey (CBECS) Data," *Energy Information Administration (EIA): Independent Statistics & Analysis*, Mar. 2016. https://www.eia.gov/consumption/commercial/data/2012/ (accessed Feb. 18, 2020).

[9] A. D. Amato, M. Ruth, P. Kirshen, and J. Horwitz, "Regional Energy Demand Responses To Climate Change: Methodology And Application To The Commonwealth Of Massachusetts," *Climatic Change; Dordrecht*, vol. 71, no. 1–2, pp. 175–201, Jul. 2005, doi: http://dx.doi.org.afit.idm.oclc.org/10.1007/s10584-005-5931-2.

[10] F. Apadula, A. Bassini, A. Elli, and S. Scapin, "Relationships between meteorological variables and monthly electricity demand," *Applied Energy*, vol. 98, pp. 346–356, Oct. 2012, doi: 10.1016/j.apenergy.2012.03.053.

[11] M. De Felice, A. Alessandri, and F. Catalano, "Seasonal climate forecasts for medium-term electricity demand forecasting," *Applied Energy*, vol. 137, pp. 435–444, Jan. 2015, doi: 10.1016/j.apenergy.2014.10.030.

[12] J. A. Dirks *et al.*, "Impacts of climate change on energy consumption and peak demand in buildings: A detailed regional approach," *Energy*, vol. 79, pp. 20–32, Jan. 2015, doi: 10.1016/j.energy.2014.08.081.

[13] N. V. Emodi, T. Chaiechi, and A. R. A. Beg, “The impact of climate change on electricity demand in Australia:,” *Energy & Environment*, May 2018, doi: 10.1177/0958305X18776538.
[14] Y. Zhou *et al.*, “Modeling the effect of climate change on U.S. state-level buildings energy demands in an integrated assessment framework,” *Applied Energy*, vol. 113, pp. 1077–1088, Jan. 2014, doi: 10.1016/j.apenergy.2013.08.034.
[15] D. P. van Vuuren *et al.*, “The representative concentration pathways: an overview,” *Climatic Change*, vol. 109, no. 1, p. 5, Aug. 2011, doi: 10.1007/s10584-011-0148-z.
[16] “Global Change Assessment Model,” *Joint Global Change Research Institute*, 2020. http://www.globalchange.umd.edu/gcam/ (accessed Feb. 29, 2020).
[17] H. Al-bayaty, T. Mohammed, A. Ghareeb, and W. Wang, “City scale energy demand forecasting using machine learning based models: a comparative study,” in *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems*, Dubai, United Arab Emirates, Dec. 2019, pp. 1–9, doi: 10.1145/3368691.3368716.
[18] M. De Rosa, V. Bianco, F. Scarpa, and L. A. Tagliafico, “Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach,” *Applied Energy*, vol. 128, pp. 217–229, Sep. 2014, doi: 10.1016/j.apenergy.2014.04.067.
[19] 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, Feb. 2019, doi: 10.1016/j.apenergy.2018.10.119.
[20] M. Q. Raza and A. Khosravi, “A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings,” *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 1352–1372, Oct. 2015, doi: 10.1016/j.rser.2015.04.065.
[21] P. Dowling, “The impact of climate change on the European energy system,” *Energy Policy*, vol. 60, pp. 406–417, Sep. 2013, doi: 10.1016/j.enpol.2013.05.093.
[22] J. Dell *et al.*, “Ch. 4: Energy Supply and Use. Climate Change IMpacts in the United States: The Third National Climate Assessment,” U. S. Global Research Program, 3, 2014. Accessed: Mar. 06, 2020. [Online]. Available: https://nca2014.globalchange.gov/report/sectors/energy.
[23] U. S. Climate Change Science Program *et al.*, “Effects of Climate Change on Energy Production and Use in the United States,” Department of Energy: Office of Biological & Environmental Research, Washington, DC, USA, 2008. [Online]. Available: https://downloads.globalchange.gov/sap/sap4-5/sap4-5-final-all.pdf.
[24] T. Hong and M. Shahidehpour, “Eastern Interconnection States’ Planning Council (EISPC): Load Forecasting Case Study,” *U.S. Department of Energy*, p. 171, Jan. 2015.
[25] S. N. Fallah, M. Ganjkhani, S. Shamshirband, and K. Chau, “Computational Intelligence on Short-Term Load Forecasting: A Methodological Overview,” *Energies*, vol. 12, no. 3, Art. no. 3, Jan. 2019, doi: 10.3390/en12030393.
[26] J. Luo, T. Hong, and M. Yue, “Real-time anomaly detection for very short-term load forecasting,” *J. Mod. Power Syst. Clean Energy*, vol. 6, no. 2, pp. 235–243, Mar. 2018, doi: 10.1007/s40565-017-0351-7.

[27] M. Capuno, J.-S. Kim, and H. Song, “Very Short-Term Load Forecasting Using Hybrid Algebraic Prediction and Support Vector Regression,” *Mathematical Problems in Engineering*, vol. 2017, pp. 1–9, 2017, doi: 10.1155/2017/8298531.
[28] J. Moral-Carcedo and J. Pérez-García, “Time of day effects of temperature and daylight on short term electricity load,” *Energy*, vol. 174, pp. 169–183, May 2019, doi: 10.1016/j.energy.2019.02.158.
[29] J. Xie and T. Hong, “Wind Speed for Load Forecasting Models,” *Sustainability*, vol. 9, no. 5, p. 795, May 2017, doi: 10.3390/su9050795.
[30] S. K. Panda, S. N. Mohanty, and A. K. Jagadev, “Long Term Electrical Load Forecasting: An Empirical Study across Techniques and Domains,” *Indian Journal of Science and Technology*, vol. 10, no. 26, pp. 1–16, Jun. 2017, doi: 10.17485/ijst/2017/v10i26/115372.
[31] L. Wenz, A. Levermann, and M. Auffhammer, “North–south polarization of European electricity consumption under future warming,” *PNAS*, vol. 114, no. 38, pp. E7910–E7918, Sep. 2017, doi: 10.1073/pnas.1704339114.
[32] L. Wang, X. Liu, and H. Brown, “Prediction of the impacts of climate change on energy consumption for a medium-size office building with two climate models,” *Energy and Buildings*, vol. 157, pp. 218–226, Dec. 2017, doi: 10.1016/j.enbuild.2017.01.007.
[33] Eisenbach Consulting, LLC, “How Cities and City Governments Acquire Electricity in the United States,” *ElectricChoice*, 2017. https://secure.livechatinc.com/ (accessed Mar. 12, 2020).
[34] N. Ismail and S. Abdullah, “Principal component regression with artificial neural network to improve prediction of electricity demand,” *International Arab Journal of Information Technology*, vol. 13, no. 1A, pp. 196–202, 2016.
[35] J. Delorit, E. C. G. Ortuya, and P. Block, “Evaluation of model-based seasonal streamflow and water allocation forecasts for the Elqui Valley, Chile,” *Hydrology and Earth System Sciences; Katlenburg-Lindau*, vol. 21, no. 9, pp. 4711–4725, 2017, doi: http://dx.doi.org.afit.idm.oclc.org/10.5194/hess-21-4711-2017.
[36] E. T. Mansur, R. Mendelsohn, and W. Morrison, “Climate change adaptation: A study of fuel choice and consumption in the US energy sector,” *Journal of Environmental Economics and Management*, vol. 55, no. 2, pp. 175–193, Mar. 2008, doi: 10.1016/j.jeem.2007.10.001.
[37] J.-L. Fan, B.-J. Tang, H. Yu, Y.-B. Hou, and Y.-M. Wei, “Impact of climatic factors on monthly electricity consumption of China’s sectors,” *Nat Hazards*, vol. 75, no. 2, pp. 2027–2037, Jan. 2015, doi: 10.1007/s11069-014-1375-1.
[38] V. Gil, M. A. Gaertner, C. Gutierrez, and T. Losada, “Impact of climate change on solar irradiation and variability over the Iberian Peninsula using regional climate models,” *International Journal of Climatology*, vol. 39, no. 3, pp. 1733–1747, 2019, doi: 10.1002/joc.5916.
[39] D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, “A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables,” *Applied Energy*, vol. 140, pp. 385–394, Feb. 2015, doi: 10.1016/j.apenergy.2014.12.011.

[40] P. Xu, Y. J. Huang, N. Miller, N. Schlegel, and P. Shen, “Impacts of climate change on building heating and cooling energy patterns in California,” *Energy*, vol. 44, no. 1, pp. 792–804, Aug. 2012, doi: 10.1016/j.energy.2012.05.013.
[41] N. Chukhrova and A. Johannssen, “Fuzzy regression analysis: Systematic review and bibliography,” *Applied Soft Computing*, vol. 84, p. 105708, Nov. 2019, doi: 10.1016/j.asoc.2019.105708.
[42] C. Kahraman, A. Beşkese, and F. T. Bozbura, “Fuzzy Regression Approaches and Applications,” in *Fuzzy Applications in Industrial Engineering*, vol. 201, C. Kahraman, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 589–615.
[43] U. Hassler and J. Wolters, “Autoregressive Distributed Lag Models and Cointegration,” p. 16, 2005.
[44] K. Dunn, “6.6. Principal Component Regression (PCR) — Process Improvement using Data,” *Learning Chemical Engineering*, Mar. 2020. https://learnche.org/pid/latent-variable-modelling/principal-components-regression (accessed Mar. 12, 2020).
[45] A. A. Kim, Y. Sunitiyoso, and L. A. Medal, “Understanding facility management decision making for energy efficiency efforts for buildings at a higher education institution,” *Energy and Buildings*, vol. 199, pp. 197–215, Sep. 2019, doi: 10.1016/j.enbuild.2019.06.044.
[46] C. E. Graafland, J. M. Gutiérrez, J. M. López, D. Pazó, and M. A. Rodríguez, “The probabilistic backbone of data-driven complex networks: an example in climate,” *Scientific Reports*, vol. 10, no. 1, Art. no. 1, Jul. 2020, doi: 10.1038/s41598-020-67970-y.
[47] C. Theusme *et al.*, “Climate change vulnerability of confined livestock systems predicted using bioclimatic indexes in an arid region of México,” *Science of The Total Environment*, vol. 751, p. 141779, Jan. 2021, doi: 10.1016/j.scitotenv.2020.141779.
[48] P. Zeng, X. Sun, and D. J. Farnham, “Skillful statistical models to predict seasonal wind speed and solar radiation in a Yangtze River estuary case study,” *Scientific Reports*, vol. 10, no. 1, Art. no. 1, May 2020, doi: 10.1038/s41598-020-65281-w.
[49] Solcast, “Global solar irradiance data and PV system power output data,” 2019. [Online]. Available: https://solcast.com/.
[50] H. J. Nussbaumer, “The Fast Fourier Transform,” in *Fast Fourier Transform and Convolution Algorithms*, H. J. Nussbaumer, Ed. Berlin, Heidelberg: Springer, 1981, pp. 80–111.
[51] H. F. Lins, “Interannual streamflow variability in the United States based on principal components,” *Water Resources Research*, vol. 21, no. 5, pp. 691–701, 1985.
[52] I. T. Jolliffe, “Discarding Variables in a Principal Component Analysis. I: Artificial Data,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 21, no. 2, pp. 160–173, 1972, doi: 10.2307/2346488.
[53] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, “Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?,” *J. Climate*, vol. 28, no. 17, pp. 6938–6959, Sep. 2015, doi: 10.1175/JCLI-D-14-00754.1.

[54] D. Maraun, "Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue," *J. Climate*, vol. 26, no. 6, pp. 2137–2143, Mar. 2013, doi: 10.1175/JCLI-D-12-00821.1.
[55] J. Ringard, F. Seyler, and L. Linguet, "A Quantile Mapping Bias Correction Method Based on Hydroclimatic Classification of the Guiana Shield," *Sensors (Basel)*, vol. 17, no. 6, Jun. 2017, doi: 10.3390/s17061413.
[56] P. A. Adedeji, S. Akinlabi, N. Madushele, and O. Olatunji, "Neuro-fuzzy mid-term forecasting of electricity consumption using meteorological data," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 331, p. 012017, Oct. 2019, doi: 10.1088/1755-1315/331/1/012017.
[57] C. D. Lewis, "Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting," *Butterworth-Heinemann*, 1982.
[58] M. A. Saunders and C. Fletcher, "Verification of Spring 2004 UK city temeprature seasonal forecasts," London's Global University, 2004.
[59] J. D. Delorit, S. J. Schuldt, and C. M. Chini, "Evaluating an adaptive management strategy for organizational energy use under climate uncertainty," *Energy Policy*, vol. 142, p. 111547, Jul. 2020, doi: 10.1016/j.enpol.2020.111547.
[60] D. J. Wuebbles, D. W. Fahey, K. A. Hibbard, D. J. Dokken, B. C. Stewart, and T. K. Maycock, "Climate Science Special Report: Fourth National Climate Assessment, Volume I," 2017. https://science2017.globalchange.gov/chapter/4/ (accessed Jun. 09, 2020).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
25-03-2021	Master's Thesis	August 2019 – March 2021

Title And Subtitle

Climate-Informed Prediction And Forecast Modeling Of Installation Total Energy Consumption

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Weiss, Scott C, 1st Lt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way, Building 640
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-MS-21-281

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Intentionally left blank

10. SPONSOR/MONITOR'S ACRONYM(S)
---

11. SPONSOR/MONITOR'S REPORT NUMBER(S) ------

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

Climate variability is an external and stochastic factor that causes energy demand uncertainty. Energy managers can use climate-based models to understand future trends of energy demand and to adjust operations, policy, and budgets accordingly. This research focuses on 1) identifying how climate attributes impact energy use, 2) creating a historically informed statistical modeling framework to skillfully predict energy use, and 3) forecasting future changes to energy use and costs, using CMIP5 temperature projections, at the campus level. After synthesizing the existing breadth of research on climate-informed energy modeling, a skillful, unbiased, climate-informed total energy consumption prediction model is developed for Wright-Patterson AFB (WPAFB) that is particularly skillful at predicting energy use during high and low use periods: the periods where impactful energy policy decisions are made (r2 = 73%, MAPE = 6.15, RPSS = 0.59). CMIP5 projections of temperature inform the model to generate energy use forecasts, which reveal significant changes to energy use within the next decade and increases in annual energy use costs by $7.3-7.9M by the end of the century. Overall, energy use predictions and forecasts can pinpoint the impact of climate factors, inform how and when to mitigate changes, and justify intervention timing and financial decisions.

15. SUBJECT TERMS
Energy, consumption, model, climate, forecast, installation

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	93	Maj Justin D. Delorit, AFIT/ENV
					19b. TELEPHONE NUMBER *(Include area code)* (937) 255-6565, ext 4648 justin.delorit@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18