Understanding Correlated Threats to Department of Defense Energy Systems

by

Katherine E. Adams

B.S. Computer Science and Engineering
Massachusetts Institute of Technology, 2020

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021

Author . . . .
Department of Electrical Engineering and Computer Science
May 20, 2021

Certified by . . . .
Mardavij Roozbehani
Principal Research Scientist
Thesis Supervisor

Accepted by . . . .
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Understanding Correlated Threats to Department of Defense Energy Systems

by

Katherine E. Adams

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 2021, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Climate change poses an existential threat to the United States military's energy systems. We researched current trends in energy, economics, and weather, translating those trends into quantifiable threats to the military's secondary power systems. We also assembled a data set about secondary power systems on domestic U.S. military bases. Because that data set was missing critical information, we formulated and then evaluated an imputation method to complete the data set. This imputation method successfully predicted expected cost for the missing installation data. We ran simulations using our quantified trends and data set on existing software to predict the effects of those trends on certain U.S. military bases. Ultimately, we identified threats that could potentially cost 150 million dollars and cause more than a week of additional electrical downtime for those select bases.

Thesis Supervisor: Mardavij Roozbehani
Title: Principal Research Scientist

Acknowledgments

I would like to acknowledge the Office of the Deputy Assistant Secretary of Defense (Environment and Energy Resilience) for their funding support and guidance in this thesis. Additionally, I thank Erik Limpaecher, Nicholas Judson and the entire Energy Systems group at MIT Lincoln Laboratory for their support of, and belief in, this thesis. I extend my deepest gratitude to Annie Weathers in that same group for her tireless work and dedication in assisting me throughout this project.

The research in this document would not be possible without the faculty of the Massachusetts Institute of Technology. Specifically, I am immensely grateful to my advisor, Mardavij Roozbehani, for his valuable time and advice. Without his input, this thesis would be lacking in much rigor and depth.

I also extend my gratitude to my friends and classmates, especially fellow AI/ML students, Nicholas Bonaker and Cole Hoffer. Finally, thank you to my family, to whom I owe everything.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The average citizen of the United States spends more time per year in blackouts than in eight other developed nations [1]. Meanwhile, the United States military is the largest in the world, with many of its essential operations occurring at home. Since most permanent military bases (i.e. installations) rely on their local commercial power grid, the outages affect them much like the rest of the country. This thesis aims to quantify the costs, in terms of downtime and expenditure, arising from increasing rates of power outages. We also investigate the effects of weather and economic trends on those same metrics.

In general, man-made and natural disasters pose a major risk to United States military installations. Even routine events like a power line falling or a heavy snowfall could disrupt crucial military operations. Specifically, any number of disturbances often disrupt power systems that installations rely on for critical operations. Downtime caused by blackouts is often unacceptable given the sensitivity of work done on installations. But, trying to determine the correct backup plan is difficult. Choosing the right backup power system depends on the anticipated outage scenarios, the acceptable power availability, and the budget. There are many solutions that can accommodate these constraints with varying degree of success.

As a result, the United States military finds itself needing to identify sufficiently resilient and cost effective systems on each of their military installations. Historically, this has been primarily managed by the local facilities administrators at each installation. However, this leaves the Department of Defense (DoD) without a holistic view into the overall readiness of its resilience systems. This is a growing blindspot as climate change is quickly changing the cost and reliability of the American commercial power grid.

The Energy Systems group at MIT Lincoln Laboratory works with the DoD to engineer better backup power systems in American military installations across the globe. As part of this mission, the Energy Systems group has developed the Energy Resilience Assessment (ERA) tool. The tool can predict the expected cost and reliability of a range of backup electricity systems over the next decade for a given installation, provided it has been given a simple power system configuration data for the installations.

In this thesis, we sought to quantify how climate change trends might affect backup power systems on DoD installations over the next decade, using the ERA tool as the backbone of our experiments. However, we faced two major challenges at the beginning of this project. The first was identifying climate change trends that have quantifiable impacts on installations, then translating them into parameters within the ERA tool. The other major hurdle was assembling a representative data set of installation configuration data to run the modified ERA tool on.

We used recent research to identify weather, economic, and energy trends that were being accelerated by climate change. Building on past data collection within the Energy Systems group, we assembled a data set of input data that describes the energy configuration on installations for the ERA tool. What data we could not access, we generated using statistical imputation. We then ran experiments in the ERA tool, testing installations under different climate change strains. In order to develop an intuition for the underlying structure of a complete energy usage data set, we clustered all domestic installations based on that usage data set. We used that clustering to inform our views of how representative our data and results were in general.

To summarize, first we developed validation tests and ran imputations to create a larger, validated data set. Second, we identified climate change and economic trends that might affect the resilience of energy systems. Next, we ran simulations to evaluate the effects of these trends on our new data set. Finally, we used the results from those simulations to identify which trends appear to be the biggest threat to DoD backup power systems over the next ten years. While this thesis is limited by the data we had access to, we believe our findings are an important step in quantifying the climate change threat to the U.S military.

Chapter 2

Literature Review

2.1 Assessing DoD Energy Resilience

US military installations that provide vital services to the Department of Defense depend on continuously delivered power, much like critical civilian facilities. Electricity is needed to power the industrial equipment, servers, fuel systems, hospitals, and communication infrastructure that underpin domestic and foreign military operations. Often, this reliance on electricity goes unnoticed until power is disrupted, and the vulnerability of backup power systems is realized.

The goal of energy resilience analysis is to understand the factors that lead to power disruptions, and to recommend methods of designing energy systems that can quickly and robustly respond to such failures. For example, within a military installation, broader American infrastructure reliability may determine the frequency of disruptions. But the ability to repair equipment, move operations to another facility, and limit electricity use may determine how well an installation is able to respond to such an event.

Resilience analysis, especially as it pertains to energy systems, takes a variety of analysis approaches, ranging from highly quantitative, probabilistic modeling that provides optimized system solutions to qualitative system assessments that emphasize softer metrics. As an example of the former qualitative approach, Ouyang, Duenas-Osorio and Min proposed a three stage resilience framework for evaluating urban infrastructure. [2] In the paper, they offer several formulas to quantify different facets of infrastructure resilience and use those to evaluate and predict threats to a Texas power grid as a case study.

Curt and Tacnet [3] exemplifies the other, qualitative approach to resilience analysis. They define four limiting factors in risk management procedures: unknown events and uncertainty, growing complexity of sociotechnical systems, poorly designed or maintained defense barriers, and errors in procedure. The paper overall focuses on framing resilience analysis as an exercise in predicting and preparing for the unexpected, rather than one in quantifying risks.

Within the context of DoD installation energy systems, probabilistic assessments often suffer from over reliance on quantitative metrics that are hard to define, and even harder to measure. More importantly, this type of analysis often fails to extend beyond foreseeable events that have well defined probabilities. In other words, these approaches are effective at modeling typical, everyday behavior, but do poorly at modeling catastrophic rare cases that can have an outsized influence on operations and public perception.

Military installations are unique from many civilian systems in that their times of greatest risk are often their periods of greatest need. For instance, when a major hurricane strikes the United States, the National Guard is often a part of the U.S. government's disaster response. Having reliable resilience systems means that National Guard locations in the region where the hurricane occurred must be fully functional. In general, U.S. military installations are meant to be assets in a crisis and resilience systems keep them in operation during many such crises. As a result, the DoD prioritizes catastrophic rare cases in the discussion of installation energy resilience.

With this in mind, this work built upon previous work by the Energy Systems group at Lincoln Laboratory in assessing the energy resilience of DoD installations. This previous work included predicting the effects of long-term power outages and different power system structures on individual installations. We expanded on that research in two dimensions. We evaluated the energy resilience of a larger number of military installations and expanded the scenarios the simulator could evaluate.

2.1.1 ERA Tool

The Energy Resilience Analysis (ERA) tool, was developed by the Energy Systems Group at MIT Lincoln Laboratory to help military leadership make informed resilience decisions. Typically used by energy systems management on installations, the ERA tool helps make architecture decisions that line up well with the cost and reliability needs of the installations.

The tool generates an expected cost and effectiveness for both the current and potential energy solutions on a given installation. An energy solution could be as complex as centralizing emergency generators into a microgrid or as simple as building a new solar farm. The expected cost includes capital expenditure (if applicable), maintenance, and fuel costs over the next ten years. The effectiveness metric measures the expected number of blackout minutes over ten years given a certain energy architecture. The ERA tool generates estimates by running failure simulations using a Monte Carlo simulation. The standard number of Monte Carlo simulations in the ERA tool is 1024 iterations. We used this number of simulations throughout this thesis.

2.2 Clustering Techniques

Given unlabeled data, we can identify underlying structure in a data set by finding data points that are quantitatively similar and grouping them together. This approach, called clustering, requires both a metric and a method for measuring similarity between data points. The metric, given two data points, will measure how similar the two data points are to one another. The method dictates the process for both grouping the data points according to the distance metric and choosing which points to inspect and when to inspect them.

2.2.1 K-Means Clustering

K-means clustering is a clustering algorithm. It uses squared Euclidean distance as its distance metric. For an n-dimensional data set, where we are measuring the squared Euclidean distance d^2 between x and y,

$$d^2(x, y) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2.$$

Centroids are the geometric center of planes, determined by the mean squared Euclidean distance of all data points. The k-means algorithm uses centroids to find a cluster for each data point.

During each iteration, it assigns each data point to its closest centroid. After, the algorithm moves the centroid to a new position determined by the plane of data points that were assigned to it in the last round. This is repeated multiple times, until some termination condition is reached. Afterwards, the clusters are determined by the final centroids. All the points that were the closest to one centroid becomes a cluster [4].

Algorithm 1 K-Means Algorithm Pseudocode

```
function KMeans(List<datapoint> X, integer k)
    Initialize length k List<datapoint> centroids randomly
    Initialize length k List<List<datapoint> clusters as empty
    while not converged do
        for datapoint x in X do
            centroid_i = nearest datapoint in centroids to x_i.
            if x in clusters, remove it
            add x to corresponding cluster_i in clusters
        for int j in range(0, k) do
            datapoint centroid_j = center of cluster_j.
    return clusters
```

Elbow Method

An important step when using a k-means clustering is the selection of k. The elbow method is a k visual selection technique, using the inertia and/or distortion of the k-means clustering. Inertia is the sum of squared distances of all data points. Meanwhile, distortion is the average of the the squared distance from each point to the centroid of the point's cluster. After graphing the distortion and/or inertia from k-means clustering over some range of k, we select a k based on the graph. We look for an elbow point, where there is a marked change in the slope of the distortion / inertia, suggesting an optimal k value [5].

2.2.2 Parameter Importance

Results from the k-means algorithm do not include which parameters most determine membership in clusters. A common approach for identifying those parameters is training a random forest classifier on the data [6]. The classifier will train on the same data points and identify the significance for each parameter in determining cluster membership. The parameter importance helps us as researchers develop intuition for the underlying selection structure of the clusters [7].

2.3 Imputation

Oftentimes, data collection methods are not complete and we are left with data sets missing important data points. Imputation allows us to complete the missing data using the other data points as references for the data completion.

2.3.1 Multiple Imputation by Chained Equations

Multiple Imputation with Chained Equations is a commonly used imputation method [8]. Also known as MICE, the model for imputing missing variables uses many chained regressions to predict the missing values. It treats each component (can be either a column or a row) with missing values as the dependent variable in a regression, using some or all of the present values in the data set as the predictors. To begin, MICE puts in a placeholder value (often the mean of the data set, column or row) into each empty data value. MICE next sets a couple of placeholder variables back to missing. Then, MICE regresses on the other variables on the imputation model, using a standard regression model. The missing variables are predicted using the recently fitted regression model. MICE uses a round robin ordering to select the missing variables. Next, it repeats the fitting and predicting from a regression model steps to predict the missing variables. The entire process is repeated through several "cycles", as each missing variable is predicted from regression models several times [9].

2.4 Hypothesis Testing

When running an experiment, often there is a control group and a group that received a treatment. It is difficult to tell if the differences between the control and treated groups is a function of stochasticity within the underlying model. Alternatively, there could be a difference in the ground truth distributions that produced the control and treated results, representing an actual effect from the treatment. We use hypothesis testing to assess the likelihood that the results from the control and treatment groups are from the same distribution.

2.4.1 Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a type of hypothesis test. It is non-parametric. In addition, while other paired tests assume that the underlying distribution is normal (namely, the paired Student's t-test), the Wilcoxon signed-rank test does not make any assumptions about the underlying distribution. Given pairs of data, often representing a treated and a control experiment on the same data point, the test evaluates the likelihood that control and treatment group results were generated by the same distribution [10]. Provided sample size N, with $2N$ total data points (accounting for the pairs), and pairs $i = 0, \ldots N-1$, we have the measurements $x_{0,i}$ and $x_{1,i}$ for a given pair numbered i and R_i is the rank, we define the Wilcoxon test statistic W as

$$W = \sum_{i=0}^{N} [\text{sgn}(x_{1,i} - x_{0,i}) * R_i$$

. After calculating W, we compare it to a critical value found in a reference table. If $|W| > W_{critical}$, we reject the hypothesis that our treated results come from the same distribution as the control results [10].

Chapter 3

Technical Approach

This thesis addresses an open-ended problem constrained by extremely limited information. In order to appropriately account for these limitations, the project required a complex design. While we discuss the mechanics of our data collection and experimentation in Chapter 4, this chapter establishes our overarching experimental design and our data sources.

3.1 Design Summary

To predict the effects of climate change strains on DoD installations, we used the ERA tool to evaluate how strain scenarios might affect specific backup power systems. As described previously (see 2.1.1), the ERA tool was designed to help installation managers make informed decisions about their backup power systems. However, by toggling certain parameters in the installation data used to run the ERA tool, we simulated the effects of specific climate change strain scenarios on those same installations.

As described in the introduction (1), there are two main challenges in using the ERA tool to predict the effects of climate change on military installations. First, the strain scenarios had to be designed. We used recent research to identify ERA tool parameters that are already affected by climate change (i.e. sunniness). Then, we wrote Python code that for each ERA tool input about an installation, would create

15 ERA tool inputs describing that installation under varying degrees of strain.

The other challenge was the limited ERA tool input data. The Energy Systems group has complete ERA tool input data for only eight installations, while there were more than 200 installations in our analysis. We assembled a larger data set by selecting certain ERA inputs sent in by installation managers to the ERA tool. We verified this new ERA input data by running it through a series of sanity checks we developed. Ultimately, our final, ground truth data set had complete ERA tool input data for 33 installations.

Even then, we only have ERA tool input data for a small fraction of the DoD installations. To make matters worse, this data is not entirely representative of the population of DoD installations. We demonstrate the lack of representation using a statistical clustering of a separate energy usage dataset (see Chapter 3.2.1) that is complete across all installations.

We combined the verified dataset with the complete energy use dataset and imputed the missing data using MICE (see Chapter 2.3.1). We selected the data normalization method for the imputation using similar tests we used to select which unverified ERA tool input to include in the ground truth data set. We tested the imputation itself by comparing the MICE predictions for the cost and reliability of the imputed ERA tool input with ERA tool's evaluation of that imputed input.

Finally, we generated 15 strain scenarios from the ground truth data set and ran them through the ERA tool. Our initial plan was to run all the data from both the verified data set and our data imputation. However, due to time constraints and concerns that our ground truth data set was not representative of all installations, we only ran verified installations from a certain cluster.

For a visual summary of this process, see Figure 3-1.

3.2 Datasets

We used several different data sources to complete this research. Some are not publicly available outside of the DoD.

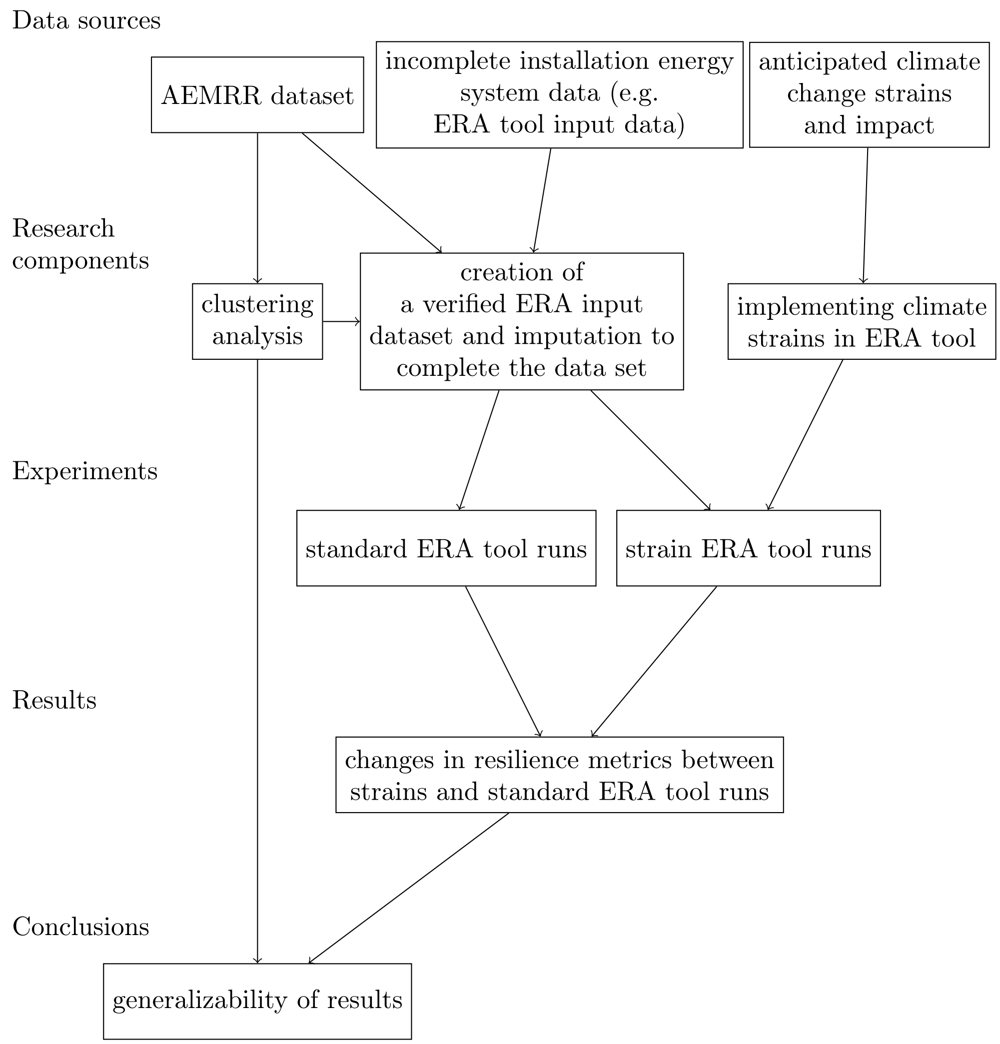

Figure 3-1: A visual representation of the expected workflow of the project

3.2.1 AEMRR Dataset

The Annual Energy Management and Resilience Report (AEMRR) is a report compiled by the Department of Defense annually. It serves as a summary for all things energy for DoD assets. [11] While available in a limited summary form on the internet, we had access to a more detailed version for this research.

We used the 2019 AEMRR report for this thesis.

Usage data

One section of the AEMRR data details the energy usage on each installation during 2019. It breaks down usage into different energy types (electricity from the grid, natural gas, fuel oil, coal, purchased steam, renewable electricity, renewable non-electric, and other) and details the consumption and cost for each of those energy types. It also segments usage data into infrastructure that needs to be meeting certain sustainability goals ("goal included") and infrastructure that does not ("goal excluded"). However, since this analysis is focused on resilience, and not sustainability, the "goal included" and "goal excluded" columns are combined.

3.2.2 EIA Price Index

The United States Energy Information Administration (often referred to as the EIA) is an analytical agency within the U.S. Department of Energy. They focus on recording and forecasting energy usage and costs. [12] Among their many publications, they release a yearly report on their forecasts for future energy costs. We used Table 3 (Energy Prices by Sector and Source) to create some of the strain scenarios for the tool. We used the industrial prices (rather than residential and commercial), because military energy pricing most closely follows the industrial costs. We specifically used price forecasts for distillate fuel oil (diesel), natural gas, and electricity. The report has a baseline estimate for the cost of each energy source, as well as predictions under different economic conditions. [13] The different economic scenarios we used were high economic growth, low economic growth, high oil price, low oil price, and different potential carbon credit programs.

3.2.3 ERA Tool Input Data

The ERA tool needs information about the baseline energy system on an installation in order to run. It allows the tool to generate an accurate Monte Carlo simulation of the installation's energy system. There are about 70 different parameters, most of them are single numerical values or booleans. Those include weather information (number of clear days, sunniness, average wind speed, etc.), cost data (capital expenditure cost for new energy systems, maintenance cost for those same systems, installation costs, fuel prices, etc.), and system configuration data (existence of a natural gas system, solar arrays, local electricity grid on the installation) There are also a handful of list based parameters (length of each historical electricity outage, failure type of each outage, capacities of each primary generator, capacities of each standby generator).

Web Application Data

The ERA tool is hosted on a private web application by the DoD. Many installation managers use this application to evaluate the resilience of their current installation configuration. Anytime that the tool is run, the tool input data is stored at MIT Lincoln Laboratory, including which installation the manager is from. While there are a handful of common sense data checks on the web application (i.e. no values are negative), there is no guarantee that the data entered by the installation manager is accurate.

3.3 Trends

There are documented energy and weather trends that have implications for the resilience of secondary power systems. We detail how we use these documented trends in our strain analysis in 4.5.

3.3.1 Wind Speed

In order to produce energy, wind turbines rely on land surface winds with high enough speeds to move the turbine blades. Wind power (p) is a function of the cube of wind speed (u), air density (ρ), swept area of turbines (s), and efficiency factor (f) according to the formula:

$$p = \frac{\rho s f}{2} u^3$$

However, wind speed is not a consistent variable when it comes to predicting wind power. As shown in Zeng, Zieglar, et. al. [14], there have been conflicting and shifting trends in wind speed in the last 4 decades. From the 1980s to 2010 there was a reported 8% global decline in land surface wind speed, referred to as global terrestrial stilling although there is still no consensus on the source of this change. On the other hand, recent studies indicate that over the past decade there has been an increase in terrestrial wind speed. While no evidence of an years-long global trend increase in wind speed has yet been compiled, there is a consistent upward trend in wind speed at many local sites over the past few years [14]. Given the inconsistency in wind speed trends, we determine that it is difficult to predict future wind speeds.

3.3.2 Sun Intensity

Photovoltaic solar energy uses solar radiation to produce electricity [15]. However, solar radiation is not a consistent resource. The mean solar radiation the Earth experiences changes year over year. Additionally, there is significant intermittency in energy generation on a shorter time scale, causing potential interruptions in service. Yin, Molini, et.al, [16] compiled the results from 11 climate change models to predict the relative change in the mean clearness index from the period 2006-2015 to the period 2041-2050. The mean clearness index measures the fraction of the irradiance from the sun the ground experiences (accounting for the scattering, absorption, and reflection of the solar radiation). The index is frequently used in assessing the energy draw possible from a photovoltaic solar arrays.

3.3.3 Power Outages

According to Climate Central [17], there has been approximately a tenfold increase in the frequency of major power outages in the United States from the 1980s to the 2010s, with the designation of major power outages is done by the non-profit NERC. Climate Central also found that the annual number of climate related outages doubled from 2003 to 2014 [17]. According to a study done by Inside Energy on outage data from 2000 to 2014, the five-year annual average of outages doubled every five years [1]. Other sources offer slightly different estimates of the rate of increase of power outages, but they both agree that there has been a marked increase over the past few decades [18][19].

It is important to note that we did not collect conclusive evidence that the increasing rate of outages is a function of climate change. While such research may exist, we did not look for it. Given that caveat, this increasing rate of power outages may be a function of aging infrastructure alone or combined function of aging infrastructure or climate change. However, for the purposes of this research, we were focusing on identifying trends and speculating their effects if they continued or accelerated. So, we are assuming that the trend of increasing outages will continue.

3.4 Addressing Incomplete Tool Input Data

3.4.1 Summary

As mentioned previously, we had limited ERA input data to run the tool on. In order to supplement that data set, we used data that the Lincoln Laboratory team did not put together. We had two sources of potential supplementary ERA input data: the web application data and a matrix imputation. To ensure that all data we used in our final analyses is reliable, we ran each ERA tool input through validation tests. Finally, to evaluate how representative our final ERA input data set is, we clustered installations based on their AEMRR usage data. Using those clusters, we identified the limitations of our verified data set.

3.4.2 Clustering

We identified clusters within the AEMRR usage data (3.2.1) using standard n-dimensional clustering methodology, as described previously. In order to understand how normalization of the data set would affect the clustering, we tried different normalization techniques prior to running the clustering algorithm. We selected the best clustering by working with the Energy Systems group to qualitatively evaluate which clustering was the best representation of different categories of installations. These clusters informed later insights into how representative our limited data set was.

3.4.3 ERA Input Validation Tests

In order to validate that any ERA input data was reasonable, we worked with the Lincoln Laboratory team to develop tests. Those tests are effectively sanity checks, making sure that a couple data points from the ERA tool input is within a reasonable range of that reliable supplementary data. For instance, one test checked (after appropriate unit conversion) that the ERA input data value for electrical load was reasonably close to the AEMRR total electrical use parameter for a given installation. The tests themselves are described in more detail in the methodology chapter.

3.4.4 Verified Data set of ERA Tool Inputs

As mentioned previously, the Energy Systems group only had complete ERA tool input data for eight installations. This is not enough data to either run the matrix imputation on, let alone make generalizations about the resilience of DoD bases from strain testing. Therefore, we ran the web application data (see 3.2.3) through the validation tests. Those ERA tool inputs that pass the tests, we combined with the data from the eight installations put together by the Energy Systems group. This final, combined data set is called the ground truth data set of ERA tool inputs throughout this thesis.

3.4.5 Matrix Imputation

We combined AEMRR usage data (3.2.1) with the verified ERA tool input data (3.4.4) into one matrix indexed by installation, with missing ERA input data parameters as null entries. We added new columns to the matrix, containing the cost and effectiveness results from running the ERA tool on the installations from the verified data set.

Next, after normalizing the matrix, we ran the MICE algorithm (2.3.1) to fill in the null entries in the matrix. We repeated this process several times, each time with a different normalization technique. We tested the normalization techniques by running the validation tests on each installation with imputed ERA input data from the matrices. After selecting the matrix that performed the best on the tests, we ran each imputed installation through the tool.

We compared the imputed values for the cost and effectiveness with the ones the ERA tool produced. The magnitude of error enabled us to evaluate how well the imputation picked up on the internal structure of the data set. If the error was large, then the imputed ERA tool values were not correct, as the ERA tool results from those imputed inputs do not line up with the imputed predictions from that imputed data. However, if the error is small, then the imputation replicated most of the underlying relationship between the AEMRR data, the tool inputs and the tool results.

3.5 Developing Strain Scenarios

3.5.1 Summary

As a first step in understanding the effects of future strains on DoD energy systems, we considered a few different potential scenarios that we would expect to change the results of the ERA tool. We chose this approach because we didn't have to make any assumptions about the probabilities of strain scenarios occurring. Instead, we developed them from current research

Another major factor in the methodology for the ERA tool experiments was time. The ERA tool is not entirely automated nor fault tolerant. Each individual scenario run for one installation takes approximately five minutes and we could reliably batch about twenty different scenario runs for one installation each run. With the number of experiments we initially wanted to run, the tool would've been running for more than 400 hours. Given that batches needed to be run and their results manually inspected approximately every other hour, this plan had to be abbreviated.

3.5.2 Strain Scenario Approach

Ultimately, we decided on developing "strain scenarios." These events were designed based on already predicted or documented trends, such as the increasing number of blackouts per year in the United States. We used the variable inputs that the ERA tool takes as our starting points for identifying energy and economic parameters that could change over the next decade, changing the ERA tool's predictions. We chose parameters for our strain scenarios from there based on the availability of reputable research that explicitly predicted changes in the parameter's value over at least the next decade. We settled on the following parameters to research and then changing their values: sunniness, wind speed, diesel cost, grid electricity cost, and natural gas cost. By sweeping through these values, we can understand the sensitivity of cost and performance of power systems to these potential future strains. We do not make any predictions about the future, instead, just using recent research to estimate the range of potential values for these relevant parameters.

Using the predictions from the EIA (see Chapter 3.2.2) and the papers referenced in Chapter 3.3, we identified three values for each parameter: a low, moderate, and high rate of change. Using the rate of outages as an example, the low rate of change for outage predictions was no change in the ERA tool, while the high range was two times the rate of outages over the next decade. We could then run ERA tool experiments where we toggled a parameter and evaluate the expected change in cost and downtime for a given installation.There were four distinct categories of strains to resilience systems.

Energy Prices

The first parameter we considered was energy prices (including grid electricity cost, diesel cost, and natural gas cost). While there are multiple ERA tool parameters tied to energy prices, the predictions come from the same set of EIA predictions. Those

EIA predictions come from the same models for potential economic growth and policy changes, meaning that for every prediction about the grid electricity cost, there's a corresponding predicted natural gas cost. This allowed us to run the tool with low/medium/high energy price predictions rather than running different permutation of electricity and natural gas costs.

Weather

We had no model for the relationship between the two weather parameters (wind speed and sunniness). So we treated the two as distinct scenarios. We toggled the sunniness separately from the wind speed in our first experiments. However, once we had compiled our final ground truth data set, we discovered that none of our data points had any wind turbines. As a result, we abandoned the wind speed analysis, as it would make no tangible difference in either of the metrics we were looking at. A handful did have solar panels, so we continued to investigate the effect of changing solar radiation.

Power Outages

The final category was the rate of power outages. We decided to run all our experiments toggling one of the other categories (ultimately, either energy prices or sunniness), along with toggling the rate of outages on the military installations. We made this decision because the rate of outages most tangibly changed cost and reliability metrics in our initial experiments. Additionally, during our research, we found the most convincing evidence for the consistently increasing rate of outages over the last two decades compared to the other strains we investigated.

Chapter 4

Methodology

4.1 AEMRR Data Set Refinement

The AEMRR 3.2.1 data set included information that was not relevant to our research, so we removed or combined the irrelevant data. In order to identify whether an aspect of the data set was useful, we focused on the end goal: identifying the reliability and cost of the military installation as a whole.

4.1.1 Duplicate Data

There were multiple columns in the AEMRR data set with duplicate columns. The columns contained the same information, just in different units. After verifying that the duplicated data was consistent after conversion, we dropped any energy column that used a unit other than megawatt hours.

4.1.2 "Goal Included" and "Goal Excluded" Data

There were several columns that mentioned either "goal included" or "goal excluded" columns. As mentioned in subsection 3.2.1, these columns described the energy use in certain assets on a military installation. If all the "goal included" and "goal excluded" assets were listed together, they would be the complete list of assets on a military installation. Because "goal included" and "goal excluded" designations are relevant for sustainability concerns, not energy usage on an installation, so we decided this was not a relevant component of the AEMRR data set. For "goal included" and "goal excluded" data, if there was not already a column that combined the information into a total, we created that column and dropped any remaining "goal included" and "goal excluded" columns.

4.1.3 Combining AEMRR Installations

The energy statistics for a single installation were often split across different rows in the AEMRR data set. An installation might be represented across several rows, with the row's data coming from different components of the installation such as:

- craft services (i.e. the kitchens and mess halls)
- administrative buildings
- training facilities

The different services or groups that use the energy on a given installation often use the same electrical grid and backup power systems. Since we're interested specifically in the performance of that shared infrastructure, we did not need the level of granularity in the AEMRR data set.

So, we combined the rows that we believed were from the same installations. When cleaning the data, provided that an installation had the same name and was in the same state as another installation, we combined the two installations by adding all their numerical data together column by column. This allowed us to quickly combine all the organizations we believed were sharing the same energy system. A quick check by a Lincoln Laboratory colleague confirmed that the number of installations after collapsing these data points was about the expected number of domestic military installations.

We also removed all installations that were not in the 50 states and District of Columbia in the United States of America. Analyzing installations in territories and in the rest of the world would add more complexity to the data set. Given the abbreviated nature of this project and therefore thesis, we removed the extra dimension of predicting strains internationally and the computational load of having even more installations to analyze.

4.2 Clustering Methodology

4.2.1 K-Means

We used k-means clustering to find the underlying structure of the refined AEMRR data. Specifically, we used the k-means clustering method from the Python package scikit-learn. We ran the algorithm with the standard settings from that package and utilized the tool's inertia measurements. Additionally, we calculated the distortion using the Euclidian distance function `cdist` from the Python package scipy. After running the k-means algorithm for values of k between 1 and 15, inclusively, we created elbow plots from the inertia and distortion measurements. We decided on the k values using those elbow plots.

4.2.2 Normalization

For the clustering, we further refined the data. Because k-means clustering uses Euclidean distances, the clustering can be overly influenced by parameters with large values. Since there were many units across the columns (United States dollar, megawatt hours), running the clustering algorithm on the data set as-is risked overweighting certain parameters. In order to avoid incorrect clusters, we developed and tested several normalization techniques.

For the baseline normalization process, we used a standard normalization technique where for each column, we found the mean and standard deviation. Then for each value in the column, we subtracted the mean from it and divided by the standard deviation. This method centers each column's values around zero and removes the units from the data, so issues of scale are reduced.

As we began running the clustering algorithm, we found that the normalization itself was producing good results. However, the clustering was clearly being dominated by the total amount of energy consumed (described in detail in section 5.1). We knew that both the amount of energy consumed is important in determining the cost and effectiveness of backup power systems, but did not represent the whole story. The different types of energy used and the proportion of that energy use matters as well.

We tried two other clustering experiments using the same normalized data set that the original clustering experiment used. For each already normalized row, we divided each value by the "Total Consumption" parameter value for that data point. We intended for this to scale each row by the total energy that was consumed on the installation, allowing the clustering algorithm to cluster based on the different energy types used and the energy costs overall. We also attempted something similar again, normalizing based on the used square footage on the military installation to cluster based on the energy consumption relative to the assets on the installation.

We trained a random forest classifier using the original refined AEMRR data set as the input data and the clusters from each normalization as the output. As described in 2.2.2, the parameters that the random forest identified as the most important in determining which installation belonged to allowed us to identify the defining characteristics of each cluster.

4.3 Assembling a Ground Truth Data Set

As mentioned in previous chapters, the Energy Systems group at MIT Lincoln Laboratory only had access to eight complete ERA tool inputs when this thesis was written. With 280 installations in our analysis, this was far too few. Therefore, we set out to create a ground truth data set that compiled all the installation energy usage and outage data we had access to.

We created a ground truth data set in several steps.

1. We combined the eight complete ERA tool inputs (each one about a different installation) with our previously assembled and cleaned AEMRR usage data set

to form the first version of the ground truth data set. This created new columns for all the ERA tool input parameters compared to the original cleaned AEMRR data set.

2. We filled in single value supplementary data (i.e. the number of substations on an installation).

3. We ran validation tests on the ERA tool input data from the web application. Any input data point that passed all the tests were added to the ground truth data set. This would not create any new rows or columns, instead just filled in empty values.

4. For every installation in the ground truth data set with complete ERA tool input data, we ran the installation through the ERA tool. We store some of the results from the ERA run as new columns in the ground truth data set.

4.3.1 Matching Installation Names

One challenge we faced as we combined our data sources was matching the names for installations to different names that referred to the same installation. Between differing abbreviation standards and bases with different names than the larger joint bases that contained them, the name matches were not always obvious.

We matched the names to one another in a multi-step process:

1. We used the string matching Python package FuzzyWuzzy to match each installation name in a supplementary data set to an AEMRR installation name.

2. We manually inspected the name match with a member of the Energy Systems Group.

3. If a name match was incorrect, we replaced it with the correct AEMRR installation name if one was available

4. If we couldn't find a correct AEMRR name, we assumed the installation was not in our analysis, and we ignored the data point.

4.3.2 Supplementary Parameter Data

The Energy Systems group had three lists that we used to supplement the ground truth data set.

The first was a master list of substations across many installations. This list contained a close to exact count of the number of substations for a fraction of the installations in our analysis. As number of substations was an ERA tool input parameter, we totalled the number of substations on each installation present in the list and imported it into the ground truth data set.

The second was a master list of backup generators. Like the list of substations, it was close to exact for the installations that it had data for. Again, the number of emergency generators was a parameter in the ERA tool input, and therefore a column in our ground truth data set. We compiled the list of emergency generators for each installation listed and imported it into the ground truth data.

The final was a list of installations with completed microgrids. This was straightforward, coming from an MIT Lincoln Laboratory report that listed all installations with microgrids on them. For the installations that were included in our analysis, we updated the ground truth data to reflect that they had microgrids. Similarly, for any installation not listed in that report, we noted that those installations did not have microgrids in the ground truth data.

4.3.3 ERA Tool Input Verification Tests

For each unverified ERA tool input (either generated using MICE or created by a non-Lincoln Laboratory ERA tool user), we needed to evaluate if the configuration was reasonable.

Using the data that we had compiled from the AEMRR and a handful of Lincoln Laboratory data sets, we checked that the unverified configuration parameter values were within a certain reasonable range. Given actual value a and the expected value x, we used this equation to measure the error E:

$$E = \frac{|a - x|}{a}.$$

For a handful of cost variables, we used a simpler test that verified the given energy cost is within the U.S. national range of commercial energy prices, as indicated from the range in the AEMRR data set.

We used a different set of tests for the data from the web application than the tests we used for the matrix imputation. All of the supplementary data we used to check the ERA web application data, we added to the ground truth data that was the source for the imputation. Therefore we couldn't use those tests that relied on the supplementary information (since the imputation didn't impute data already existing in the ground truth data set).

4.3.4 Selecting ERA Tool Inputs from the Web Application

There was data from the ERA tool web application (see section 3.2.3) that we used to fill in more of the ground truth data set. The ERA tool web application is used by stakeholders to investigate the resilience of their installation and potential architecture improvements.

However, we had no guarantee that ERA tool input data entered into the web application was representative for the given installation. Using the tests described previously, we ran the most recent web application input for each installation through our tests. Each ERA tool input that passed all tests (i.e. had a test value between the lower and upper bounds) was considered verified.

The verified inputs were then added into the ground truth data set. There were no new columns or rows added, instead they went into the already existing ERA tool parameter columns. If there was data from the supplementary Lincoln Laboratory data sets already there, it was overwritten by the ERA web application tool data.

Actual value	Expected value	Test type	Lower bound	Upper bound
ERA tool input substation count	LL substation count (if known)	error equation	-10.00	10.0
Sum of ERA tool input generator capacity	ERA tool input average electrical load	error equation	-3.00	3.0
ERA tool input average electrical load	AEMRR consumption ÷ 8760 hours	error equation	-1.00	1.0
ERA tool input critical thermal load	ERA tool input cogen + steam capacities	error equation	-10.00	10.0
ERA tool electricity cost		within range	0.01	0.5
ERA tool natural gas cost		within range	0.10	30.0
ERA tool average PPA cost		within range	0.01	0.5

Table 4.1: ERA input data verification tests (for web application data)

4.3.5 Running the ERA Tool on Complete Input Data

The final component of our ground truth data set were the results of an ERA tool run. We selected two of the summary outputs from the ERA tool and added them as two more columns in the ground truth data set. The two columns were the total expected cost of the installation's backup power system (including fuel and capital expenditures) and the expected downtime, both over the next decade.

We ran the verified entries (i.e. the eight Lincoln Laboratory ERA input installations and the ones selected from the web application) in the ERA tool. For these

Actual value	Expected value	Test type	Lower bound	Upper bound
ERA tool input average electrical load	AEMRR consumption ÷ 8760 hours	error equation	-0.1	0.1
ERA tool input grid electricity cost	AEMRR electricity cost ÷ (100 × elec. consumption)	error equation	-0.1	0.1
ERA tool input natural gas cost	AEMRR natural gas cost ÷ (10 × n.g. consumption)	error equation	-0.1	0.1
ERA tool input average PPA cost	ERA tool input grid electricity cost	error equation	-2.0	2.0

Table 4.2: ERA input data verification tests (for imputed data)

specific ERA tool runs, we did not use the strain scenarios we designed. We selected the seven day black sky outage because it is a common outage duration used when designing installation energy systems. We imported the results from the ERA tool run into the ground truth data set.

4.4 Matrix Imputation

Even with the additional data from the web application, we didn't have ERA tool inputs for a majority of the installations. Using MICE (see 2.3.1), we ran a matrix imputation on the ground truth data to generate ERA tool inputs for the installations with missing data.

4.4.1 Data Preparation

MICE only runs on numerical data and will impute any missing numerical values. The ground truth data set included Boolean, list, string, and numerical data. To create a data set that MICE could run on, we removed all non-numerical data.

All the string parameters the tool needed, we either already had or we could generate easily.

The Booleans were mostly based on numerical values. We inferred those Booleans in post processing. Another Boolean indicated if the installation had a microgrid or not. We already had a list of which installations had microgrids, so we used that list prior to running the imputation. The other Booleans were settings that were consistent across installations (i.e. including the black sky outage).

There were four list-based parameters. Two were outage related, and the other two were lists of generators (prime and emergency specifically). While the ERA tool could estimate the number of outages, it could not predict the generator lists for a given installation. We created two new numerical columns where we summed the prime and emergency generator capacities, respectively.

We also changed certain empty values to zero. For instance, any empty values in rows of verified installations were set to zero. Since the tool inputs were complete, we knew that those empty values meant zero. Because we did not want MICE trying to predict those values, it was imperative to change the values from empty to zero.

Normalization Techniques

Like the clustering, we used different normalization techniques to try to improve the MICE results. When applicable, we normalized over columns (i.e. over the parameters in our matrix, not the installations).

Standard normalization: No normalization at all.

Z-score normalization: Given mean μ and standard deviation σ of a column and a value v from that same column, the normalized value v_n is

$$v_n = \frac{v - \mu}{\sigma}$$

.

Standard deviation normalization: Given standard deviation σ of a column and a value v from that same column, the normalized value v_n is

$$v_n = \frac{v}{\sigma}$$

.

Logarithmic normalization: Given value v, the normalized value v_n is,

$$v_n = \log_2(n)$$

.

4.4.2 Imputation

We used the Iterative Imputer method from the Python package scikit-learn. We ran the imputer with some custom parameters. We set maximum iterations to 2000 iterations of MICE. We also set the minimum predicted value of 0, since no value in the data set should be negative. Finally, we had an error tolerance of 1e−3. This value was selected by hand and the one we selected was the lowest value that improved the results. Using the normalized and prepared data, we ran the imputer. The imputer outputted the results of running MICE procedure on the data set. We also used the Simple Imputer method from the Python package scikit-learn. We ran our entire preprocessing, imputation, and validation process on the results from the simple imputer. This allowed us to evaluate the sensitivity of our validation tests.

4.4.3 Data Post-Processing

We took several steps to process the data after the imputation and return it to its intended state.

1. We reversed the normalization by running the inverse of the normalization function on the imputation output.

2. We converted certain columns from floats to integers, based on which columns contained integer parameters (e.g. the number of substations).

3. We restored the Boolean columns. For the Booleans that depended on another column (usually the capacity of a certain technology) we inferred it from the imputed value. For the Booleans that were constant across installations (e.g. including the black sky simulation), we set those to our predetermined values.

4. For installations without supplementary generator data, we took the sum of generator capacities (either prime or emergency) and divided it by 300. Rounding the value we received, we listed that number of 300 kW generators for the installation. The 300 kW number was chosen as a representative, standard generator size

4.4.4 Data Validation

We validated the different normalization methods using two techniques. The first verified that the normalization was reversed correctly after the imputation was run (e.g. there was no substantial change in the non-missing values). Selecting only the data from the installations with complete ERA tool inputs, we checked the difference from the data before the normalization and after it. Then we took the mean and standard deviation of those errors, to verify that there wasn't too much data loss from the MICE procedure and/or the normalization technique.

The next validation step was running the tests (the tests themselves are enumerated in section 4.3.3). For each normalization technique, we ran each imputed ERA tool input through the tests. We stored which imputed inputs passed the tests and scored the normalization techniques based on the proportion of installations that passed.

4.4.5 ERA Tool Experiment

After selecting the normalization techniques that performed the best, we used the imputed data set they had produced as our main imputed data set. Then, we ran all the imputed ERA tool inputs through the tool. We evaluated the imputation performance based on the difference between the tool's output for the two output columns (total expected hours of downtime and total expected cost) and the imputed values for those same columns. The rationale for this method of testing is described in depth in the previous chapter (see 3.4.5).

4.5 Strain Analysis Methodology

4.5.1 Installation Selection

As described in the previous chapter, we were limited by the sheer amount of time it took to run the ERA tool (see 3.5.1). Also, two thirds of the strain experiments for a given installation relied on having outage history. More importantly, we trusted the ground truth data set more than the imputed data set. With this all in mind, we selected 26 installations to run through the strain experiments. We only chose the installations that were complete in our original ERA ground truth data set, that had outages listed, and were in our dominant cluster.

4.5.2 Experiment Configuration

As described previously, the strains experiments toggle either a price change or a sunniness change for a given ERA input. Then, those toggling experiments are all rerun with different (increasing) rates of outages (see 3.5.1 for the rationale). Ultimately, including a "standard run" control where the price and sunniness is not toggled, there are 15 experiments.

Parameters

For each strain variable (sunniness, price change, rates of outages) we had three rates of change: low, medium, and high. For all the strains, we generated the medium value by finding the mean between the high and low rate of change.

For the sunniness, the rates of change came from the paper discussed previously (see 3.3.2). There were two different graphs of sunniness. One was a prediction of the sunniness in the 2040s and one was a report of the sunniness in the 2010s. We found the low rate of change by subtracting the lowest value in the range of the 2010s from the predicted sunniness in the 2040s, then dividing by the difference in years from the middle of each periods. We found the high rate using the same procedure, just choosing the highest values rather than the lowest.

The price change values came from the EIA data (for more detail, see 3.2.2). We were interested in several different EIA forecasts: high economic growth, low economic growth, high oil price, low oil price, and different potential carbon credit programs. We were looking at three different industrial energy prices: distillate fuel oil (diesel), natural gas, and electricity. When looking at the graphs provided by the EIA to visualize their different predictions, we noticed that most of the time, a rise or a fall in one energy type price, causes a similar change in the other price predictions. Therefore, we "bundled" the energy prices, assuming for the purposes of the strain experiments, that the lowest energy prices would co-occur, as would the highest. We selected the high rates of change for each of the energy types by finding the highest predicted price in 2031 (the end of the strain simulation's time frame) and dividing it by ten. We did the same to find the low rate of change, except we found the lowest, rather than highest, predicted price in 2031.

The exact rate that power outages are increasing across the United States is unknown and it is not monotonically increasing. However, the consensus from the literature review (described in detail in 3.3.3) is that they are increasing decade over decade. We decided to assume that there were double and triple the number of outages occurring each decade, for the moderate and high outage rates respectively.

As noted previously, this might chance given serious investment in the U.S. energy grid infrastructure, but we are interested in predicting effects of current trends continuing.

In order to convert the sunniness and price change rates of change into usable strain experiments, for a given ERA tool input value v corresponding to a strain with an expected rate of change r, the strain value v_s is

$$v_s = v \times r^5$$

. The rate r is raised to the fifth to approximate the amortized value over the decade that the experiment is run for.

Strain Scenario	low rate of change (per year)	high rate of change (per year)
Diesel Cost	-3.02	3.88
Electricity Cost	-0.37	1.09
Natural Gas Cost	1.15	7.16
Mean Clearness (Sunniness)	-0.23	-0.15

Table 4.3: Rates of change for generating strain scenarios

Outage Generation

In order to generate the increasing rate of outages strains, we also used an amortized rate measurement. The moderate increase in outages approximated the rate of outages doubling by the end of the decade. The high increase in outages was the same, except it approximated the rate of outages tripling by the end of the decade. Because we only have data for the rate of commercial utility outages, this rate increase only includes outages described as type 1 outages by the ERA tool.

We first created a list of only the type 1 outages from the original list of outages.

For each set of strains with a moderate increase in outages, we created a new list .6 times the length of the type 1 list. This list was created using sampling with replacement from that same list of type 1 outages. Finally, the new sampled list was appended to the original list (including all outages of type 1, 2, and 3). The same procedure follows for strains with a high increase in outages, except the list created by sampling is 1.4 times the length.

4.5.3 Hypothesis Testing

After running the ERA tool on all the different strain scenarios, we had two important statistics for the effects of each strain scenario on a certain installation:

- the expected number of hours of downtime for the next decade
- the expected expenditure on maintaining and running the backup power system over the next decade

We used hypothesis testing, specifically the Wilcoxon signed-rank test, to evaluate if the strains were having a significant effect on the backup power system cost or downtime across the installations.

Chapter 5

Results

5.1 Clustering

We selected the k value for the cluster based on the elbow plot (see Figure 5-1). We noted that the elbow plots produced by the cluster using a standard normalization did not have the pronounced elbow that would indicate a particularly descriptive k value. All three clusterings had one cluster of more than 200 installations and all other clusters smaller than 50 installations (see Figure 5-2). As shown in Figure 5-3, the feature importance for the standard normalization clustering was dominated by the total consumption and cost values. However, for the clusterings produced by the other two normalization methods, there is a much more equitable split of feature importance across the top ten features.

Ultimately, we decided to use the clustering produced by the normalization by total consumption as our primary clustering. Therefore, when we could only use one cluster designation (i.e. selecting which cluster to use to run strain experiments), we used the normalized by total consumption clustering as our default clustering. Three factors contributed to our selection of this clustering. First, its elbow plots had well defined inflection points. Secondly, it had the most well distributed parameter importance graphs. Finally, there were concerns from the Energy Systems group that the square footage measurements in the AEMRR might be inflated for certain installations, while the total consumption metric is much more reliable.

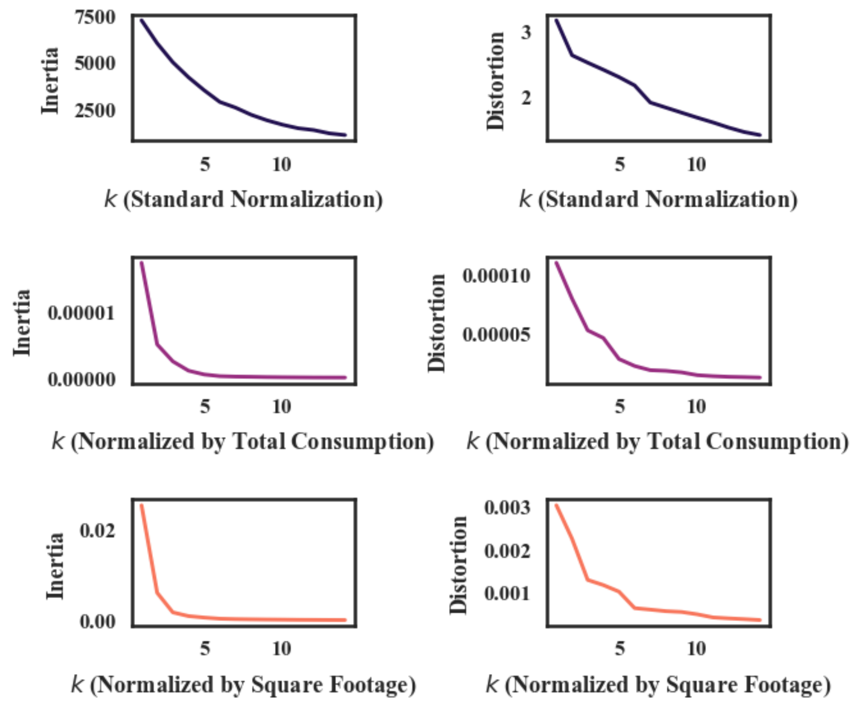

Figure 5-1: K-means elbow plots for different normalization techniques

5.2 Supplementary ERA Input Data

Once we added the 26 new ERA tool inputs from the web application to our ground truth data, we had a verified data set of ERA tool inputs with 33 data points. There were 33, not 34, because there was an installation in the web application data that we already had in the Lincoln Laboratory data. Our verified data is close to representative of the clusters (see Table 5.2). However, as shown in Figure 5-4, cluster one is just substantially bigger than clusters zero and two, and the verified data is very small compared to the ERA tool input data that is incomplete.

5.3 Imputation

After running the matrix imputations, we validated the results using the tests we designed. We also ran a simple mean imputation, which began the same way as the MICE (placing the mean in all empty entries), but terminated immediately after that. The simple mean imputation performed the best, with every installation passing the

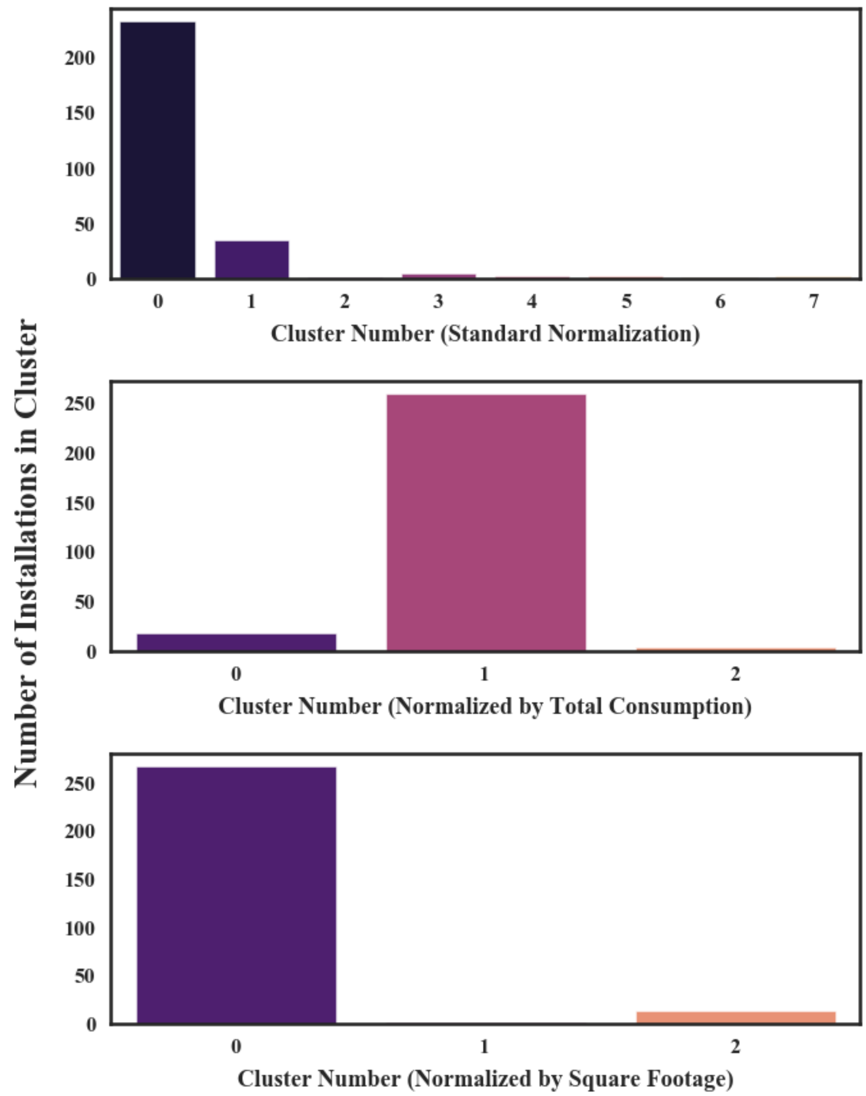

Figure 5-2: Number of installations in each cluster for a given normalization technique

tests, with no normalization and standard deviation normalization trailing behind with the second and third best scores, respectively. See Table 5.2 for the specific scores.

We knew that the result of the simple mean imputation is not a good imputation. Very small installations do not have the same number of substations, same electricity cost, etc. as large installations. As a result, we were wary of the validation test results. We ran both the standard deviation normalization and the no normalization imputation results through the ERA tool.

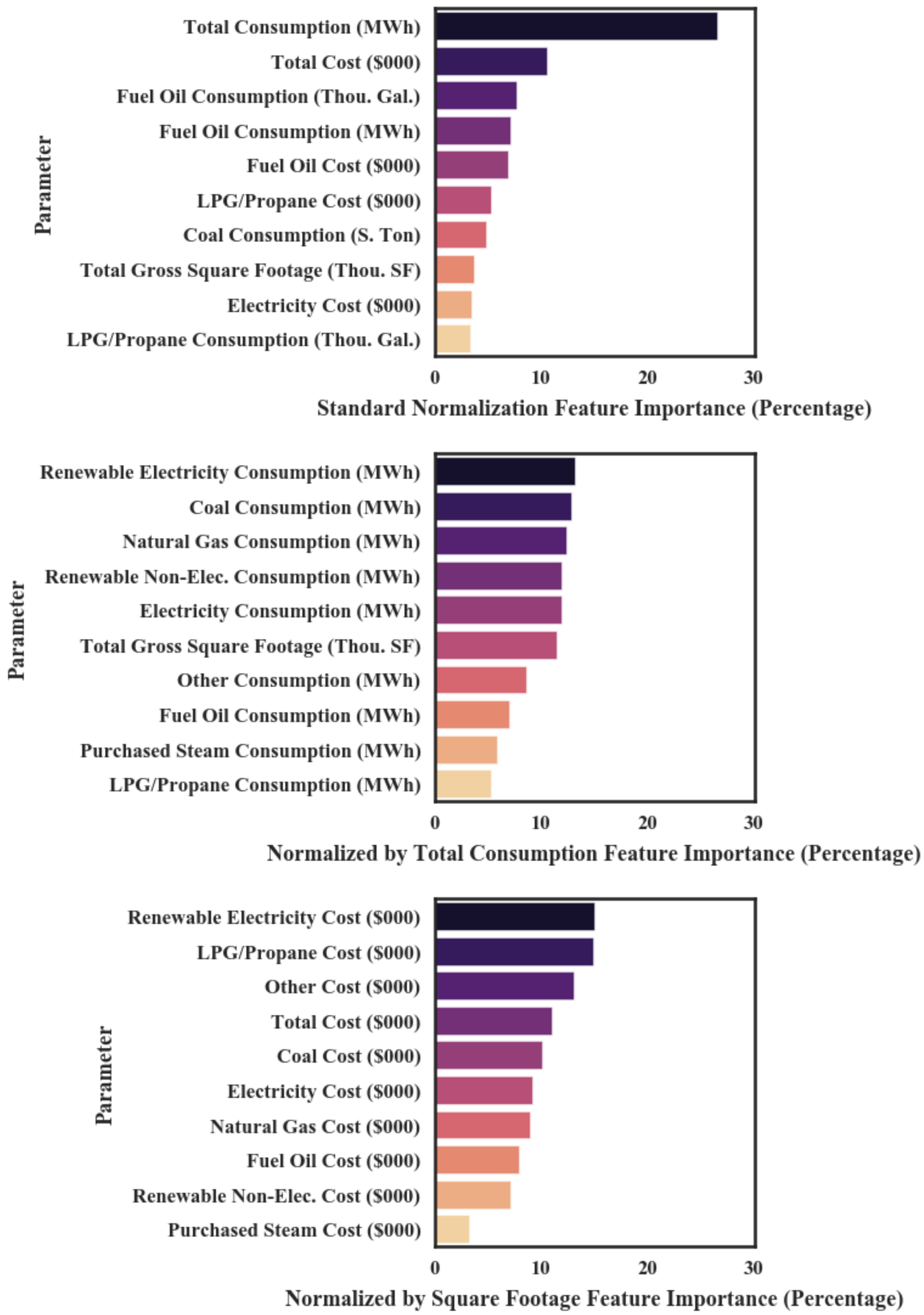

Figure 5-3: Feature importance for each clustering for a given normalization technique, only including the top ten parameters by percentage

Despite no normalization imputation performing better on the validation tests, the standard deviation imputation performed significantly better on predicting the ERA tool results. While both imputations performed poorly in predicting downtime,

Status	Cluster ID	Number of Installations	Percent of Installations in Cluster, by Status
Unverified	0.0	16	6
Unverified	1.0	227	92
Unverified	2.0	4	2
Verified	0.0	2	6
Verified	1.0	31	94
Both	NaN	18	6
Both	NaN	258	92
Both	NaN	4	1

Table 5.1: Breakdown of total cluster membership by installation verification status

the standard deviation imputation predicted the operational expenditure with an R^2 value of .94 (see 5-5). A possible cause of the worse downtime prediction performance for the imputation is how most of the AEMRR data is about energy consumption, not outages or downtime.

5.4 Strain

When we ran the strain scenario experiments, we decided to only run the verified ERA tool data input data that came from installations in cluster one. 31 of the 33 verified data points were from installations in cluster one. Given how few data points we had for other clusters, we were only confident in claims we would make about installations in cluster one, since we had so many more data points. Cluster one is defined by predominantly higher consumption than the other two clusters, typically

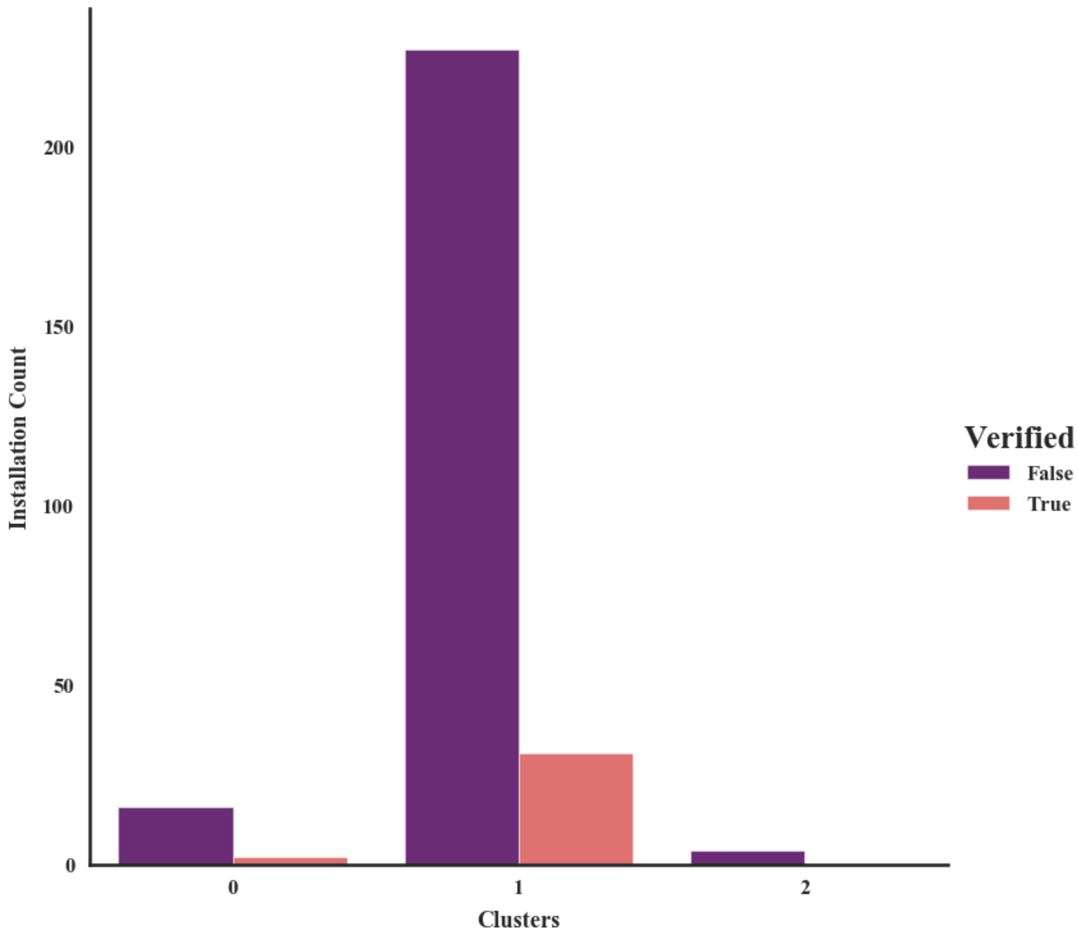

Figure 5-4: Number of installations in each cluster, broken down by verification status

higher cost, and a wide range of fuel oil consumption (see Figure 5-7). Additionally, cluster one represents a very strong majority. After selecting only the installations with outage data, we had 26 total installations, representing just under 10% of total installations in cluster one.

Once we ran the strain experiments, we inspected the cost and downtime of the installations under strain (see Figures 5-8 5-9). Comparing them to the "standard run" where the installation is not under strain, some of the installations were only seeing marginal changes in costs or downtime under strain. However, there was a shifting in the distribution and medians for most of metrics from installations under the strains.

After evaluating the distributions of differences in metrics from the standard ex-

	no normalization	log scale normalization	standard deviation normalization	z score normalization	simple mean imputation (no normalization)
mean data change during imputation	0.086	-0.159	0.089	0.096	0.086
standard deviation of data change	0.142	0.318	0.151	0.162	0.142
percentage of imputed installations passing tests	32.8	0.0	23.1	3.2	1
number of iterations before stopping condition reached	126	432	309	343	NaN

Table 5.2: Imputation performance on verification tests

periment to each strain experiment, (see Figures 5-10 and 5-11) we determined that the distributions are not normal. Because the distributions were not normal, we selected and then ran the Wilcoxon signed-rank test for each metric on each of the strain results, comparing it to the standard run for that same installation. For the majority of the strain scenarios, we rejected the hypothesis that the strain scenario results came from the same distribution as the standard run for at least a $p < .05$ (see Table 5.3). The strain scenarios where we failed to reject the hypothesis were all in the no outage change group and all except for one were measuring the change in downtime. It would make sense that energy price changes and sunniness changes would not make a huge difference on downtime, given that relatively few installations rely on solar power during grid outages and cost changes alone should not impact downtime substantially.

Overall, the price changes and outage increases have the biggest impact on predicted expenditure. Additionally, when low energy price changes co-occurred in a strain experiment, the average price change from the standard strain was negative, meaning that the cost went down. In terms of changing the predicted downtime, the biggest factor was the rate of outages. While sunniness changes affected costs significantly, it did not affect downtime significantly, suggesting that currently the biggest impact of solar power in these installations is in reducing costs, rather than in backup power generation.

We totalled the expected expenditure and downtime from each strain across the 26 installations we analyzed. The data is all in Table 5.4, but we highlight a few values here. The expected total expenditure of a high outage increase alone is 20 million dollars more than the standard expected expenditure, just for these 26 installations. A high energy price change alone is 126 million dollars more than the standard expenditure, representing a 6% change in expenditure overall. Across the 26 installations, there would be almost a weeks worth more hours of downtime than in the standard scenario under a high outage increase.

Strain Scenario	Mean Percent Change in Downtime	Level of Significance of Δ in Downtime	Mean Percent Change in Expenditure	Level of Significance of Δ in Expenditure
standard run (no strain)	0.00	NaN	0.00	NaN
low energy price change	-8.53	not significant	-3.62	***
high energy price change	-5.79	not significant	4.74	***
less sunniness	-2.29	not significant	-0.00	not significant
more sunniness	-3.81	not significant	0.03	*
standard run (moderate outage increase)	39.20	**	0.93	***
low energy price change (moderate outage increase)	43.91	***	-2.86	***
high energy price change (moderate outage increase)	43.96	***	5.87	***
less sunniness (moderate outage increase)	39.94	***	0.95	**
more sunniness (moderate outage increase)	43.69	**	0.90	**
standard run (high outage increase)	69.46	***	2.50	**
low energy price change (high outage increase)	65.20	***	-1.59	***
high energy price change (high outage increase)	65.92	***	7.73	***
less sunniness (high outage increase)	72.90	***	2.48	**
more sunniness (high outage increase)	64.95	***	2.48	***

Table 5.3: Aggregate statistics of strain experiment results.

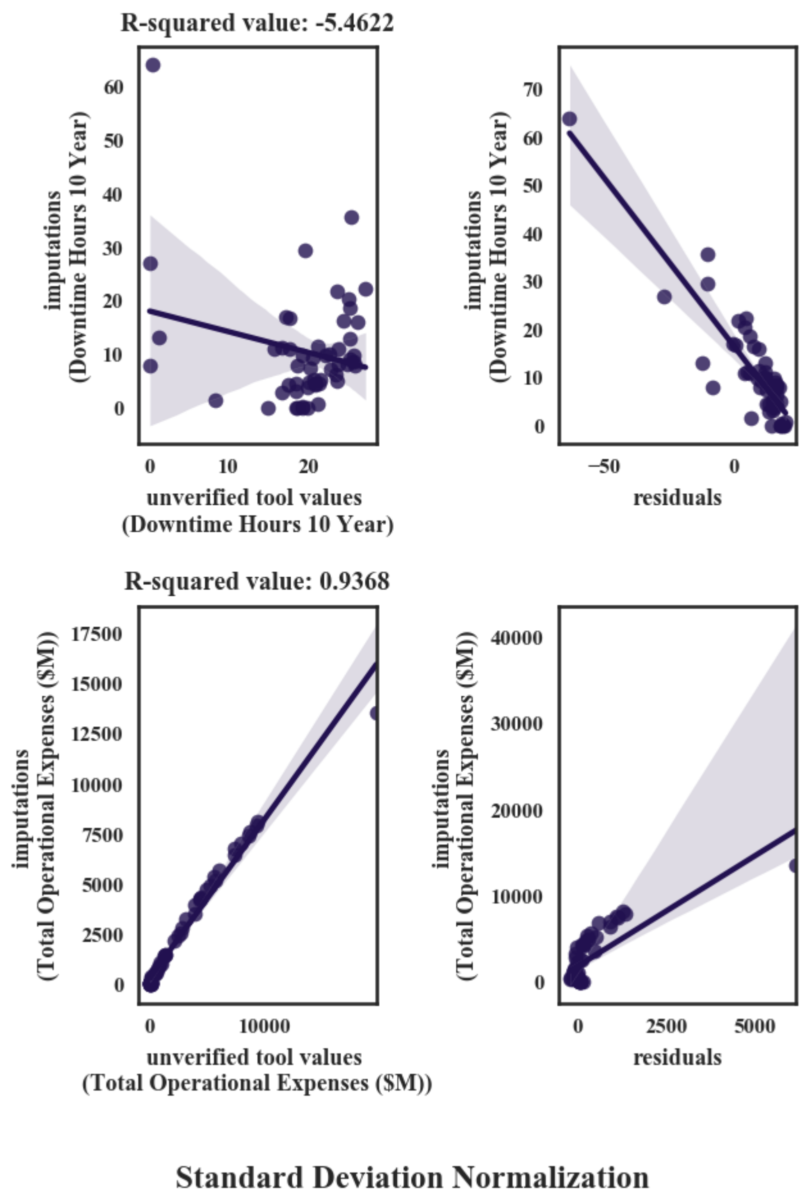

Figure 5-5: Parity and residuals plots for evaluating the performance of the imputation on data with normalized by standard deviation. The truth values are the ERA tool results from running the ERA tool on the imputed data

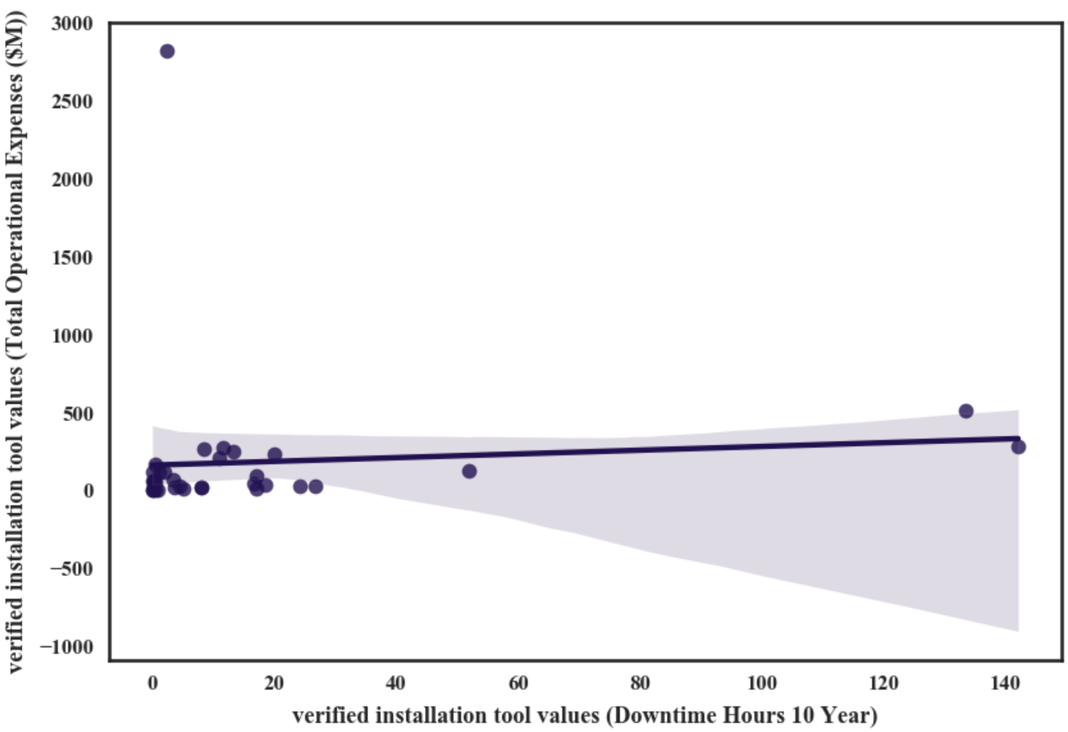

Figure 5-6: Plot of correlation between ERA tool predictions of downtime and expenditure for verified installations

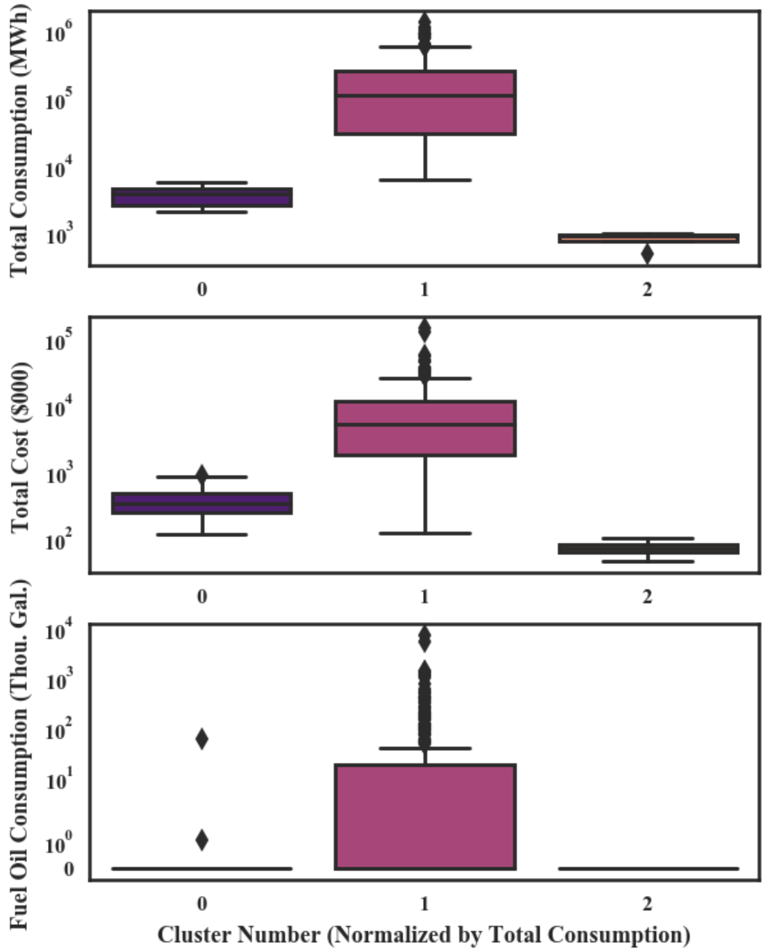

Figure 5-7: Distribution of top 3 defining features of the clusters from the clustering run on the data that was normalized by total consumption

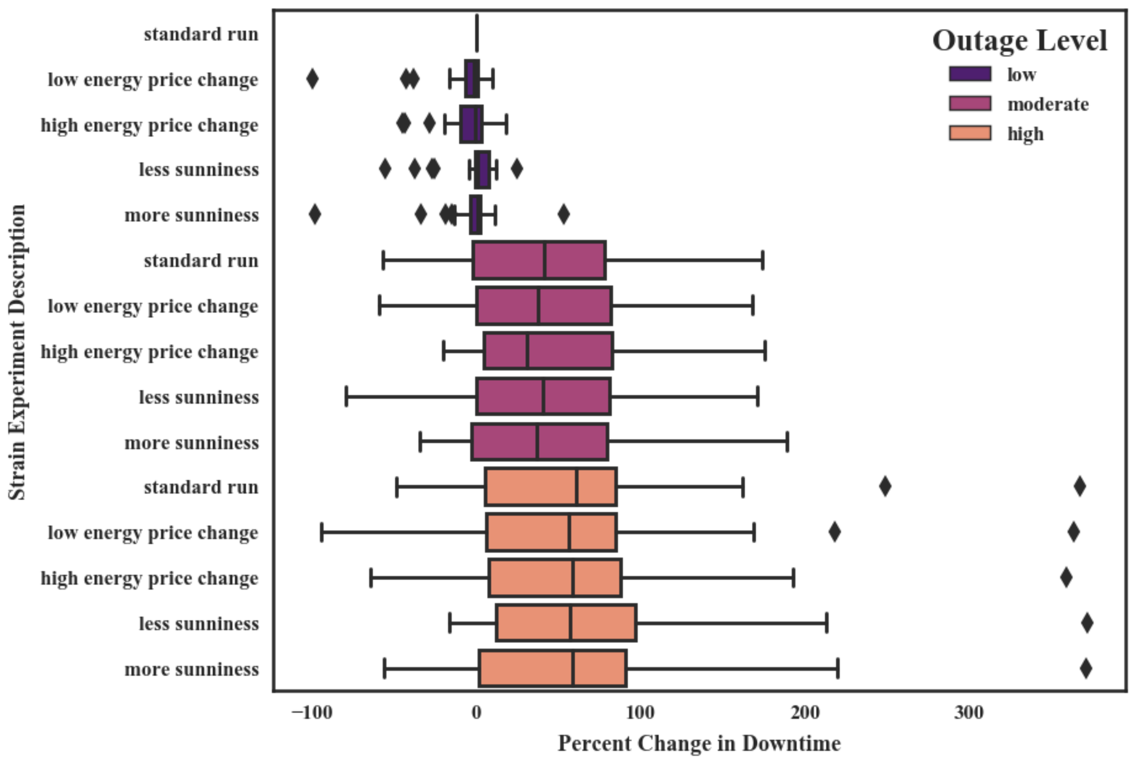

Figure 5-8: Box plot of the percent change in expected downtime for each strain experiment from the control ERA tool run

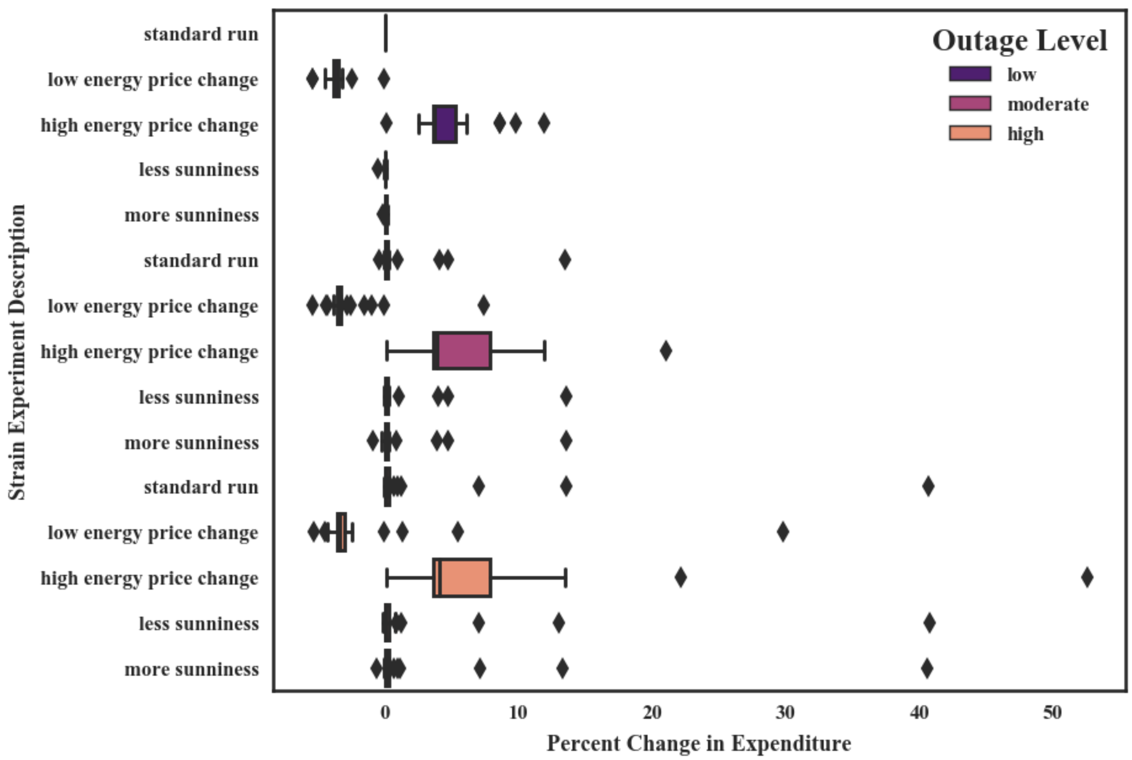

Figure 5-9: Box plot of the percent change in expected expenditure to administer the installation for each strain experiment from the control ERA tool run

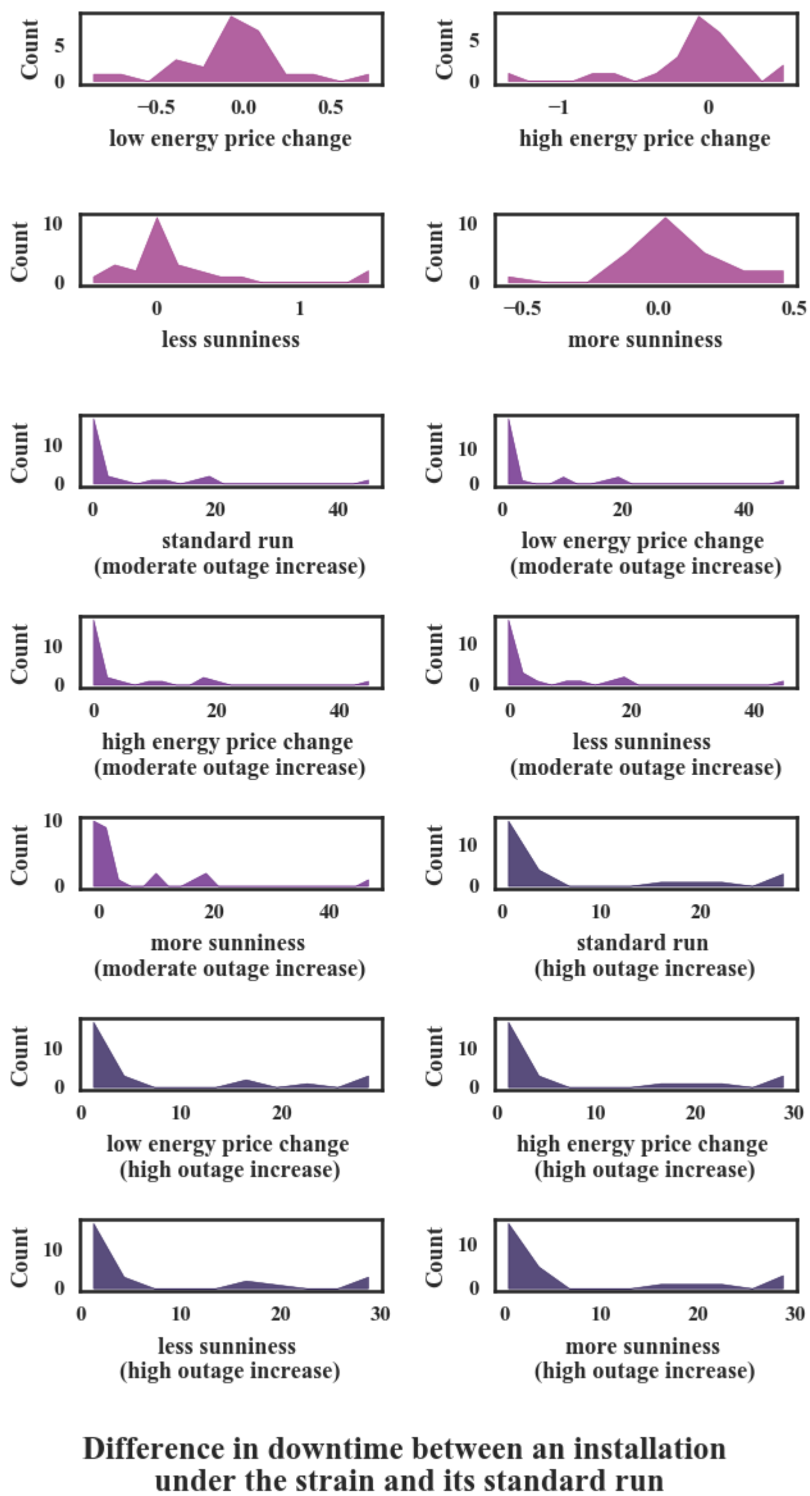

Figure 5-10: Distributions of the change in downtime of the installations from the control run to the strain experiment run in the ERA tool, separated by strain experiment

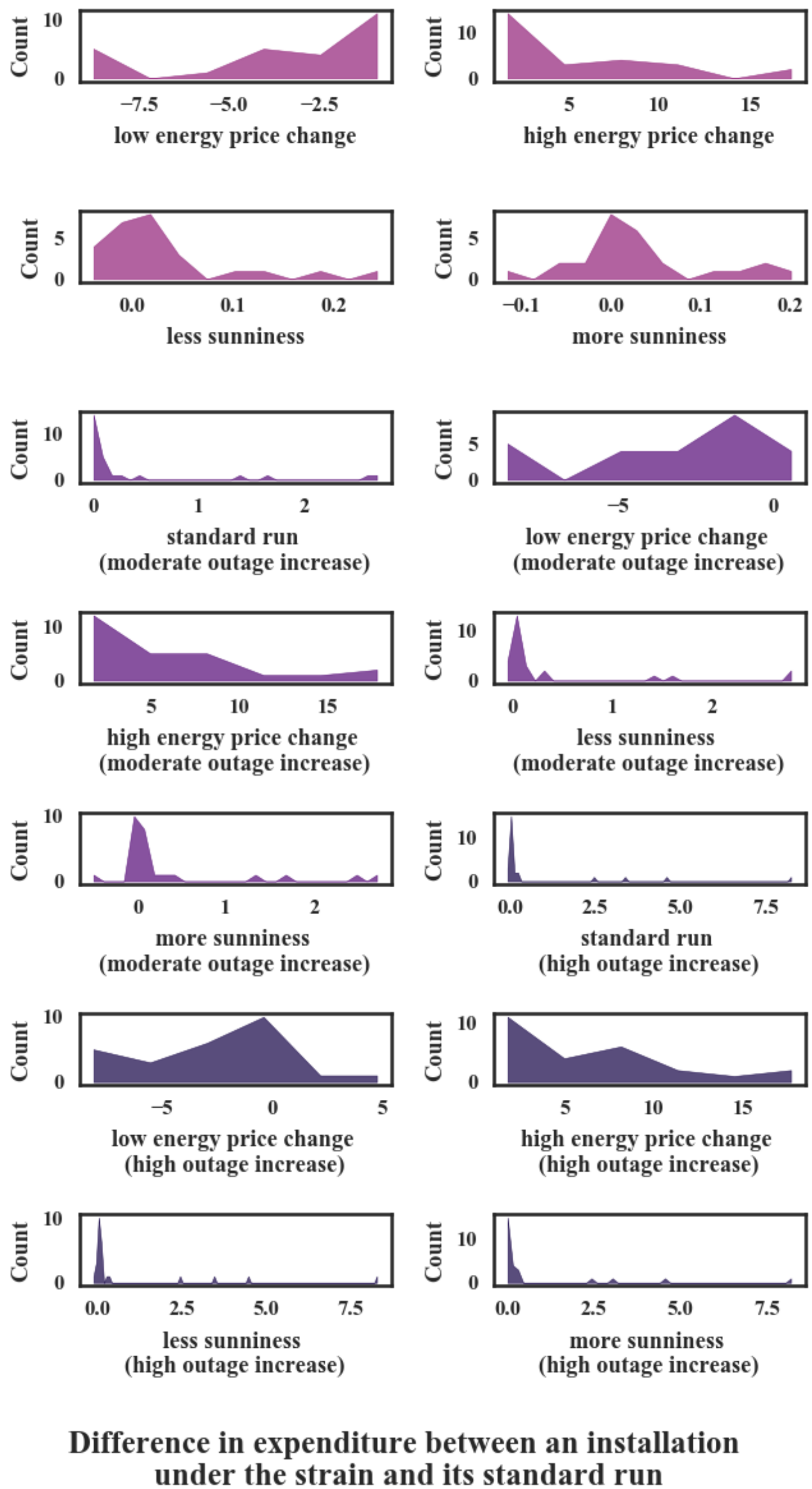

Figure 5-11: Distributions of the change in expenditure from the control run to the strain experiment run in the ERA tool, separated by strain experiment

	Total Downtime Across Tested Installations	Total Δ in Downtime (hours) Across Tested Installations	Total Expenditure Across Tested Installations	Total Δ in Expenditure ($M) Across Tested Installations
standard run	216.0	0.0	5054.0	0.0
low energy price change	214.0	-2.0	4968.0	-86.0
high energy price change	214.0	-2.0	5180.0	126.0
less sunniness	220.0	4.0	5055.0	1.0
more sunniness	217.0	1.0	5055.0	1.0
standard run (moderate outage increase)	346.0	130.0	5064.0	10.0
low energy price change (moderate outage increase)	350.0	134.0	4975.0	-79.0
high energy price change (moderate outage increase)	348.0	132.0	5191.0	137.0
less sunniness (moderate outage increase)	344.0	128.0	5064.0	10.0
more sunniness (moderate outage increase)	346.0	130.0	5063.0	9.0
standard run (high outage increase)	379.0	163.0	5074.0	20.0
low energy price change (high outage increase)	375.0	159.0	4983.0	-72.0
high energy price change (high outage increase)	382.0	166.0	5203.0	149.0
less sunniness (high outage increase)	379.0	163.0	5074.0	19.0
more sunniness (high outage increase)	378.0	161.0	5074.0	20.0

Table 5.4: Totals across the 26 installations in the strain experiment

Chapter 6

Conclusion

6.1 Summary

In this thesis, we compiled a partially complete data set of military installation energy systems and identified underlying structures in it using clustering. Also, we formulated a method for handling the large amount of missing data in our data set and then implemented that imputation methodology. At the same time, we ran the complete parts of our data set through simulations to evaluate the effects of climate change on the U.S. military's backup power system.

The results of the different clusterings that we generated showed us that most military installations are similar in energy use and consumption, with a handful of extreme outliers. The imputation methodology shows promise, as it predicted one of our two metrics very well. Given better validation tests and enough data to run leave-one-out testing, it could be a great tool for approximating the energy systems on installations. Finally, almost all the strain experiments we ran had statistically significant effects on both the expected downtime and expenditure of backup military power systems in our simulations.

6.2 Future Work

Stronger reporting standards and better naming conventions in the DoD could improve opportunities for energy usage and resilience analytics. With a consistent naming scheme, data sets could be combined more easily. Additionally, with more reported data, we could gain more insight into the performance and possibilities for each installation.

There are many new high performance sparse matrix completion methods that could greatly improve the imputation performance. Similarly, we had only about 10% of the data set as training data. Given more data, we could improve, and more extensively test, the imputation results. With either of these improvements, the imputation methodology could enable resilience surveys across domestic DoD installations using the ERA tool. Since the ERA tool evaluates many potential architectures, rather than just the current one, the DoD could investigate the long term savings from different renewable energy sources across all domestic military installations using the imputed results.

Additionally, these imputation methods could be used for other incomplete DoD data streams. For example, a highly relevant area of investment currently is protecting against power-related cyber attacks. One of the first steps in preventing such attacks is understanding the inventory and condition of networked power components. Often, these data sets are incomplete, hindering or stopping analysis. With imputation, we could build out more complete data sets of configurations and installed components, which are crucial for identifying cyber-security weaknesses.

The ERA tool currently offers black sky predictions, where it tests the effects of a multi-day outage, along with the standard outages. The strain scenarios we developed could be added to the ERA tool as an option, along with the black sky predictions. Building strain simulations into the tool would allow stakeholders across the military to view the effects of potential economic and climate shifts on their installation's resilience. They could also see if changing prices or outage rates affect their decisions as they're deciding on potential architecture changes.

6.3 Potential Impact of this Work

The DoD wants to be able to make holistic and informed investments and decisions about energy infrastructure. In order to make those decisions, the DoD needs information about the configuration and performance of energy systems across all installations managed by the department. However, at present, only a small, non-representative group of installations have the complete power system data needed to inform these decisions and investments.

Part of this thesis addressed this lack of usable data. Most of data that the DoD, and therefore MIT Lincoln Laboratory, has access to is self-reported data without quality assurance. So, we built tests to verify the validity of installation power systems data. Some data needed for trend and reliability analysis was missing and we had no process for finding the data. In order to solve this problem, we used an imputation methodology to infer the missing installation power system data. We showed that the methodology is reliable by running comparisons between the imputed results and simulated results, where there were high correlations for one of our two metrics. Finally, DoD installations are unique and are difficult to make generalizations about, especially from the few data points we had. Using one of the few complete, quality assured data sets we had access to, we identified clusters in the installations. We used these clusters to characterize the types of installations that our data set had verified data for.

In short, we have developed multiple avenues (clustering, self reported data verification tests, and imputation) for characterizing and generalizing from these limited data sources. Next, we showed how these larger data sets could be used in analytics.

We focused on the potential future effects from current trends in economics, weather, and infrastructure reliability. Using the data set we created, we were able to predict that at current rates of change from these trends, there would be more than 5 additional days of power outages and 10 million dollars in additional expenditure over 26 installations. If these results are representative of all of the installations in their cluster, we could see more than a month in additional total downtime and 100 million dollars in additional expenditure over all those installations.

As discussed in the future work section, we believe that this approach of analyzing potential strains to the cost and reliability of critical infrastructure is extendable beyond the work we have shown here. Evaluating the sensitivity of DoD systems to potential strains of any kind can inform better investment of limited resources in improving the military's energy infrastructure.

In the race to prepare critical infrastructure for climate change, better visibility and analytics is an important component.

Appendix A

Tables

	Cluster Number (Standard Normalization)	Number of Installations in Cluster
0	0	232
1	1	35
2	2	1
3	3	5
4	4	2
5	5	2
6	6	1
7	7	2

Table A.1: Cluster membership totals for standard normalization

	Cluster Number (Normalized by Total Consumption)	Number of Installations in Cluster
0	0	18
1	1	258
2	2	4

Table A.2: Cluster membership totals for normalized by total consumption

	Cluster Number (Normalized by Square Footage)	Number of Installations in Cluster
0	0	266
1	1	1
2	2	13

Table A.3: Cluster membership totals for normalized by square footage

	Number of installations with supplementary data added	percent of installations with supplementary data added
Total Number of Substations	94	38.1
Generator List	18	7.3
Verified Web Application Data	26	9.6

Table A.4: Amount of supplementary data added to installations with missing data

Appendix B

Figures

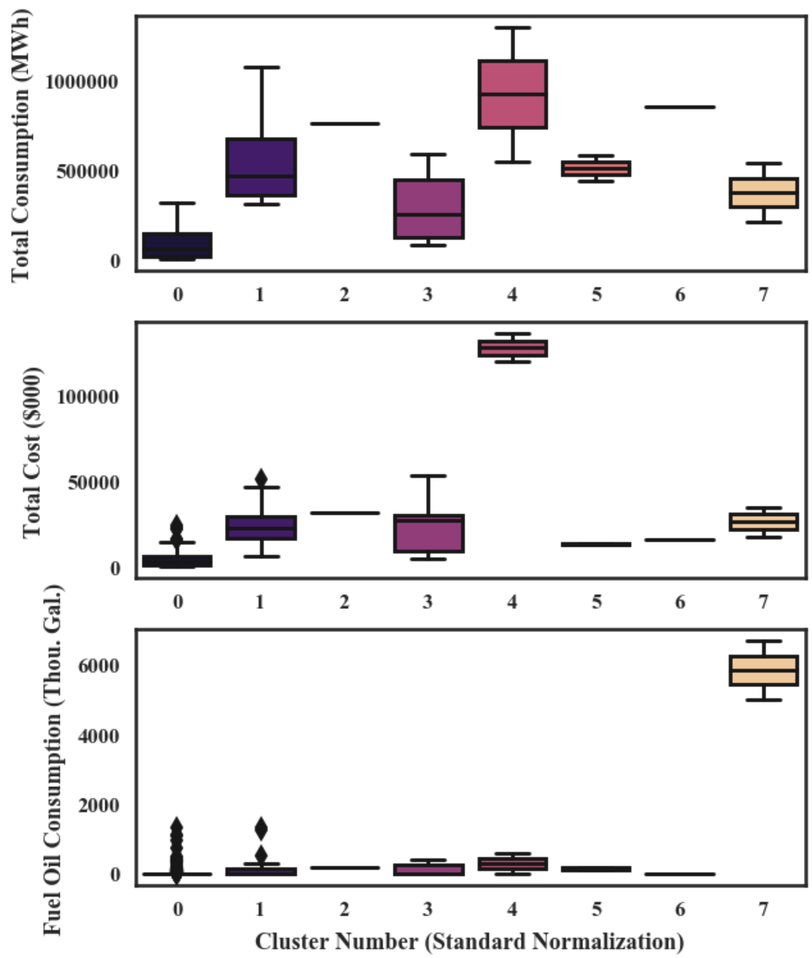

Figure B-1: Distribution of top 3 defining features of the clusters generated by clustering on data normalized using only the standard method

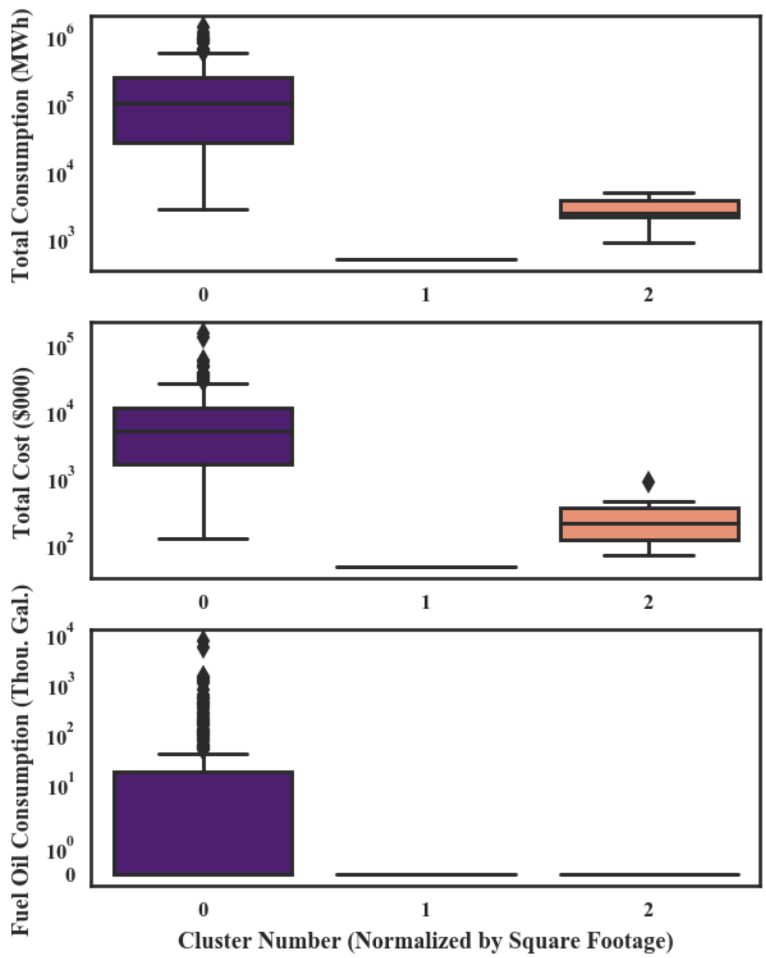

Figure B-2: Distribution of top 3 defining features of the clusters generated by clustering on data normalized using square footage and the standard method

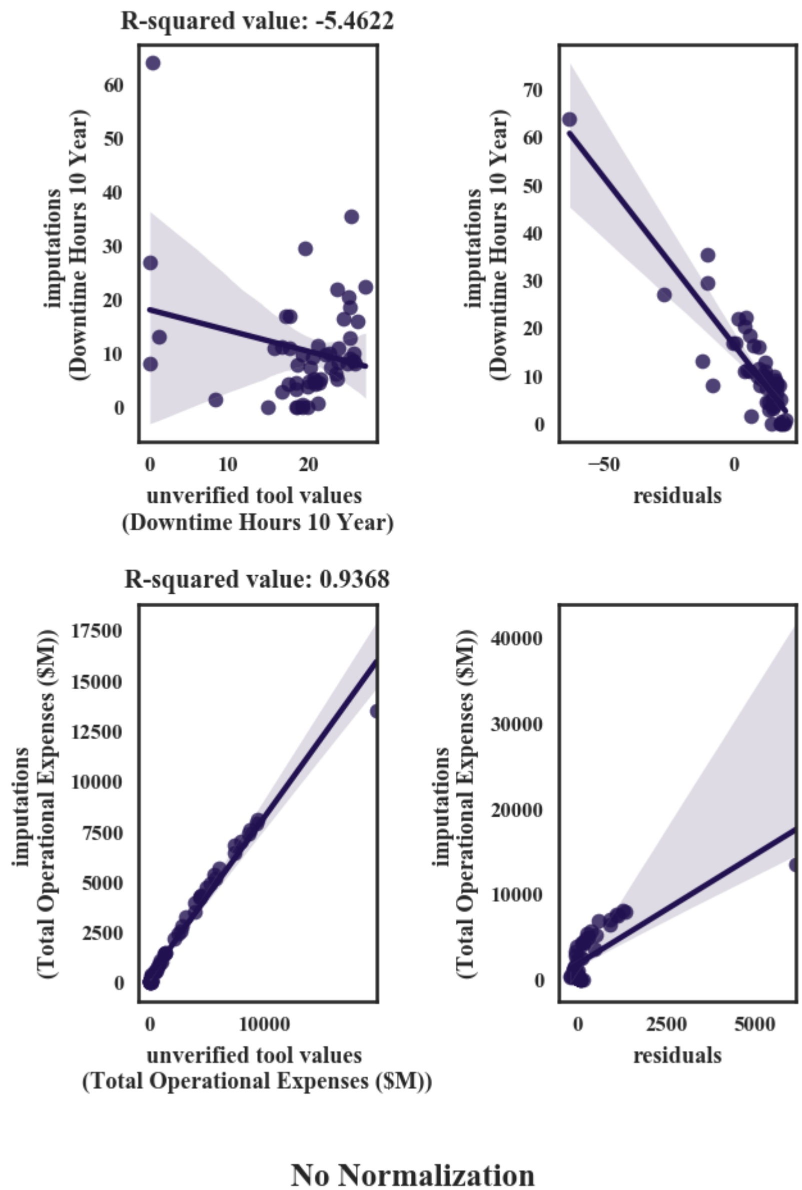

Figure B-3: Parity and residuals plots for evaluating the performance of the imputation run on data with no normalization

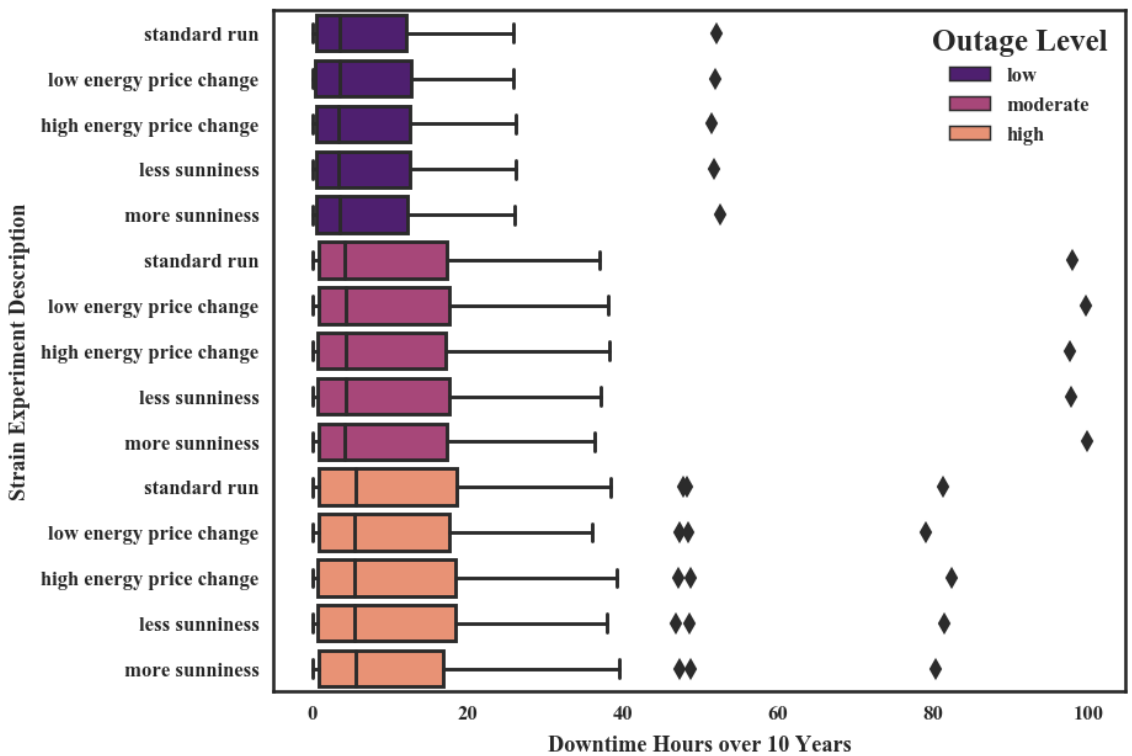

Figure B-4: Box plot of the expected downtime of the installation (in hours) generated by the ERA tool, separated by the strains the installations were run under

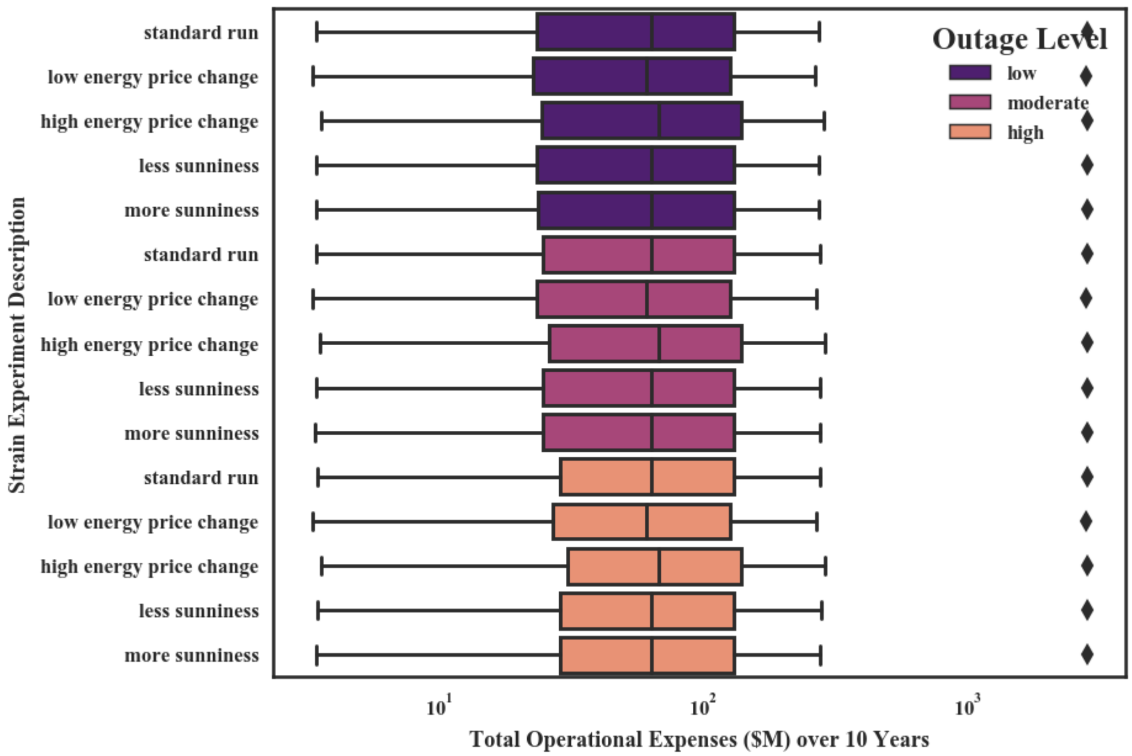

Figure B-5: Box plot of the expected expenditure to administer the backup power system on an installation (in millions of dollars) generated by the ERA tool, separated by the strains the installations were run under

Bibliography

[1] M. Clark, "Aging U.S. power grid blacks out more than any other developed nation," *International Business Times*, 2014.

[2] M. Ouyang, L. Dueñas-Osorio, and X. Min, "A three-stage resilience analysis framework for urban infrastructure systems," *Structural Safety*, vol. 36-37, pp. 23–21, May-July 2012.

[3] C. Curt and J.-M. Tacnet, "Resilience of critical infrastructures: Review and analysis of current approaches," *Risk Analysis*, vol. 38, no. 11, pp. 2441–2458, 2018.

[4] C. Piech, "K-means algorithm," *Stanform CS 22 Handouts*, 2013.

[5] P. Dangeti, *Statistics for Machine Learning.* O'Reilly, 2017.

[6] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2225–2236, 2010.

[7] G. Gazzola and M. K. Jeong, "Dependence-biased clustering for variable selection with random forests," *Pattern Recognition*, vol. 96, December 2019.

[8] M. Azur, E. Stuart, C. Frangakis, and et. al, "Multiple imputation by chained equations: what is it and how does it work?," *International Journal Methods Psychiatric Research*, 2011.

[9] S. Buck, "A method of estimation of missing values in multivariate data suitable for use with an electronic computer," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1960.

[10] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[11] Department of Defense, "Annual energy management and resilience report (AEMRR)," June 2020.

[12] U.S. Energy Information Administration, *About EIA.* 2021. From the EIA.gov website.

[13] U.S. Energy Information Administration, *Annual Energy Outlook 2020.* 2021. From the EIA.gov website.

[14] Z. Zeng, A. Ziegler, T. Searchinger, and et. al., "A reversal in global terrestrial stilling and its implications for wind energy production," *Nature Climate Change*, vol. 9, no. 12, pp. 979–985, 2019.

[15] Iberdrola, "How do photovoltaic plants work?," 2021.

[16] J. Yin, A. Molini, and A. Porporato, "Impacts of solar intermittency on future photovoltaic reliability," *Nature Communications*, vol. 11, no. 1, p. 4781, 2020.

[17] A. Kenward and U. Raja, "Blackout: Extreme weather, climate change, and power outages," *Climate Central*, 2014.

[18] U. Chrobak, "The U.S. has more power outages than any other developed country. here's why.," *Popular Science*, 2020.

[19] J. Wirfs-Brock, "Power outages on the rise across the U.S.," *Inside Energy*, 2014.