Title
Machine Learning to Scale Fault Detection in Smart Energy Generation and Building Systems

Author
Hu, Rong Lily

Publication Date
2016

Peer reviewed|Thesis/dissertation

Machine Learning to Scale Fault Detection in Smart Energy Generation and Building Systems

by

Rong Lily Hu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alice Agogino, Co-chair
Professor David M. Auslander, Co-chair
Professor David Brillinger

Fall 2016

Machine Learning to Scale Fault Detection in Smart Energy Generation and Building Systems

Abstract

Machine Learning to Scale Fault Detection in Smart Energy Generation and Building Systems

by

Rong Lily Hu

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor Alice Agogino, Co-chair

Professor David M. Auslander, Co-chair

Data-driven techniques that extract insights from sensor data reduce the cost of improving system energy performance through fault detection and system health monitoring. To lower cost barriers to widespread deployment, a methodology is proposed that takes advantage of existing sensor data, encodes expert knowledge about the application system, and applies statistical and mathematical methods to reduce the time required for manual configurations.

Renewable energy technologies as well as building energy management systems have upwards of hundreds of existing sensor data points used for control and monitoring. Furthermore, innovations in "Internet of Things" (IoT) devices have led to connected power meters, lights, occupancy sensors, and appliances that are capable of data collection and communication. This data presents a valuable opportunity to extract meaningful information and take data-driven action.

The motivation for transforming data from these devices into actionable information is to improve operations, monitor system health, increase energy generation, and decrease energy waste. The development and widespread use of energy conservation and renewable energy technologies are critical to minimizing negative environmental consequences. To that end, increasing profitability for users and decreasing costs of these technologies enables market penetration and widespread adoption. On the energy generation side, operations and maintenance accounts for up to 30% of the cost of wind generation, and unexpected failures on a wind turbine can be extremely expensive. On the energy demand side, commercial buildings consume 19% of US primary energy [1]. Of this, an estimated 15% to 30% of energy used in commercial buildings is wasted by poorly maintained, degraded, and improperly controlled equipment [42].

However, one cannot achieve scalable deployments of analytics and applications across systems if deploying solutions requires vendors and domain experts to install sensors and information technology infrastructures that require tailoring each solution for each deployment. Today, even well-established commercial offerings are not deployed at scale because

costs are prohibitive. Thus, a major challenge to scalability is reducing hardware and software installation costs, manual configuration requirements, and manual monitoring.

To address this challenge, a methodology is proposed that leverages machine learning techniques to configure automated fault detection systems and controls. The approach combines sensor data points and encodes engineering knowledge that is generic to the application system but independent of a particular deployment. The resulting data can be input into numerous machine learning and optimization algorithms. Furthermore, the procedure selects data points and demonstrates that only a small number of sensors are necessary for fault detection with high accuracy rates. Applications to a wind turbine, a commercial building chiller plant, and residential buildings demonstrate the proposed methodology. The results are implementable and realizable using off-the-shelf algorithms, libraries, and tools. The goal is to enable an application that can be written once and then widely deployed with little additional cost or effort. The results of analyses can also inform policy decisions for stakeholders.

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation would not have been completed without the guidance of Professor Alice Agogino, Professor David M. Auslander, and Professor David R. Brillinger. Professor Agogino played a vital part in my journey starting and completing my doctoral studies at Berkeley. She helped me navigate the procedures, ensure I was on-track, and combine my research interests. Professor Auslander taught me how to frame my research in the context of academia as well as the real world. Professor Brillinger challenged me to think across discipline boundaries. I would also like to thank Professor David Dornfeld and Professor Duncan Callaway for serving on my qualifying exam committee.

Many thanks to Dr. Jessica Granderson for mentoring me as an undergraduate student, for her encouragement in my doctoral pursuits, and for her support during my studies. This work would not have been possible without her. I would also like to thank Carla Trujillo for writing advice. From the teams I worked with at the Lawrence Berkeley National Laboratory, I would like to thank Marco Bonvini, Guanjing Guo, and Oren Schetrit. Thanks to Thierry Nouidui, Xiupeng Feng, and Liping Wang for help with Dymola, Modelica, and the Molecular Foundry model respectively.

I would like to thank my co-authors for permission to use reference [37] in Chapter 5. Thanks to Kevin Leahy for the data and for help understanding the wind turbine industry. Thanks to Ioannis Konstantakopoulos for feedback and knowledge of previous work. Thanks also to Professor Agogino, Professor Auslander, and Professor Costas Spanos.

I want to thank my co-authors for permission to use reference [39] in Chapter 7. Thanks Professor Yinyu Ye for giving me the opportunity to explore the topic. Thanks to Robert Entriken for insight into the regulatory and utility industries. Thanks to Ryan Skorupski for running the experiments and extensions. Thanks to Nicole Taheri for initial onboarding help.

My thanks go to Aaron Culich and Berkeley Research Computing, and to Drew Sabelhaus for help and access to computing power.

This work was supported in part by US NSF ACI-1541349 and OCI-1246396, the University of California Office of the President, the California Institute for Telecommunications and Information Technology (Calit2); and the Berkeley Research Computing Program at the University of California, Berkeley.

The numerous professionals and experts I connected with over the course of this dissertation provided valuable background information. I appreciate their time in speaking with me.

I want to thank my family. The now senior Dr. Hu made me realize that attaining a PhD was possible. Learning about his journey and his career inspired me and shaped my perspective on professional life. My mother taught me the value of understanding mathematics decades ago and she has also been a constant source of emotional strength. My little sister has been a source of amusement and a reminder of my creativity and curiosity that initially motivated me in my studies. Most recently, my husband has both helped and challenged me in computation and coding while being a source of stability and emotional support. Lastly, thanks to my friends for making this multi-year endeavour fun, enjoyable, and memorable outside of research.

Chapter 1

Introduction

1.1 Chapter Overview

This introduction chapter summarizes the motivation for the research, key research contributions, the organization of chapters in the dissertation, and defines terminology.

1.2 Motivation

The development and widespread use of energy conservation and renewable energy technologies are critical to minimizing negative environmental consequences from climate and for environmental sustainability. To that end, increasing profitability for users and decreasing costs of these technologies enables market penetration and widespread adoption. Data-driven techniques that extract insights from sensor data reduce the cost of improving system energy performance through fault detection and system health monitoring. The motivation for transforming data from these devices into actionable information is to improve operations, monitor system performance and health, increase energy generation, and decrease energy waste.

On the energy generation side, operations and maintenance accounts for up to 30% of the cost of wind generation, and unexpected failures on a wind turbine can be extremely expensive. On the energy demand side, commercial buildings consume 19% of US primary energy [1]. Of this, an estimated 15% to 30% of energy used in commercial buildings is wasted by poorly maintained, degraded, and improperly controlled equipment [42]. The median whole building energy savings is 16% and just 13 of the most common faults in U.S. commercial buildings in 2009 were believed to have caused over $3.3 billion of wasted energy [51]. Much of this waste can be prevented with fault detection and diagnostics (FDD). Therefore, an enormous opportunity for energy savings is in monitoring of energy use and continuous fault analysis. However, for continuous fault analysis, manual fault detection can be tedious, making automated processes of fault detection desirable. Thus, there is great opportunity in the integration of machine learning to automate fault detection of grid distributed demand and generation resources.

Tremendous challenges remain along the pathway of widespread integration of techniques – such as machine learning – for the analysis of massive data sets to better operate energy systems. These challenges include i) the high cost of distributed sensors and manual configuration for data analysis techniques to individual systems and ii) methods to encode first principles engineering knowledge of energy systems into data analytics. One cannot achieve scalable deployments of analytics and applications across systems if deploying solutions requires vendors and domain experts to install sensors and information technology infrastructure and tailor each solution for each deployment. Today, even well-established commercial offerings are not deployed at scale because costs are prohibitive. Thus, a major challenge to scalability is reducing hardware and software installation costs, manual configuration, and manual monitoring.

The goal of this dissertation is to lower the barrier of adopting energy analytics to improve energy efficiency and reliability by addressing the challenges above for energy generation units such as wind turbines as well as energy demands in buildings. To this end, a methodology is proposed that leverages machine learning techniques to configure automated fault detection systems and controls. The approach is to convert existing data off of energy control systems into a form suitable for off-the-shelf machine learning algorithms and to encode engineering knowledge of energy systems into this input. This is necessary because machine learning algorithms traditionally assume independent samples of one variable type, while energy systems data is time series data, that includes both continuous and binary variables, contains correlated or irrelevant variables, and has other challenges.

Renewable energy technologies as well as buildings have upwards of hundreds of existing sensor data points used for control and manual monitoring. Furthermore, innovations in "Internet of Things" (IoT) devices have further led to connected power meters, lights, occupancy sensors, and appliances that are capable of data collection and communication. This data presents a valuable opportunity to extract meaningful information and take data-driven action.

In the field of energy management, the current trend of increasing integration of Information and Communication Technology has enabled market actors to develop technologies that has engendered the electric grid as a cyberphysical system — the smart grid. This enables the development and use of applications such as the smart meter, bidirectional communication, advanced metering infrastructure (AMI), home automation, and home area networks [44]. Measurement and recording of electricity consumption and two-way communication between the meter and the utility's system can help adjust energy consumption patterns, and thereby achieve economic benefits.

The electric grid as a cyber-physical system results in the creation of large data sets, including for example, electric meter readings, electricity end-use consumption, electricity price signals, consumer preferences, and device control signals, among others. Using this data, big data methods such as mathematical programming can manage electricity demand more efficiently and assess the likely sources of value to be derived from smart grid technologies under electricity programs such as demand response.

To meet energy needs on the grid, demand response is a valuable resource because it can help reduce the volatility of electricity prices, mitigate market power of generators, and enhance grid reliability. Demand response achieves these objectives by lowering the peak demand for energy, which reduces the need to construct new and expensive generation units, and by providing ancillary grid services such as regulation and reserves to reliably integrate variable resources such as renewable generation [23]. In 2013, the potential contribution of demand response resources in the United States was 28,798 MW in Regional Transmission Organization (RTO), Independent System Operator (ISO), and Electric Reliability Council of Texas (ERCOT) markets [22], which represents an increase of 5.9% since 2009 [23]. Note that these values have fluctuated recently, due to economic impacts on electricity consumption. Demand response technologies can be and are considered as components within a broad range of energy supply scenarios [23]. Correspondingly, there are significant policy implications associated with the development of demand response technologies (DR technologies).

While DR technologies are already prevalent in the commercial and industrial sectors of the U.S. economy, the market for DR technologies in the residential sector is currently in a nascent stage [23]. Accordingly, this research examines the potential economic benefits associated with the application of DR technologies in the residential sector. Specifically, this dissertations offers (1) an optimal load shifting (OLS) model that can manage electricity demand more effectively; and (2) an assessment of the most likely sources of value to be derived from DR technologies.

The desired outcome is to enable the economic integration of energy analytics to improve efficiency and reliability while deepening the larger research community's knowledge of applied statistics and mathematics for energy analytics.

1.3 Research Contributions

The key dissertation research contributions are:

- A framework for expanding control system sensor measurements with additional features derived from expert knowledge that is general to the type of system and independent of the individual installation. In the wind turbine fault detection literature, for example, fault detection methods either focus on using control systems data, or focus on using a few selected features for detection of (usually specific) faults in specific system components. There does not exists a public encoding of general expert knowledge for openly available measurements for these systems. This knowledge encoding creates new virtual sensors, virtual in the sense that they do not necessarily correspond to physical sensors.

- A methodology that automatically recommends data points to use for fault detection from among all sensors and virtual sensors without the use of knowledge unique to the individual deployment.

- A procedure that uses the expanded data set of both sensor measurements and derived features as variables in the fault detection problem framed as a statistical learning problem. This allows for the use of tools and knowledge developed in the statistical learning community, such as data driven feature selection and machine learning algorithms, to address the problem of fault detection.

- A model that leverages optimization algorithms to enable positive return on investment by managing electricity consumption under time varying electricity prices. Importantly, the model has a modular design that is flexible and can be configured to accommodate added capabilities. Its quick performance is desirable for real-time dynamic management.

- A smart grid valuation framework that can be used to interpret the results produced by the model with respect to the efficiency of smart appliances and their respective prices.

1.4 Organization of the Dissertation

This dissertation begins with an overview of fault detection, wind turbine maintenance, building maintenance, building energy use, and demand response. Next, relevant methodologies for fault detection in these domains are proposed. Then, the proposed methodologies are demonstrated on data, with description of the data collection, results, and analysis. Finally, this dissertation concludes with findings, recommendations, and future work. The breakdown by chapter is as follows:

Chapter 1 summarizes the motivation for the research, the terminology used, and the organization of chapters.

Chapter 2 describes background and academic literature in fault detection, wind turbine maintenance, commercial HVAC maintenance, home energy management systems, and industry practises as well as relevant statistical and mathematical concepts.

Chapter 3 presents the proposed methodology on using applied statistics and applied mathematics to improve energy performance.

Chapter 5 demonstrates the methodology on fault detection in wind turbines.

Chapter 6 illustrates the performance of the methodology by detecting faults in commercial building chiller plants.

Chapter 7 provides application of managing energy demand in residential buildings by scheduling energy use.

Chapter 8 summarizes the results of the research, presents findings and recommendations, and identifies avenues for future research.

1.5 Terminology

In the domain areas that intersect with this work, terms may have different meanings. Conversely, different terms may be used to refer to the same concepts, or have different degrees of overlap. Furthermore, terms from one domain area may be foreign to another domain area. To avoid possible confusion, the following terminology will be employed in this study, with some attempt at identifying potential confusion.

Fault detection and the verb **detect** [a fault] refers to indicating the occurrence of a fault at time t given a data set at time t. **Fault prediction** and the verb **predict** [a fault] refers to determining the occurrence of a fault at some future $t + \epsilon$ given a data set at time t.

A potential for misunderstanding is that in the statistical prediction field, the verb "predict" need not be in the future. The statistical prediction framework is "given X, predict Y", where X is some variable(s) and Y is some other variable, and often both X and Y are independent of time. For example, a classic problem is given X is an email, predict Y that email is spam or not. Thus, someone from a statistical learning background may interpret given X data from a wind turbine, "predict" Y the wind turbine has a fault during that time, while someone from a fault detection or wind turbine background may interpret the use of the term "predict" to mean anticipating a fault in the future.

Measurements are collected from sensors and are measured quantities. In different fields, other possible terms one may come across are "process data", "data points", "sensor data points", "variables", and "parameters".

Variables will refer to the "raw" input variables. Variables can include but are not limited to measurements. In the literature, similar terms are "measurements", particularly when the measurements are the variables, "features", or "parameters".

Features will refer to variables constructed from the "raw" input variables. For example, a feature could be a linear combination of variables. Other similar terms are "variables" and "parameters".

Given the previous definitions, **variable selection** is choosing from among the set of variables. **Feature selection** is constructing and selecting features, as there are an infinite set of possible features.

In the fault detection field and in application domains such as wind turbines, HVAC, and chillers, variable and feature selection commonly involves using expert knowledge. In the machine learning field, variable and feature selection commonly involves using mathematical and statistical techniques, as in [32]. As a result, the terms "variable selection" and "feature selection" may carry those connotations in those respective fields and have assumptions about the selection methodology.

Furthermore, while there are many potential benefits of variable and feature selection — such as facilitating data visualization and data understanding, reducing storage requirements, reducing computation times, etc. — the purpose of variable and feature selection in this study is to construct and select subsets of variables and features that are useful for automated fault detection.

Chapter 2

Background and Literature Review

2.1 Chapter Overview

The reviewed literature is organized into the following sections for the following reasons:

- Fault Detection - this section covers general fault detection techniques, drawn from the areas of process engineering, chemical engineering, performance and health management, etc.
- Fault Detection in Wind Turbines - wind turbines are examples of high-cost energy systems, which is the target application of this study. Wind turbines are used as a demonstration application.
- Fault detection in Building Systems, Chiller Plants, and Chillers - chiller plants are examples of high-cost energy systems, which is the target application of this study. Chiller plants are an example of a building energy system. Within a chiller plant, one of the most important pieces of equipment is the chiller.
- Overview of Relevant Applied Statistics and Mathematics - this section briefly introduces relevant concepts in statistics and mathematics that are used in the proposed methodology. These concepts include machine learning, optimization, time series, and variable and feature selection.

2.2 Overview of Fault Detection

Important terms are defined below from [31]. A fault is another word for a problem in a system. More precisely, a **fault** is an unpermitted deviation of a least one characteristic property or parameter of the system from acceptable, usual, or standard condition. **Fault detection** is indication of a presence of a fault in a system and time of detection. Following fault detection is **fault isolation**, which is the determination of the kind, location, and time

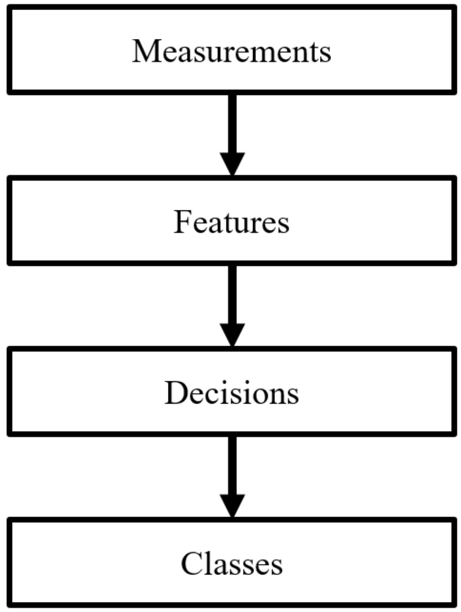


Figure 2.1: Transformations undergone for fault detection and diagnostics

of a fault. **Fault identification** is determination of the size and time-variant behavior of a fault. fault diagnosis, which follows fault isolation and identification, is determination of the kind, size, location, and time of a fault.

A review of fault detection and diagnosistics for processes can be found in [72] [70] [71]. The first paper [72] in this three part series of papers describes a common set of criteria with which to compare and evaluate fault detection and diagnostic methods and provides and describes a taxonomy for the methods.

The fault diagnostic process can be viewed as a series of transformations on measurements, illustrated in Figure 2.1, adapted from [72]. [72] describes the measurement space as the space of measurements with no a priori problem knowledge which is input into the diagnostic system. The feature space is obtained as a function of the measurements by using a priori knowledge about the problem. The decision space is made up of decision variables. The mapping from the feature space to the decision space is usually designed to meet an objective function such as minimizing misclassifications, that is, false positives and false negatives. The class space indicates categorically to which failure class a given measurement pattern belongs. [72] notes that the transformations from feature space to decision space, and from decision space to class space are performed using either threshold functions, template matching, or symbolic reasoning. However, in this dissertation, concepts from machine learning are used.

Methods are categorized as qualitative, quantitative, or based on process-history, with each paper in the series reviewing one of the major categories. Qualitative and quantitative methods are subsets of model based methods. Model-based methods are usually developed based on some fundamental understanding of the physics of the process. Quantitative methods express this understanding in terms of mathematical relationships while qualitative methods express this understanding qualitatively. While quantitative and qualitative methods are model-based approaches that require a priori knowledge, process-history based methods only need a large amount of historical process data.

The process-history based methods covered in [71] include qualitative process-history methods and online methods such as process control charts. Also discussed are principle component analysis (PCA) which transforms a number of related process variables to a smaller set of uncorrelated variables which are linear combinations of the original variables, partial least squares (PLS) which extracts latent variables that not only explain the variation in the process data but also that variation in the process data that is most predictive of the process quality data. The paper also touches on wavelets, and some statistical classifiers such as Parzen windows, mixtures of multivariate gaussians. Also discussed are Bayes classifiers, neural networks, and clustering algorithms.

There also exist review papers that survey specific categories of fault detection, such as supervisory methods, model based techniques, and trends in applications of model based techniques [40]. Another review paper surveys fault diagnostics with multivariate statistical models [80]. A further paper comments on the era of "big data" in the context of analyzing process history data [57].

2.3 Fault Detection in Wind Turbines

Wind Turbine Manufacturer

Wind turbine manufacturers not only sell wind turbines but also offer wind turbine warranties, extended warranties, and maintenance and service plans. A typical service contract is five years. While under warranty, the manufacturer conducts almost all the scheduled and unscheduled maintenance. Outside of warranty, the manufacturer conducts about 80%, while third parties, known as independent service provides (ISP) account for about 20%.

Personnel Training

Wind turbine installers and maintenance technicians and engineers who work for the manufacturer are trained in-house. The background of new incoming personnel varies and can include no experience,or graduates of a wind school program from a technical college, or some experience working on a wind turbine, or some experience in maintenance. Once hired, they take classes in-house and follow the manufacturer's curriculum, which covers different levels and is required for promotion. The teaching material, which includes alarm code reports, troubleshooting procedures, and maintenance manuals, are internal to the manufacturer. This training curriculum ensures that the personnel act the same and follow the same procedures and standards across visits to different sites and also when different personnel visit the same wind turbine site.

Data in Wind Turbines

Each wind turbine has its own server which collects data and stores data from the individual wind turbine. Furthermore, each wind turbine has a dedicated optical fiber data line from the wind turbine to on-site servers which collect data across the wind farm and sends data out to, for example, the manufacturer's data center. If connectivity breaks down along any of these fiber lines, data will be sent when reconnected. All this fiber may be expensive, but the cost of fiber is not so expensive when compared to the cost of construction of roads to access the wind farms and wind turbines or the cost of the electrical hookup to the electricity grid.

Manufacturers' Fault Detection

Data

Fault detection solutions provided by manufacturer's use 10 minute aggregated data sent from individual wind turbines and farms to the manufacturer's data center. This 10 minute data is usually the average over 10 minutes but can also include the min and max values over 10 minutes. This 10 minute aggregated data includes about 200-500 data points. The trend is towards increasing the number of data points on wind turbines. Older wind turbines have approximately 100 data points while newer ones have approximately 600 data points. This data is overlaid with event data logs and the data is stored in a master database, which can be a SQL database.

The event data is a log with new entries created when a change occurred. The event log has precision of up to one second. For an event, the log also includes some related values of interest at that time. For example, if an inverter overheats at 8:32:40, the temperature of the inverter is also reported at that timestamp.

For the most part, from the manufacturers' point of view, it is sufficient to use 10 min data. There is because there are very few phenomena that happen instantaneously, for example, due to the thermal inertia. Their belief is that there isn't much need to see within every 10 minute sample, because if a failure fails that fast, there is no time to react to it anyway.

For customers, the basic option is to allow customers to download SCADA data. SCADA is the Supervisory Control and Data Acquisition system, an industrial computer system that monitors and controls the building. Depending on the contract between the customer and the manufacturer, it is also possible to stream the data, or store it in a database accessible to the customer.

Fault Detection Method

Most fault detection is based on rules, thresholds, outlier detection, and anomaly detection.

For example, for most faults, the 1 second max, 10 second max, and 100 second max are checked with different alerts at each level. The faults are not instantaneously detected. For example, if the generator bearings overheat, the temperature sensor will not register the high temperature instantaneously because the sensors are outside.

For alerts and faults, there are associated internal FAQ documents within the company. The FAQ includes what to investigate, what parts to order, and other associated information for a given fault code, type, and equipment. When a fault or alert is raised, an FAQ is generated. The fault can be reset for a specified number of times. After the limit is exceeded, the wind turbine needs to be manually reset by climbing up the turbine.

At the data center, metrics that are calculated and tracked include **lost production factor** which is the percent of power the manufacturer is contracted to provide for but lost, **spinning performance** which is the deviation from the wind turbine's power curve, and the assignment of lost power/energy to specific faults or component.

One reason that machine learning isn't used from the manufacturer's point of view is because historical data has not sufficiently built up. Many of the customers are wind turbine farms and the benefit with wind turbine farms is that the turbines are at the same place, with the same ambient conditions, same model of wind turbine, and therefore one turbine can be compared against the rest of the farm. For example, the farm average can be calculated and the number of standard deviations of each turbine from the farm average can be calculated. One standard deviation from the average may be treated as a yellow level of fault, two standard deviations is orange, and three standard deviations is red. Alarm levels can be set below the fault levels to trigger the start of short term monitoring to protect the wind turbine.Use outlier detection and then compare to the global fleet. A surveillance group within the manufacturer breaks down the data by parts, platform, and parts consumed. After an alert or alarm is raised, this surveillance group can do a dive-down using a toolkit to specify the sampling rate of data to the millisecond resolution. Thus, the approach is to use outlier detection with comparisons to the global fleet of wind turbines.

Condition Monitoring System (CMS)

The aim of a CMS is to detect relevant changes in the condition of monitored components at an early stage to hinder a potential premature breakdown of these components. Components are monitored through the addition of accelerometer and vibration sensors. A condition-based maintenance strategy can save up to 20-25% of maintenance costs vs. scheduled maintenance of wind turbines [28].

For Vestas, CMS systems are third party systems which include additional acceleration sensors and and additional data. This was a popular option 4-5 years ago when about 2/3 of consumers electing to get one because there were many failures. Now, turbines are running much better and this has dropped down to 1/3 with about a 1% failure rate for main component per year. The CMS is part of the sales negotiation. The CMS data is handled differently, with higher resolution and a separate data stream. The data is sent to the CMS

3rd party, who interpret the data, make recommendations, the results of which are sent to the Vestas monitoring system.

For Siemens, CMS is standard and installed on all Siemens wind turbines. All wind turbines in a wind farm are monitored by the CMS. The system monitors the state of main wind turbine components by evaluating vibration data recorded from accelerometers placed in the wind turbine nacelle. The results, detailed analysis, and reprogramming of the CMS can be carried out using a standard web browser. The vibration data is reduced and compressed on-location at the wind farm before being sent to Siemens.

On-Site Maintenance

A wind farm may also have its own dedicated, on-site maintenance staff. If a wind turbine farm chooses not to keep on buying service contracts from the manufacturer, then during the service contract with the wind turbine manufacturer, the manufacturer's technicians may work next to the wind farm's technicians. Then, when the service contract expires, the wind farm's technicians take care of servicing the wind farm.

As an example of an on-site maintenance team, a farm with 44 wind turbines (Siemens Wind Turbine 2.3) may have a staff of 8-9 full time maintenance employees, each of whom is also on call 1-2 nights or weekend days. This staff reviews warmings, alarms, and alerts from the wind turbine, examines the SCADA data, and physically visits and investigates the wind turbines. An on-site maintenance employee's workday typically involves coming into the office in the morning and then spending the rest of the day out on the field. This on-site maintenance solution may be combined with a contract with the wind turbine manufacturer to remotely monitor the wind turbines — for example, outside working hours during the night to notify the on-call wind farm maintenance employee. The level of sophistication may be as simple as handwriting work orders and maintenance logs. A company that oversees multiple wind farms may have individual maintenance teams at each wind farm that work independently from the other wind farm maintenance teams.

Research in Fault Detection in Wind Turbines

Several review papers summarize and list many related papers on the topic of fault detection in wind turbines. A review of fault detection systems and condition monitoring systems in wind turbines, including common parameters and algorithms, broken down by methodology and components of the wind turbine can be found in [33]. Another review on condition monitoring in wind turbines chronologically describes papers one by one [61]. There also exists a review that focuses on condition monitoring and fault detection on wind turbine gearboxes, a subsystem in a wind turbine [53].

Fault detection methods applied to wind turbines include modeling the power curve under normal conditions from which a deviation from the power curve could be used to give an indication of a developing fault. Methods of modeling the power curve include using copula statistics [27] and kernel methods from which deviations from the power curve correspond to periods of poor turbine performance without deifferentiation between faults [64]. Modeling the power curve using Gaussian Process models has been able to show a performance degradation which began three months in advance of a main bearing failure on the turbine [6]. Another paper used a binning approach to modelling the power curve, followed by interpolating between the bins and shifting of the curve to find alarm limits [54].

Although Condition Monitoring Systems (CMSs) have been widely successful in other applications, CBM has not been taken up extensively by the wind industry, despite the supposed benefits [28]. A number of reasons exist for this [79, 78]. The capital cost of retrofitting sensors, as well as data collection and analysis can be quite high - upwards of Euro 13,000 per turbine.

However, there already exist a number of sensors on modern turbines related to the Supervisory Control and Data Acquisition (SCADA) system. In recent years, there has been a concerted effort to apply condition monitoring (CM) techniques to wind turbines by analysing data collected by the SCADA system. SCADA data is typically recorded at 10-minute intervals to reduce transmitted data bandwidth and storage, and includes a plethora of measurements such as active and reactive power, generator current and voltages, wind speed, generator shaft speed, generator, gearbox and nacelle temperatures, and others [82]. By performing statistical analyses on various trends within this data, it is possible to detect when the turbine is entering a time of sub-optimal performance or if a fault is developing. This is all done without the added costs of retrofitting additional sensors to the turbine [79].

Machine learning approaches to CBM of wind turbines include principal component analysis and neural networks or other pattern recognition methods, usually applied to high resolution data from retrofitted vibration or oil particulate sensors. One such approach used a modified version of the classification method K nearest neighbors on 100Hz sampled data from a simulated benchmark model of a wind turbine [67].

2.4 Fault Detection in Building Systems, Chiller Plants, and Chillers

Current Practises

In practise, detecting faults in building HVAC systems is currently a labour intensive process. Typically, building operators use intuition and rules of thumb to identify the problem, and the labor intensive nature of these tasks is such that they are not routinely performed. Commercial offerings of fault detection products remains very small.

The traditional approach to fault detection in buildings is through complaints from occupants, routine or unscheduled maintenance, and alarms on the HVAC control system. Faults are often found while tracking down the causes for occupant discomfort or during equipment inspections. However, faults that do not cause uncomfortable space conditions may be overlooked, while continuing to waste energy. Examples of such faults include simultaneous heating and cooling, and hunting and excessive cycling around a setpoint.

The current standard for software-based building fault detection offerings are single-variable alarms from the HVAC control system. There alarms are based on threshold values or durations on measured or computed variables. However, the technical personnel who are responsible for programming the building control systems are interested in setting up the controls for the HVAC, and not specifically to save energy, and they typically do not have training on how to write alarms that identify energy inefficiencies.

Data in Building Systems

Data from buildings is not extensively archived and reviewed to become useful, actionable information that can then be used for planning and execution purposes. This lack of data-driven action in buildings is largely due to the time and effort required. Building operators are often reactively responding to complaints and emergencies instead of pro-actively planning and acting to save energy.

On the other hand, the existing data and potential volume of data from buildings include utility bills, manual repair and maintenance logs, and extended logging of control system measurements. Modern building automation systems (BAS) are able to store, trend, and plot system-level operational or control data. However, BAS are often not configured with the capabilities enabled or used.

In terms of accessing the data on a control system, some control systems are built on a standard SQL database which can facilitate continuous data exchange with third party software. Others allow access to data in bulk through comma delimited files or spread sheets. Another method for extracting data from buildings is through communication protocols such as BACnet and LONtalk.

Aside from BAS, meter visualization tools and energy information systems (EIS) include interval data on electricity and gas consumption at whole-building or submeter level [29]. While it is also possible to integrate energy meters in BAS, it is not common practise.

For IT personnel, it is not a priority to ensure data flow and support for the BMS. The priority for IT personnel is to ensure that the critical systems, such as employee clocking in/out, sales, security, etc., are up and running, and not to save energy. Also, the controls company that installed the BMS is often not the same as the manufacturer of the HVAC equipment, such as the AHU, the pumps, and the sensors. Thus, the data is not seamlessly integrated. Furthermore, each equipment has its own data format and data protocals, which are hard to harmonize and integrate. Buildings are usually single instances, unlike wind turbines where there are many wind turbines that are all the same that are installed at a park.

Fault Detection in Building HVAC Systems

HVAC stands for heating, ventilation, and air conditioning. In practise, faults are detected through alarms on a building's HVAC control, complaints and feedback from building occupants, and discoveries of faults during routine or un-scheduled maintenance. Alarms in

HVAC control systems are usually set on single variables. These alarms on threshold values and duration are set by control system installers, designers, or by building personnel. When occupants complain due to uncomfortable space conditions, faults are sometimes detected when tracking down the source of the discomfort. Maintenance includes physical inspection of equipment by an experienced technician as part of periodic maintenance.

Third parties also offer solutions for identifying faults and inefficiencies. These solutions include monitoring services by experts that generate reports to inform building management stakeholders about efficiency opportunities. However, the high cost of these third party solutions have discouraged wide spread adoption.

A review of fault detection, diagnostics, and prognosics as applied to building systems can be found in [42][43]. This two part series of papers presents a taxonomy for methods and describes the different categories of methods. Methods are organized by the modelling and inference techniques. Quantitative model based methods are formulated from physical or first principle engineering models and faults were detected based on the deviation between model predictions and measurements. Another category is process history based methods, which detects faults in current data based on historical data using techniques such as pattern recognition. The final category is qualitative model-based methods which include expert inference rules and limits and alarms based on engineering first principles and practical experience.

The literature also includes proposed fault detection algorithms for or applied to building systems. A common approach in the literature is to identify specific faults of interest in specific building system components of interest, identify the relevant sensor and data inputs, develop a fault detection algorithm, and then assess the performance of the algorithm on the specific faults.

Quantitative methods to fault detection rely on physical modelling of the system thermodynamics, heat transfer, and mass transfer. Proposed methods in this category differ in the choice and detail of physical quantities that are modelled, and the algorithms for using such physical models. Typically, the quantitative modeling approached continuously computes residuals between predicted and measured system performance. Then, thresholds and simple discrimination functions convert these residuals into fault signals. For example, one method identifies six performance indexes for centrifugal chillers based on theoretical analysis of first principles [10], another example proposes performance indices for sub-system components along with a reference regression model [74]. Another method uses the frequency response of the building [81]. Quantitative models are prone to difficulties from transient interactions, un-modeled behaviours, over-fitting to a specific system, and the need for additional sensors.

Qualitative approaches include rule-based systems and models based on qualitative physics. Qualitative physics based models rely on qualitative relationships or knowledge bases to draw conclusions regarding the state of a system and its components. Rule based methdods are typically build from anecdotal and best practise guidelines. Desirable characteristics include ease of development and transparent reasoning towards a conclusion. However, expert systems tend to be highly specialized to a particular system and are prone to failure in situations beyond the boundaries of their programmed knowledge. Expert rules derived from mass and energy balances in air handling units are proposed by [59] and [36]. Anomaly detection is also used on daily readings of energy consumption and peak energy consumption using mean and standard deviation [60]. Identification of sensors related to energy performance and sensors related to energy balance enabled the use of principle component analysis (PCA) [73].

In the hybrid approach category, one proposed method of fault detection and diagnostics combines statistical learning tools with building energy models using minimal sensors and little customization [24]. Results indicate that the algorithms can detect and diagnose several common faults but additional work is required to reduce false positive rates and improve diagnosis accuracy. Other hybrid approachs uses expert rules, performance indexes, and statistical process control models [56] or a model combined with a support vector machine [48].

Fault detection approaches using machine learning has also been applied to HVAC systems. Methods include dynamic bayesian networks [30] , neural networks and adaptive subtractive clustering [15], density-based spatial clustering of applications with noise (DBSCAN), K-means, classification and regression tree (CART) [45].

Much of the research on chiller applications focuses on internal chiller components, for example, oil levels and refrigerant leaks within the chiller. This paper, on the other hand, analyzes faults at the plant level in the context of sensor data available in the controls systems. The benefit of this approach is that newer chillers already have FDD for internal components built in by the manufacturer, but system level FDD is rarely used in current business-as-usual operation and maintenance of commercial buildings.

Furthermore, existing fault detection built into equipment, or people themselves, can detect hard faults — failures that occur abruptly and either cause the system to stop functioning or to fail in meeting comfort conditions. On the other hand, soft faults that degrade performance but allow continued operation of the system are more difficult to detect and diagnose, and can continually waste energy [9].

2.5 Relevant Applied Statistics and Mathematics

Machine Learning

Machine learning, also called statistical prediction, is concerned with systems that learn to solve information processing problems. In this context, learning refers to the use of experience to improve performance. Experiences include data, queries, interaction, and experiments. Information processing problems include classification, regression, and ranking. An introductory textbook on machine learning is [41] and a more detailed textbook is [25]. The particular machine models employed are described in Chapter 3 Methodology. A particular area of machine learning relevant to this methodology is variable selection and feature selection, which is discussed in the next section.

Variable Selection and Feature Selection

While application domain knowledge can be used to construct features and select variables and features, there exist techniques that operate with little or no application domain knowledge. These techniques are commonly used in statistical learning and machine learning problems.

For data-driven methods, an overview of variable and feature selection that focuses on constructing and selecting subsets of features that are useful to building a good predictor — a good predictor in the "given X, predict Y" sense, not necessarily in the temporal sense — is provided in [32]. The paper covers methods for ranking variables, embedded and wrapper methods that use prediction performance to assess the relative usefulness of subsets of variables, methods for reducing the dimension — the number of variables — of data, methods for validating the selection, as well as some examples and open topics.

A contextual-based feature selection algorithm for fault detection is introduced in [7] to determine the interaction between features and select a minimum relevant feature subset, which is then demonstrated on fault detection in a chemical plant.

Feature selection for fault detection has also been done using decision trees for the machine learning methods of Proximal Support Vector Machine (PSVM) and Support Vector Machines (SVM) for faults in roller bearings [65]. Another paper used artificial neural networks (ANN) for fault detection in roller bearings, with time domain input features and the use of the ANN for feature construction and selection [58]. The details of the variable selection and feature selection methods used are described in Chapter 4 Methodology.

Time Series Data

Many machine learning methods and approaches are time invariant and assume that data are non-sequential. On the other hand, time series data, such as the data from sensor measurements, are examples of sequential data. A textbook on the modeling and analysis of time series data is [62]. For the use of sequential data in problems framed as statistical learning problems, [13] formalizes the principal learning tasks and describes machine learning methods that have been developed to address these problems.

Chapter 3

Methodology

3.1 Chapter Overview

First, this chapter presents a method for generalized machine fault detection, which is motivated by the energy saving opportunities described in Chapter 1. The proposed method uses existing data supplemented with expert engineering knowledge that is generic to the class of equipment. Second, this chapter proposes an intelligent algorithm that can effectively optimize energy consumption. The applications for this methods include wind turbines and buildings, which are described in Chapter 2 and will be demonstrated in the following chapters.

3.2 Application Domain Knowledge

Using the data variables from the control system, new features are created that incorporates knowledge about the application. The creation of new features derived from existing data variables in the control system saves money by avoiding the purchase, installation, and setup of new sensors. New features are also included over simply using the data variables as a means to include application specific information which could improve fault detection performance.

To create application domain knowledge features, experts were interviewed about metrics that were useful to monitor the condition of equipment and energy performance. These experts included building scientists, wind turbine researchers, technical leads at wind turbine manufacturers, and wind farm operators. Proposed features were also shown to experts to double check proposed features.

Control system data is supplemented with application domain knowledge to create a data matrix X of n rows of samples and m columns of features as shown in Figure 3.1. This process is novel in that it does not simply use the original sensor data treated as independent samples, but also includes time series characteristics and application domain knowledge; Furthermore, the original data is combined with this new knowledge in a simple

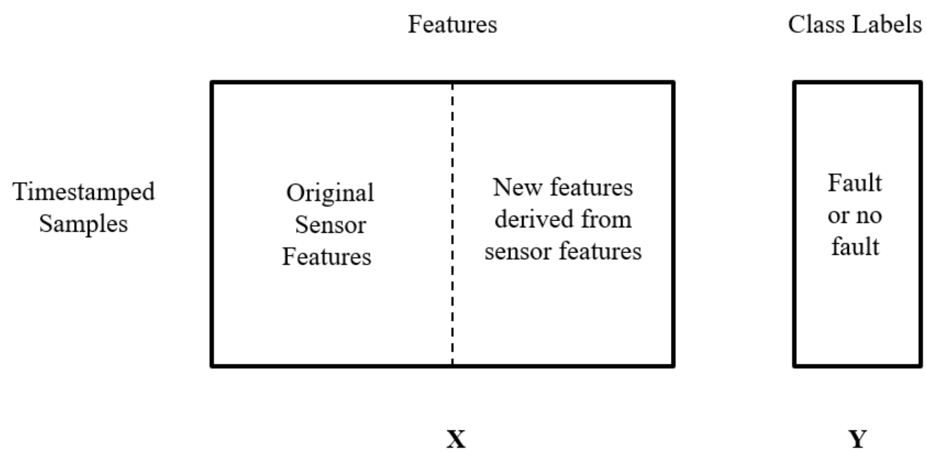


Figure 3.1: Data matrices X and Y for machine learning models. Matrix X is the control system data supplemented with application domain knowledge

form that is appropriate as input into numerous machine learning algorithms — a data matrix X of n rows of samples and m columns of features in a form that can be input into various machine learning algorithms.

The new features created using application domain knowledge include time series features, statistical features, and engineering knowledge features.

Engineering Knowledge Features

The original features are supplemented with new features created from the original features using knowledge of the application. Specifically, knowledge of the quantities that the original features correspond to, the location of where data for those features is collected, and an understanding of the operations of the application.

For example, given knowledge that on the wind turbine, there are ambient temperature features associated with the nacelle housing that correspond to the two ambient temperature sensors on the nacelle, it is expected that the two nacelle ambient temperatures would have similar values. Thus, a new feature is the difference between the two nacelle temperatures, and taking the difference can be done because the original features have the same units. Domain knowledge was acquired through interviews with manufacturers, controls companies, independent service providers, researchers in the application domain, previous work experience in the industry, and basic science and engineering knowledge.

Time Series Features

Because the applications are physical systems, it is known that the original features and derived features are physical quantities that form time series, hence timestamped data samples are not independent from one another. It is desirable to have the machine learning model work across time and to include correlation across sample columns. To that end, the data samples are converted to rolling time series representations of one hour using lagged features. The use of lagged features allows for the approximation of derivatives, an important aspect of physical systems. This is similar to a finite impulse response filter (FIR), specifically a discrete time and digital FIR.

To represent the data as time series, lagged features — also called delays — for each original feature are created to include the data from time $t - 60min$ to t. For example, if the original feature "Nacelle Ambient Temp" is at time t and the sampling resolution is 10 minutes, one new feature is "Nacelle Ambient Temp at time $t - 10min$", another new feature is "Nacelle Ambient Temp at time $t - 20min$", etc. all the way to "Nacelle Ambient Temp at time $t - 60min$". This procedure creates new features. The total number of new features created from this process is: (number of original features)*(order).

The n, which is called the order of the lag, is selected using Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC).

AIC measures the goodness of fit for a particular model by balancing the error of the fit against the number of parameters in the model. AIC is defined as [62]:

$$\text{AIC} = log\hat{\sigma}_k^2 + \frac{n + 2k}{n}$$

where k is the number of parameters in the model and $\hat{\sigma}_k^2$ is the maximum likelihood estimator for the variance:

$$\hat{\sigma}_k^2 = \frac{SSE_k}{n}$$

where SSE_k denotes the residual sum of squares under regression model with k regression coefficients.

The first term of AIC is thus a measure of the error of the fit and the second term of AIC penalizes the error variance proportionally by the number of parameters. This is because $\hat{\sigma}_k^2$ decreases monotonically as k increases.

The penalty term in AIC is not the only choice for a penalty term. Another penalty term based on Bayesian arguments is BIC, which is defined as [62]:

$$BIC = \log \hat{\sigma}_k^2 + \frac{k \log n}{n}$$

Simulation studies tend to verify that BIC does well on finding the correct order in large samples while AIC is superior on small samples where the relative number of parameters is large [62].

In this dissertation, for each feature, orders recommended by both AIC and BIC are calculated and the max order between AIC and BIC is selected.

Statistical Features

Statistical features are created from the original features. The first statistical feature is the rolling average. For example, the rolling average of "Nacelle Ambient Temp" at sample time t is the average of "Nacelle Ambient Temp" between $t - n$ to t. The rolling average is also a finite impulse response filter commonly called a boxcar filter. This has the benefit of suppressing high frequency noise. The second statistical feature is the rolling standard deviation. The size of the rolling window is the size of order selected for delay variables.

3.3 Feature Selection and Data Reduction

Feature selection methods are conducted on the original variables and new features to find subsets of features useful for prediction of faults. Reducing the number of features is also desirable to reduce the data requirements and computation time. The feature selection methods used include removal of zero-variance variables and features, and the use of MRMR. Of interest is to observe if and how many of the new features are among the top selected features to validate the usefulness of the expert knowledge features to detect faults.

First, the variance is calculated for each feature and features with zero variance are removed.

Minimal-Redundancy-Maximal-Relevance criterion

Next, the mutual information based minimal-redundancy-maximal-relevance criterion (mRMR) is run on the derived features and original features. mRMR is a feature selection method that is used to find a subset of features useful for prediction of the faults. mRMR reduces redundancy in the features and selects those most relevant to prediction [14]. Both Mutual Information Difference (MID) and Quotient (MIQ) schemes are used.

First, the feature set S with the maximum relevance to the target class c is chosen. The following criteria is used to search for features based on maximal relevance, based on the average mutual information $I(.)$ between each individual feature x_i and class c:

$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i \epsilon S} I(x_i; c) \tag{3.3.1}$$

However, features selected according to maximum relevancy could also have high redundancy, i.e., a lot of highly dependent features. If two features are highly dependent, removing one of the features would not have much of an effect on distinguishing between classes. So, the following minimum redundancy condition is added [55]:

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \epsilon S} I(x_i, x_j) \tag{3.3.2}$$

mRMR combines the above two constraints:

$$\max \Phi(D, R), \Phi = D - R \tag{3.3.3}$$

In practice, the near-optimal features defined by $\Phi(.)$ can be found by the incremental algorithm that optimizes the following condition:

$$\max_{x_j \epsilon X - S_{m-1}} \left[I(x_j; c) - \frac{1}{m-1} \sum_{x_i \epsilon S_{m-1}} I(x_j; x_i) \right] \tag{3.3.4}$$

The number of new derived features among the features selected by mRMR is then compared against the total number of selected features. Comparisons are made between the number of selected new features and the expected value of selected new features if features were randomly selected. The rankings of the new features are also compared against the rankings of original features.

3.4 Machine Learning Model Creation

To see if the new derived features improve fault detection performance, machine learning models are trained using the new feature set and the original feature set.

Creation of Training and Testing Sets

When testing on time series, care must be taken to not use future data to predict past data. To this end, the following procedure is used to select training and testing subsets of the data.

The entire data set is sorted by ascending time and divided into twelve continuous time blocks of almost equal size. Twelve is selected because then each time block approximately corresponds to a month of data. This training and testing process is done from $m = 1$ to $m = 11$ and the averages of the training and testing scores are calculated. An illustration of this process is shown below:

Train on [Jan] test on [Feb]

Train on [Jan, Feb] test on [March]

Train on [Jan, Feb, March] test on [April

etc.

Sample Weights

Samples are weighted to increase the importance of certain samples during the training process. This is done because of the unequal numbers of samples for faults and normal operations. A second reason is to weight the more recent data more strongly than older data. Thus, the selected sample weights are products of a term to account for the class – fault or no-fault – and a term to account for the time of the sample:

$$\text{Sample Weight} = (\text{Class Weight})(\text{Time Weight})$$

The class weight balances out the classes such that a class with fewer samples is treated as importantly as a class with more samples. The class weight is inversely proportional to class frequencies in the input data. For class i, the formula for the class weight is:

$$\text{Class Weight}_i = \frac{\text{Number of samples}}{(\text{Number of classes})(\text{Number of samples in class}_i)}$$

The time weight weights the most recent data twice as heavily as the oldest data. The weights for the other samples follow an exponential curve between these two end points. The equation of the exponential curve is:

$$\text{Time Weight} = Ne^{\lambda t}$$

Solving for the constants such that the newest data is weighted twice as heavily as the oldest data gives the following:

$$\text{Time Weight} = e^{\frac{\ln(1/2)}{t_{initial} - t_{final}}(t - t_{final})}$$

Support Vector Machines

The specific machine learning model used is the support vector machine. Support vector machines (SVMs) are supervised learning models with associated learning algorithms that solve for separating hyperplanes. SVMs are primarily used for classification and regression. Given training data that is labelled, SVMs solve for an optimal hyperplane that can classify new samples, depending on which side of the hyperplane the new samples are located. An illustration of a SVM is shown in Figure 3.3 where the two classes are circles and squares. The hyperplane is optimal in the sense that it maximizes the margin between the two classes and a larger margin generally lowers the generalization error of the classifier. The support vectors are the samples that help define the seperating hyperplane. In 3.3, the support vectors are the shaded-in samples.

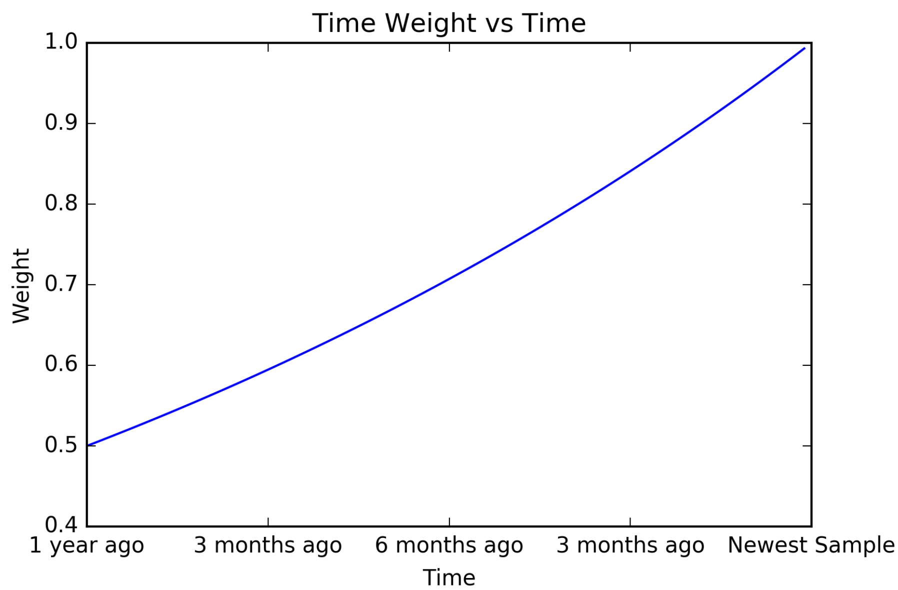


Figure 3.2: Graph of time weight

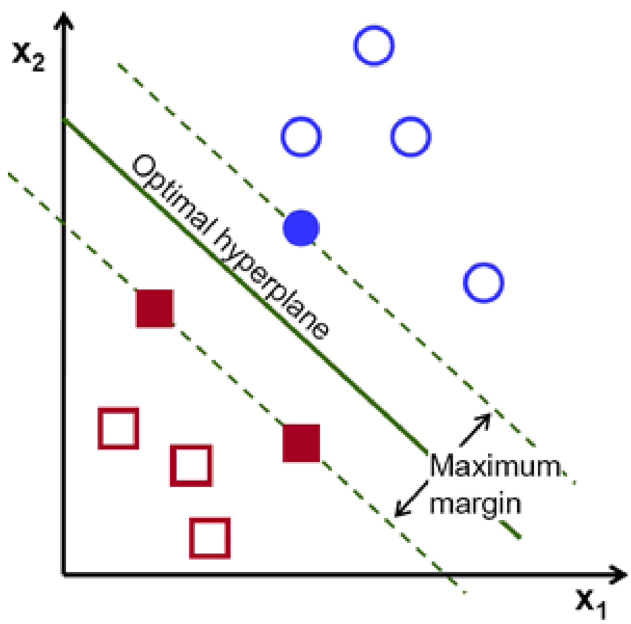
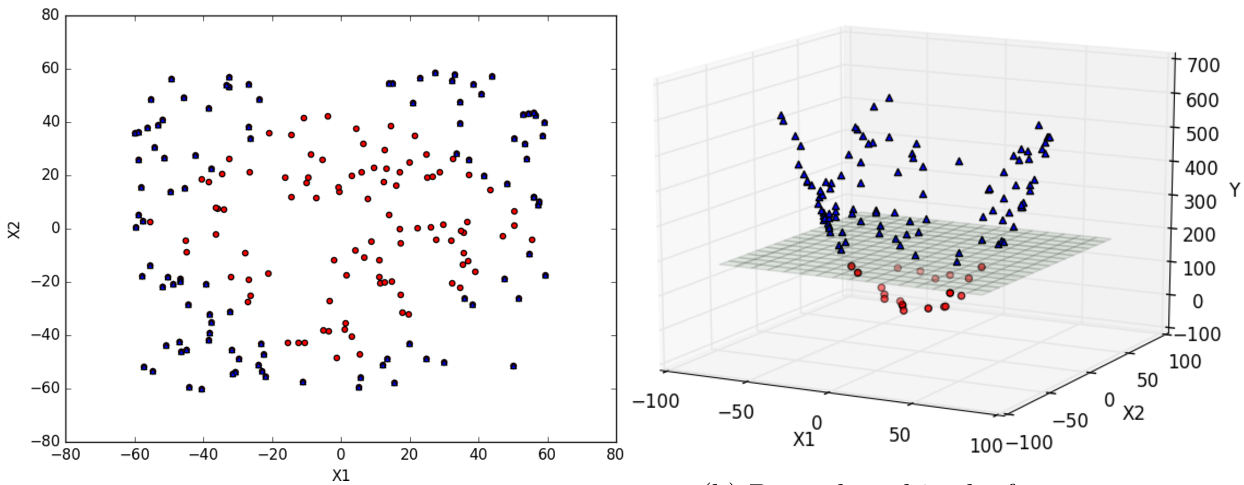


Figure 3.3: Illustration of a SVM. Image source: http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

Model Tuning and Cross Validation

A set of hyperparameters are also used to specify an SVM model. The hyperparameters that are searched over are C, γ, and the kernel. These hyperparameters are tuned using



(a) Data plotted in the input space

(b) Data plotted in the feature space

Figure 3.4: Example of using a quadratic kernel on data where 3.4a is the original data and 3.4b shows the original data plotted in the kernel feature space, resulting in the data from the two classes becoming linearly separable

10-fold cross validation on the training set to find the hyperparameters which yielded the highest f1 classification score. The C parameter is a penalty on misclassification of training samples, which must be traded off against simplicity of the decision surface. A low value for C makes the decision surface smooth, while a high value for C aims to correctly classify all training samples. The γ parameter defines the size of the influence of any single training sample. The larger the γ, the closer other samples must be to be affected. The kernel is a similarity function that calculates the similarity between two inputs to the function. The purpose of the kernel is to take the input data and transforms it in feature space to simplify the learning problem. For example, data that may not be linearly separable may become linearly separable in polynomial feature space. Figure 3.4 illustrates how data may not be linearly separable in the original input space but becomes becomes linearly separable in the quadratic feature space.

3.5 Performance Metrics

The overall accuracy of the classifier on the test set is a poor choice as a metric when the amount of data in different classes is imbalanced. For example, if 4,990 samples were correctly labelled as fault-free, and only 20 fault samples were also incorrectly labelled as such, the overall accuracy of the classifier would still stand at 99.6%. The formulae for calculating specificity, precision, recall and the F1-score are:

$$Recall = \frac{tp}{tp + fn} \tag{3.5.1}$$

$$Precision = \frac{tp}{tp + fp} \tag{3.5.2}$$

$$F1 = \frac{2tp}{2tp + fp + fn} \tag{3.5.3}$$

$$Specificity = \frac{tn}{fp + tn} \tag{3.5.4}$$

where tp is the number of true positives, i.e., correctly predicted fault samples, fp is false positives, fn is false negatives, i.e., fault samples incorrectly labelled as no-fault, and tn is true negatives.

The fault detection performance of the models that use the new features is compared against the fault detection performance of the models that use the raw variables for different numbers of features.

3.6 Energy Scheduling Procedure

The problem of managing energy consumption to save energy and money is formulated as mathematical optimization models. The models are solved using mathematical programming. In broad terms, mathematical programming can be defined as a mathematical representation aimed at programming or planning the best possible allocation of scarce resources. More specifically, mathematical programming concerns the optimum allocation of limited resources among competing activities, under a set of constraints. These constraints can reflect considerations due to finances, organization, technology, among others. When the mathematical representation exclusively uses linear functions, the resulting model is called a linear programming model. Likewise, a non-linear programming model includes non-linear functions and integer programming includes variables that are restricted to be integers.

Scheduling energy consumption can be framed as a mathematical programming problem to plan for the allocation of energy consumption among energy end-uses with energy treated as a scarce resource. First, the cost function is defined, which penalizes both the quantity of energy consumed, as well as the time of consumption based on time of use electricity rates. Constraints on the problem are also defined. These constraints include the physical behaviour of the electrical end-uses, electricity consumption patterns of end-uses, thermal characteristics, and occupant preferences and behaviour. This formulation is demonstrated on the residential data to calculate energy savings.

Next, a smart grid valuation framework is created that can be used to interpret the results produced by the mathematical model with respect to the efficiency of smart appliances and their respective prices. The framework is centered on four key questions: (1) which smart appliance provides the greatest overall savings?; (2) which smart appliance provides the greatest incremental savings?; (3) which smart appliance has the highest benefit/cost ratio?; (4) what incentives do smart appliances provide for behavioral changes?

Additional details on the formulation of the model is in Chapter 7

Chapter 4

Data

4.1 Chapter Overview

The previous chapter presented the proposed methodologies. This chapter first introduces the applications of wind turbines, commercial buildings, and residential buildings. Then, this chapter describes the sources of data used to execute and demonstrate the methodologies. This data covers 1) wind turbine data, 2) commercial building chiller plant data, 3) residential building energy end use, price, and weather data, and 4) expert knowledge of the above application domains. Details of the data processing and assumptions are described in the respective application chapters, Chapters 5 to Chapters 7.

4.2 Wind Turbines

Wind energy is a renewable resource which comes from the movement of air flowing across the earth's surface. A wind turbine converts the kinetic energy of the wind into usable energy, such as electrical energy (electricity) or mechanical energy (i.e., for pumping water or crushing grain). The power in the wind is extracted by allowing the wind to blow past blades. The resulting rotation of the rotor blades drives an electrical generator to produce electricity. Wind turbines can be classified based on the orientation of the rotational axis. The blades of a horizontal axis wind turbine (HAWT) rotate on an axis parallel to the ground. The blades of a vertical axis wind turbine (VAWT) rotates on an axis perpendicular to the ground.

The wind turbine used in this study is a horizontal axis wind turbine. Because sensor names and fault names will be referred to throughout this study, a brief overview of some components of the horizontal axis wind turbine relevant for this study is given below and shown in Figure 4.1.

The rotor blades are attached to a rotor. A rotor consists of multiple rotor blades attached to a hub. The rotor converts the movement of the wind into rotational movement.

The nacelle houses all the turbine machinery. This includes the drive train components: rotor shaft with bearings, transmission, generator, coupling and brake. To allow the nacelle

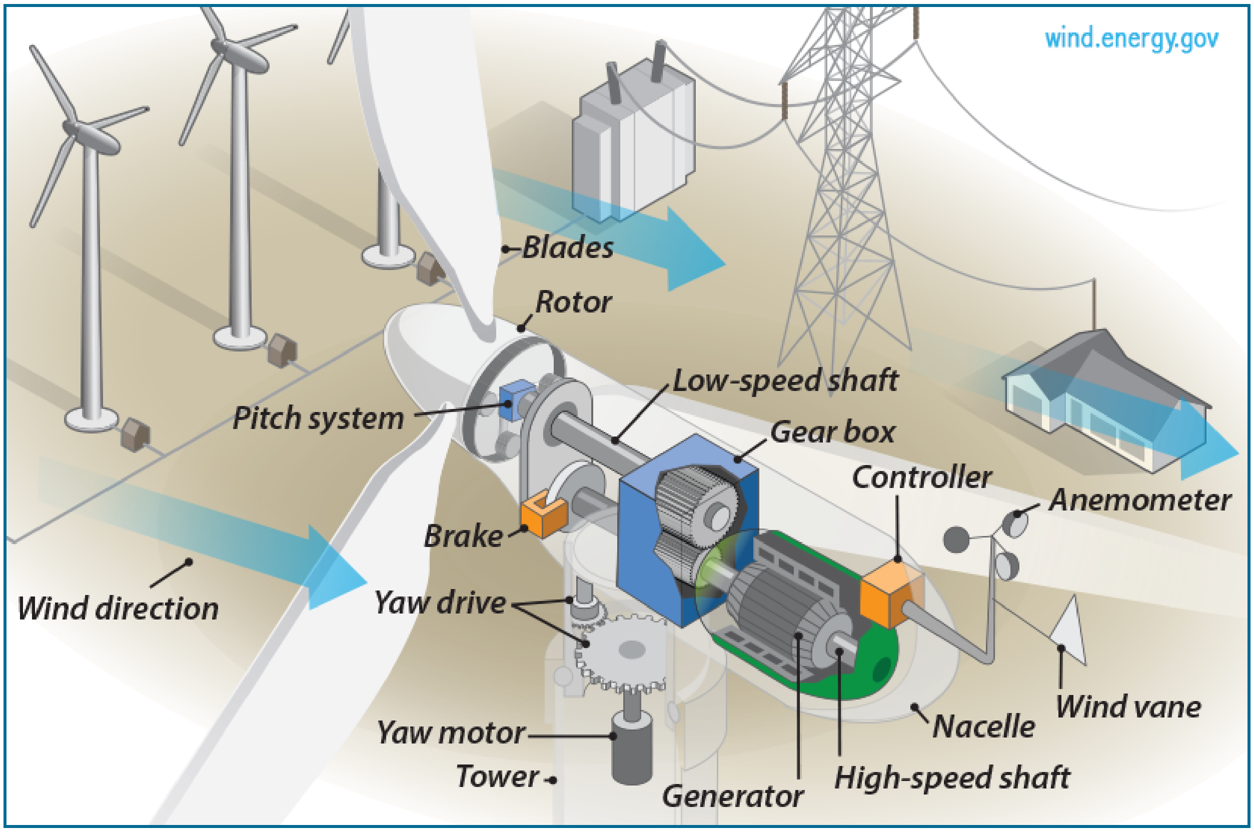


Figure 4.1: Components of a wind turbine. Credit: U.S. Department of Energy

to rotate to follow the wind direction, the nacelle is connected to the wind turbine tower via bearings.

The tower is the structure that elevates and carries the weight of the nacelle and the rotor blades, as well as absorbs forces from the varying wind.

The inverters are electrical components that convert the electricity from the generator into the specification required by the electric grid. This includes the required frequency of the electric grid.

4.3 Wind Turbine Data

Measured data was collected from a 3 MW direct-drive wind turbine near the coast in the South of Ireland. The turbine is manufactured by Enercon and the model is E-101. The data collected from the turbine covers an 11 month period from May 2014 - April 2015. This wind turbine supplies power to a major biomedical devices manufacturing plant.

The data is downloaded from a wind turbine's SCADA — Supervisory Control and Data Acquisition — system and is comprised of three separate datasets: "operational" data, "status" data, and "warning" data. These data sets are described in the following sections.

Table 4.1: 10 Minute Operational Data

TimeStamp	**Wind Speed (avg.)** m/s	**Wind Speed (max.)** m/s	**Wind Speed (min.)** m/s	**Power (avg.)** kW	**Power (max.)** kW	**Power (min.)** kW	**Bearing Temp (avg.)** °C
09/06/2014 14:10:00	5.8	7.4	4.1	367	541	285	25
09/06/2014 14:20:00	5.7	7.1	4.1	378	490	246	25
09/06/2014 14:30:00	5.6	6.5	4.5	384	447	254	25
09/06/2014 14:40:00	5.8	7.5	3.9	426	530	318	25
09/06/2014 14:50:00	5.4	6.9	4.5	369	592	242	25

Operational Data

The turbine control system monitors many instantaneous parameters such as wind speed and ambient temperature, power characteristics such as real and reactive power and various currents and voltages in the electrical equipment, as well as temperatures of components such as the generator bearing and rotor. The average, minimum and maximum of these values over a 10 minute period is then stored in the SCADA system with a corresponding timestamp. This is the "operational" data. A sample of this data is shown in Table 4.1. The initial operational data contained roughly 45,000 datapoints, representing the 11 months analysed in this study.

Status Data

There are a number of normal operating states for the turbine. For example, when the turbine is producing power normally, when the wind speed is below u_c, or when the turbine is in "storm" mode, i.e., when the wind speeds are above u_s. There are also a large number of statuses for when the turbine is in abnormal or faulty operation. These are all tracked by status messages, contained within the "Status" data. This is split into two different sets: (*i*) WEC status data, and (*ii*) RTU status data. The WEC (**W**ind **E**nergy **C**onverter) status data corresponds to status messages directly related to the turbine itself, whereas RTU data corresponds to power control data at the point of connection to the grid, i.e., active and reactive power set points. Each time the WEC or RTU status changes, a new timestamped status message is generated. Thus, the turbine is assumed to be operating in that state until the next status message is generated. Each turbine status has a "main status" and "sub-status" code associated with it. See Table 4.2 for a sample of the WEC status message data. Any main WEC status code above zero indicates abnormal behaviour, however many of these are not associated with a fault, e.g., status code 2 - "lack of wind". The RTU status data almost exclusively deals with active or reactive power set-points. For example, status 100 : 82 corresponds to limiting the active power output to 82% of its actual current output.

Table 4.2: WEC Status Data

Timestamp	Main Status	Sub Status	Status Text
13/07/2014 13:06:23	0	0	Turbine in Operation
14/07/2014 18:12:02	62	3	Feeding Fault: Zero Crossing Several Inverters
14/07/2014 18:12:19	80	21	Excitation Error: Overvoltage DC-link
14/07/2014 18:22:07	0	1	Turbine Starting
14/07/2014 18:23:38	0	0	Turbine in Operation
16/07/2014 04:06:47	2	1	Lack of Wind: Wind Speed too Low

Warning Data

The “Warning” data on the turbine mostly corresponds to general information about the turbine, and usually isn’t directly related to turbine operation or safety. These “warning” messages, also called “information messages” in some of the turbine documentation, are timestamped in the same way as the status messages. Sometimes, warning messages correspond to a potentially developing fault on the turbine; if the warning persists for a set amount of time and is not cleared by the turbine operator or control system, a fault is raised and a new status message is generated.

The “operational” data has 64 features with a sampling resolution of every 10 minutes. These features include wind speed, ambient temperature, power characteristics such as real and reactive power, and temperatures of components in the wind turbine such as the generator bearing and rotor. For some of the quantities above, the features include the average, minimum and maximum over the 10 minute period.

The “status” and “warning” data are event logs. The “status” data records changes in the status of the wind turbine. The “warning” data is also called “information messages” and mostly corresponds to general information about the turbine. The “status” and “warning” data is used to create labels and to process the “operational” data, which is described in Chapter 5.

4.4 Commercial Building Chiller Plants

A commercial buildings is “a building with more than 50 percent of its floor space used for commercial activities, which include stores, offices, schools, churches, libraries, museums, health care facilities, warehouses, and government buildings except those on military bases” [17].

The air conditioning and cooling for commercial buildings as well as industrial buildings are often provided by chiller plants. A chiller plant is a large facility that provides centralized chilled water to cool large buildings and multi-building campuses. Within the chiller plant,

Figure 4.2: One case study is the Molecular Foundry at LBNL

the chillers are the refrigeration machines that provide the chilled water. The chilled water that circulates through the buildings is called 'chilled water'. A chiller removes heat from the 'chilled water' and ejects that heat into 'condenser water'. The 'condenser water' is then circulated through a cooling tower to eject the heat into the atmosphere. Cooling tower fans aid in this process. The movement of water is accomplished via pumps.

4.5 Commercial Building Chiller Plant Data

Simulated data was obtained from simulations of one chiller plant. The chiller plant is the Molecular Foundry at the Lawrence Berkeley National Laboratory. The Molecular Foundry, pictured in Figure 4.2, is a 6-floor nanoscience research facility that is served by a chiller plant that has three chillers and two cooling towers that are operational for 24 hours a day, seven days a week.

Simulated data for the chiller plant was obtained from physics-based simulations. It is undesirable to introduce faults in real plants to collect measured data because introducing faults would interfere with normal operations of the facilities and damage the expensive equipment. The physics-based model was created and calibrated by research collaborators for the three chiller plants. The physics-based model was created using the Modelica modeling language [4] within the Dymola modeling and simulation environment [49]. A graphical overview of the model for the Molecular Foundry is show in Figure 4.3. The model was calibrated to sensor measurements collected from the facility.

The simulation was run for a year using standard weather data that represented an average year in the building's location of the San Francisco Bay Area. The simulation contained numerous variables required for simulation with user-specified sampling frequency. Because the goal of this research was to use existing sensors common across systems, only a subset of the variables and data in the simulation were used. Experts were interviewed about which data points were common across chiller plants. The common data points in chillers are sum-

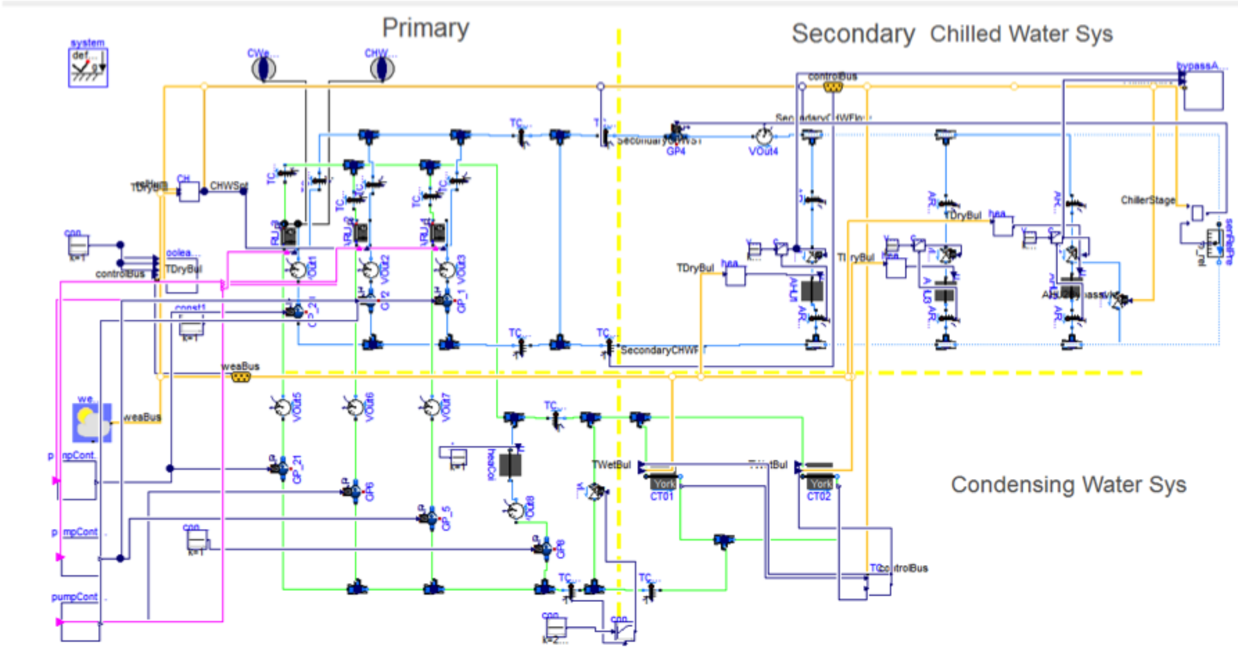


Figure 4.3: Visual overview of the physics-based model created using Modelica of the Molecular Foundry, which was one of the chiller plant models used to generate training data under various fault conditions

marized in Table 4.3. Only the commonly available data points were used from the simulation sampled at a frequency available in real buildings. Data from valves was not available in the simulation and the control signals and status signals were the same in the simulation. For the chiller plant, there are 54 measurement variables sampled every 5 minutes.

Table 4.3: List of common measurements in chiller plants

Location	Measurement
Chiller	Entering chiller water temp
	Leaving chilled water temp
	Chilled water set point
	Chiller status
	On off control signal
	Compressor motor currents
	Chilled water valve percent opening
	Entering condenser temp
	Leaving condenser temp
	Condenser water valve percent opening
Primary Chilled Water Loop	supply chilled water temp
	return chilled water temp
Primary Loop Pump	Pump on off status
	Pump on off signal
Secondary Loop Pump	On off status
	Variable frequency drive set point
	On off control signal
	Motor current
Secondary Loop	Supply chilld water temp
	Return chilled water temp
	Secondary loop chilled water mass flow rate
	Differential pressure measurement
	Differential pressure set point
Condenser Water Loop Pump	On off status
	On off control signal
Cooling Tower	Tower status
	Control signal
	Variable frequency drive fan speed
	Condenser water supply temp
	Condenser water return temp
	Supply and return valve status
	Supply and return valve open command
	Fan motor current
Weather	Outside air humidity
	Dry bulb outside air temp

4.6 Residential Building Data

The residential building data include both measurement data and simulated data, which is based on assumed appliance loads and end-use timings.

The measured data is the end-use data found in the Reference Energy Disaggregation Data Set (REDD) [46]. This data set is publicly available, possesses many of the desired traits described in Chapter 7, and is intended for research purposes. This time series data was measured from a residential house in Boston, MA for two months in 2011 and includes data on the electricity consumed by end uses such as the dishwasher, furnace, washer, and dryer.

The simulated data set is based on a combination of design choices about the location, outside temperature range, and available electricity prices. Weather temperature data are the actual measured temperatures during the modeling period. The temperatures were acquired from the Weather Underground, a commercial weather service that provides real-time weather information [77]. Their site collects most of its data from the National Weather Service. For appliances, wattage ranges were taken from a U.S. Department of Energy report EG. For modeling an electric vehicle, the Mini E was selected and it's performance specifications were used [69]. Electricity prices for day ahead prices and time of use prices were acquired from the local utilities: Ameren[3] and Austin Energy [16]. Further details are provided in Appendix A.

4.7 Application Domain Expert Knowledge

To collect domain application knowledge, experts were contacted for interviews, group brainstorms, documentation, and verification of proposed new features. These experts include: a performance specialist and a director power plant engineering of a major wind manufacturer, a business development engineer at another major wind manufacturer, a wind farm operator, a wind turbine researcher, six building scientists, a data engineer at a building controls company, and one building manager. A total of eleven interviews were conducted.

Chapter 5

Wind Turbine

5.1 Chapter Overview

This chapter demonstrates the proposed procedure on wind turbines. First, the data source and collection process are described. Then, processing and computation on the data are explained. Finally, results are presented and analysed. This chapter is based on [37]. Previous published work by the author is [47].

5.2 Data

Data Collection

The data is downloaded from a wind turbine's SCADA system and is comprised of three separate datasets: "operational" data, "status" data, and "warning" data.

The "operational" data has 64 features with a sampling resolution of every 10 minutes. These features include wind speed, ambient temperature, power characteristics such as real and reactive power, and temperatures of components in the wind turbine such as the generator bearing and rotor. For some of the quantities above, the features include the average, minimum and maximum over the 10 minute period.

The "status" and "warning" data is used to create labels and to process the "operational" data. The "status" and "warning" data are event logs. The "status" data records changes in the status of the wind turbine. The "warning" data is also called "information messages" and mostly corresponds to general information about the turbine.

The processing of the data to clean the data and to create labels is described in the next section.

Data Cleaning and Pre-Processing

Samples that are in between the 10 minute sampling times are removed and samples with the same time stamp are averaged. The labels for the samples were determined based on the "status" data.

The data is normed and balanced class weights are used for training of machine learning models. There are a total of 54 original features.

Creation of Classification Labels

To conduce supervised learning, it is necessary to develop a set of class labels for the data. Class labels are created for the "operational" data using the "status" data. For this, a list of frequently occurring faults was made, shown in Table 5.1. For these faults, status messages with codes corresponding to the faults were selected. The start and end of these turbine states was used to match up the associated 10-minute operational data.

In this paper, we attempt two levels of classification: fault/no-fault, where samples are classified as faulty or fault free; and fault diagnosis, where samples are classified as a specific fault, or fault-free. The process for labelling data is explained in this section.

No-Fault Dataset

Unless a data sample is labelled as a fault, it is labelled as a no-fault data sample.

All Faults Dataset

To develop a data set for fault operations, a list of frequently occurring faults was made. For these faults, status messages with codes corresponding to the faults were selected. Next, a time band of 600s before the start, and after the end, of these turbine states was used to match up the associated 10-minute operational data. The 10 minutes timeband was selected so as to definitely capture any 10-minute period where a fault occurred, e.g., if a power feeding fault fault occurred from 11:49-13:52, this would ensure the 11:40-11:50 and 13:50-14:00 operational data points were labelled as faults. The faults included are summarised in Table 5.1. Note that the fault frequency refers to specific instances of each fault, rather than the number of data points of operational data associated with it, e.g., a generator heating fault which lasted one hour would contain 6 operational data points, but would still count as one fault instance. *Feeding faults* refer to faults in the power feeder cables of the turbine, *excitation errors* refer to problems with the generator excitation system, *mains failure* refers to problems with mains electricity supply *to* the turbine, *malfunction aircooling* refers to problems in the air circulation and internal temperature circulation in the turbine, and *generator heating faults* refer to the generator overheating.

Table 5.1: Frequently occuring faults, listed by corresponding status code

Fault	**Main Status Code**	**Fault Incidence Frequency**	**Operational Data Points**
Feeding Fault	62	92	251
Excitation Error	80	84	168
Malfunction Aircooling	228	20	62
Mains Failure	60	11	20
Generator Heating Fault	9	6	43

Specific Fault Datasets

For specific faults, the same methodology for all faults was used, but this time single status codes for each fault code in Table 5.1 were used. Again a time band of 600s before the start and after the end of each fault status was used to match up corresponding 10-minute operational data.

The wind turbine faults included in this data set are summarised in Table 5.1. While the types of faults present in the data affect the features that are selected in the feature selection process, the methodology presented here can be applied to other faults. When reading Table 5.1, note that the fault frequency refers to specific instances of each fault, rather than the number of data points of operational data associated with it, e.g., a generator heating fault which lasted one hour would contain 6 operational data points, but would still count as one fault instance.

Thus, five types of faults are included in this data set. There are 213 occurrences of faults during the observation period, resulting in a total of 437 fault data samples.

5.3 Methodology

Overview

The "operational" data and corresponding features collected from the wind turbine are supplemented with additional features derived from the "operational" data based on domain knowledge. This process, is novel in that it does not simply use the original sensor data treated as independent samples, but also includes time series characteristics and application domain knowledge; Furthermore, the original data is combined with this new knowledge in a simple form that is appropriate as input into numerous machine learning algorithms — a data matrix X of n rows of samples and m columns of features.

Application Domain Knowledge Features

The original features in the "operational" data is supplemented with new features created from the original features using knowledge of wind turbines. Specifically, knowledge of the quantities that the original features correspond to, the location of where data for those features is collected, and an understanding of the operation of the wind turbine. For example, given knowledge that there are two nacelle ambient temperature features that correspond to the two ambient temperature sensors on the nacelle, it is expected that the two nacelle ambient temperatures would have similar values. Thus, a new feature is the difference between the two nacelle temperatures, and taking the difference can be done because the original features have the same units. This domain knowledge about wind turbines was acquired through interviewing wind turbine manufacturers and independent service provides, previous work experience in the wind turbine industry, and basic science and engineering knowledge.

A summary of the new features derived from the original features using knowledge of the application domain of wind turbines is shown in Table 5.2, which represents 67 derived features.

Time Series Features

The wind turbine is a physical system and it is known that the original features and derived features are physical quantities that form time series, hence timestamped "operational" data samples are not independent from one another. It is desirable to have the machine learning model work across time and to include correlation across sample columns. To that end, the "operational" data samples are converted to rolling time series representations of one hour using lagged features. The use of lagged features allows for the approximation of derivatives, an important aspect of physical systems. This is also similar to a finite impulse response filter (FIR), specifically a discrete time and digital FIR.

To represent the "operational" data as time series, lagged features — also called delays — for each original feature are created to include the data from time $t-60min$ to t. For example, if the original feature "Nacelle Ambient Temp" is at time t and the sampling resolution is 10 minutes, one new feature is "Nacelle Ambient Temp at time $t - 10min$", another new feature is "Nacelle Ambient Temp at time $t - 20min$", etc. all the way to "Nacelle Ambient Temp at time $t - 60min$". This results in 324 time-lagged features.

Statistical Features

Statistical features are created from the original features. The first statistical feature is the 2-hr rolling average. For example, the rolling average of "Nacelle Ambient Temp" at sample time t is the average of "Nacelle Ambient Temp" between $t - 2hr$ to t. The rolling average is also a finite impulse response filter commonly called a boxcar filter. This has the benefit of suppressing high frequency noise. The second statistical feature is the 2-hr rolling standard deviation, together resulting in 108 statistical features.

Feature Selection

A summary of all the new features used is found in Table 5.2

Given that the number of features are greatly increased, feature selection methods are conducted to find a subset of features useful for prediction of the faults.

After the derived features are created, feature selection methods is run on the derived features and original features to see if the derived features are selected.

First, the variance is calculated for each feature and features with zero variance is removed. None of the features had zero variance.

Next, a univariate chi-squared test and f-test is run on the derived features and original features. The features are ranked according to the chi-squared statistics and f-statistics to select the top ranking features.

Finally, mRMR is run on the derived features and original features.

The top ranked features selected are shown in Table 5.3. Most of the top ranked features are derived features - note that time lagged features are actually multiple features that correspond to different time lags - and few of the original data features make the top of the lists of top ranked features.

Among the selected features, the derived features are often ranked higher than the features that they are derived from. For example, the rolling average of “avg blade A angle”s, and the standard deviation of “spinner temp” rank higher than the original features. Some time lagged features also rank higher than the original features. Moreover, the time-lagged features are grouped together in the rankings, which make intuitive sense as the features correspond to the same physical sensor and measured quantity.

Given that the number of features are greatly increased, the mutual information based minimal-redundancy-maximal-relevance criterion (mRMR) feature selection method is used to find a subset of features useful for prediction of the faults. mRMR reduces redundancy in the features and selects those most relevant to prediction [55]. Both Mutual Information Difference (MID) and Quotient (MIQ) schemes are used.

Machine Learning

To see if the new derived features improve fault detection performance, support vector machines (SVMs) are trained using the new feature set and the original feature set for different numbers of features. The measured data is randomly separated into training and testing sets with 80% of the data in the training set and 20% of the data in the testing set. To train each SVM, a randomised grid search is performed over hyperparameters using 10-fold cross validation to find the hyperparameters which yielded the highest F1 classification score. The hyperparameters that are searched over are C, which controls the number of samples allowed to be misclassified, γ which defines how much influence an individual training example has, and the kernel.

Due to the asymmetry in the number of samples for the two classes, the class weights are balanced, that is, the classes are weighted inversely proportional to class frequencies in the

Table 5.2: Features from Knowledge of the Application

Average Of	Inverter Cabinet Temperatures belonging to one system Front and Rear Bearing Temperatures Pitch Cabinet Blade Temperatures Rotor Temperatures Stator Temperatures Nacelle Ambient Temperatures
Difference Between	Max and Min of (wind speed, rotation, power, reactive power) Max and Average of (wind speed, rotation, power, reactive power) Min and Average of (wind speed, rotation, power, reactive power) Available Power (from wind, technical reasons, force majeure reasons, force external reasons) Average Power and Available power (from wind, technical reasons, force majeure reasons, force external reasons) Inverter Cabinet Temperatures Inverter Cabinet Temperatures and Average Inverter Cabinet Temperature by system Front and Rear Bearing Temperature Average Bearing Temperature and (Front and Rear Bearing Temperature) Pitch Cabinet Blade Temperatures Average Pitch Cabinet Blade Temperature and Pitch Cabinet Blade Temperatures Rotor Temperatures Average Rotor Temperature and Rotor Temperatures Stator Temperatures Average Stator Temperature and Stator Temperatures Nacelle Ambient Temperatures Average Nacelle Ambient Temperatures and Nacelle Ambient Temperatures Nacelle Temperature and Nacelle Cabinet Temperatures Ambient Temperature and (Nacelle Temperature, Nacelle Cabinet Temperatures, Main Carrier Temperature, Rectifier Temperature, Inverter Cabinet Temperature, Tower Temperature, Control Cabinet Temperature, Transformer Temperature) Generator Temperature and Nacelle Temperature
Ratio of	Average power to Available Power (from wind, technical reasons, force majeure reasons, force external reasons)

input data. The models are evaluated using specificity, precision, recall and F1-Score.

Table 5.3: Top 10 features selected by mRMR under the MIQ and MID scheme with feature rankings

Feature	MIQ	MID
Difference between P technical and P external	1	1
System 1 inverter 1 cabinet temp t-30min	2	NA
2hr mean of average blade angle A	3	NA
2hr stddev of spinner temp	4	NA
Difference between P technical and P majeure	4	NA
2hr mean of RTU ava Setpoint 1	6	2
2hr stddev of rear bearing temp	7	3
2hr mean of rotor temp 1	8	NA
Difference between avg Power and P from wind	9	NA
Average Nacel position including cable twisting t-60min	10	9
Difference between rotor temps	NA	4
Difference between P from wind and P technical	NA	5
Min windspeed t-60min	NA	6
Min windspeed t-20min	NA	7
Difference between nacelle ambient temps	NA	8
Difference between average and min rotation	NA	10

5.4 Results and Discussion

Selected Features

In the feature selection process, the new derived features are favoured more than the original features. The top ten features as selected by mRMR under the MIQ and MID schemes are shown in Table 5.3. All of the top ten features are new features. Furthermore, all the different types of new features — application domain knowledge features, time series features, and statistical features — are among the top ten features. There is some overlap between the features selected under the MIQ and MID schemes. Differences in rankings and selected features are expected because to select the features, the MIQ scheme uses a quotient while the MID scheme uses a difference.

Figure 5.1 plots the number of new derived features among the features selected by mRMR under the MIQ and MID schemes, along with their respective rankings. The dotted red line traces the expected curve if all the top features are new derived features. The actual curves follow this reference line closely. The deviations of the curves from the dotted reference line are the number of original features among the top selected features. This deviation is less than the number of original features selected if features were selected randomly from the new feature set.

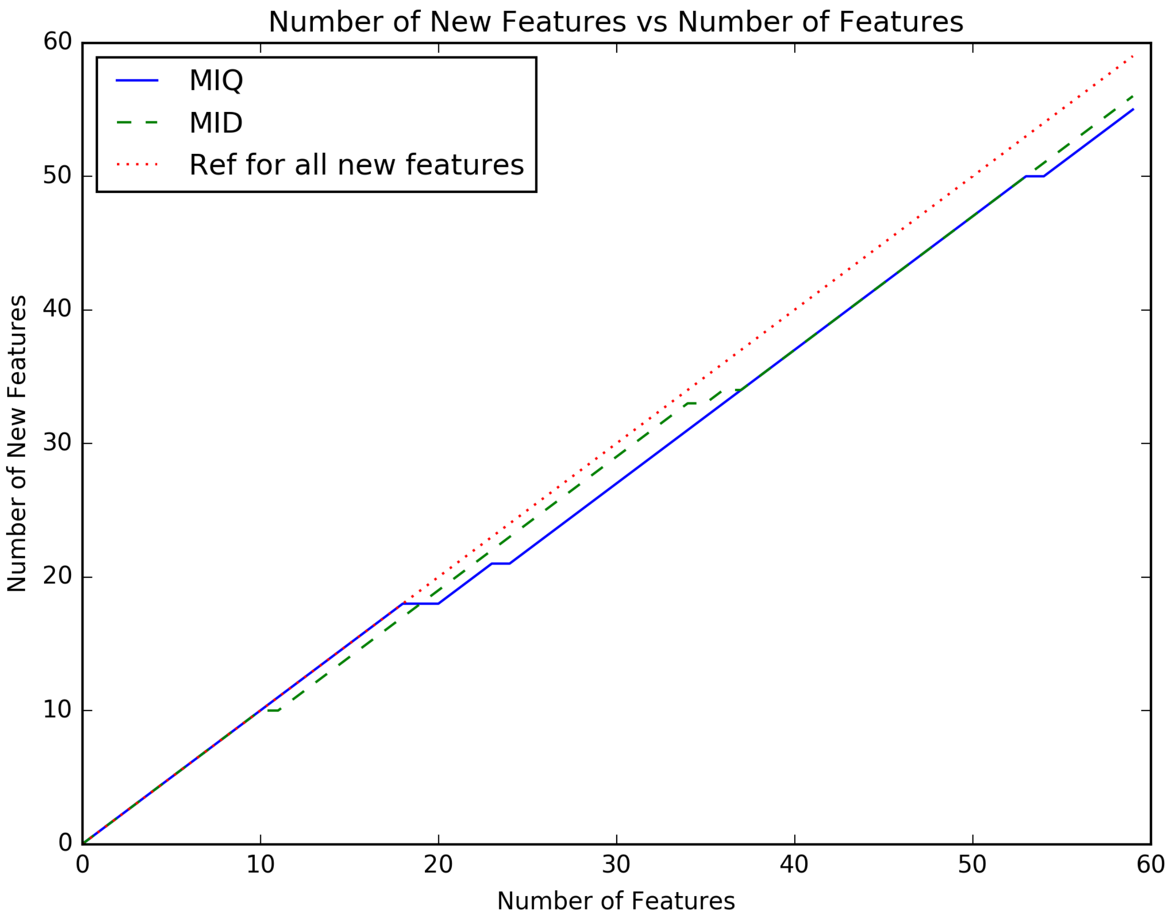


Figure 5.1: Number of new features among the top features with a dotted reference line indicating the expected line if all the features were new features

Among the selected features, the derived features are often ranked higher than the features that they are derived from. The histogram in Figure 5.2 tallies the differences in rankings between the new derived features and their respective original features. To create Figure 5.2, the original features are binned together with their derived features. That is, features associated with the same original sensor reading are grouped together into a bin. The feature selection ranking of the first derived feature selected from that bin is compared to the feature selection ranking of the original feature by calculating the difference between the rankings. For example, if the standard deviation of "rear bearing temp" is the first derived feature selected from the "rear bearing temp" bin and has a ranking of 4, and the "rear bearing temp" feature has a ranking of 54, then the relative ranking is 50. Positive differences in ranking indicate that the derived feature is ranked higher than the original feature that it was derived from. The magnitude of the difference indicates how many rankings higher the

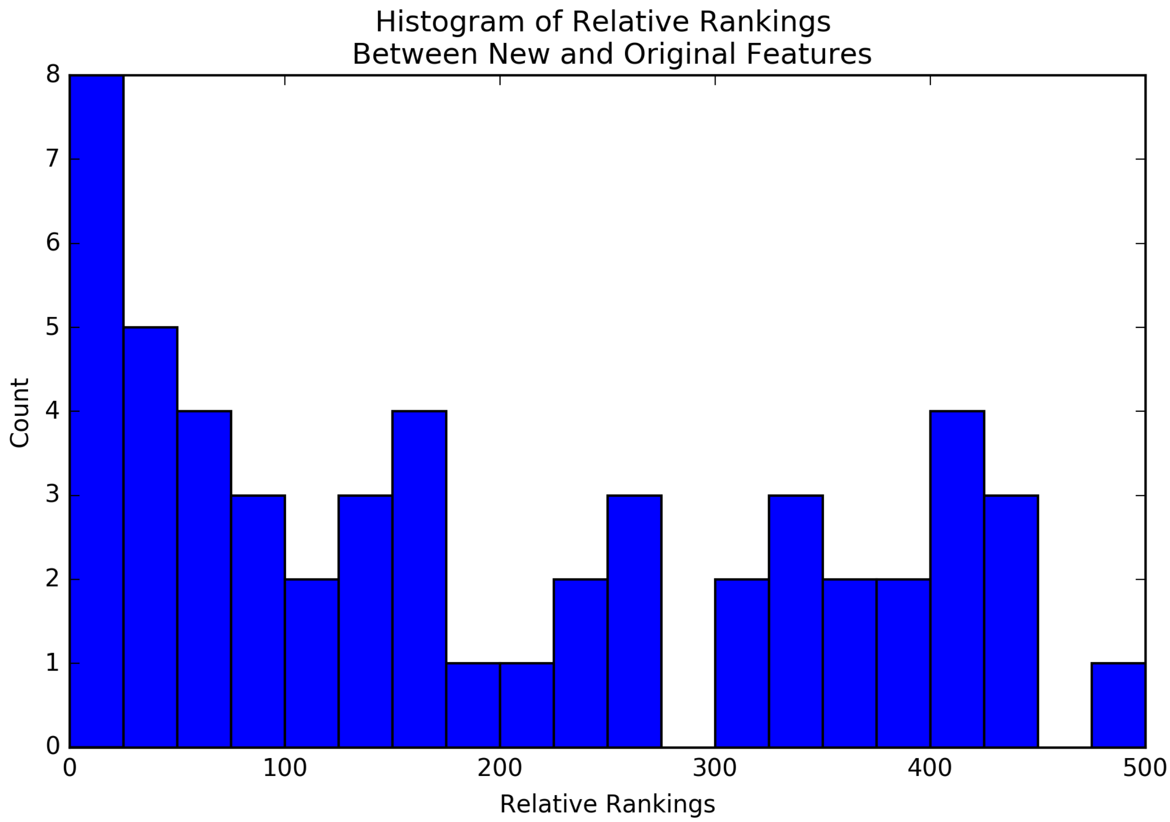


Figure 5.2: Histogram of relative ranking between original features and binned features

derived feature is ranked compared to its original feature.

In Figure 5.2, all the ranking differences are positive, indicating that derived features are selected before the original features that they are derived from. Furthermore, the derived features are sometimes ranked substantially higher than the original features, with ranking differences in the hundreds.

Model Prediction Results

Inclusion of the new derived features shows improvement in classification scores over the use of the original feature set for classification of faults and no-faults.

The classification results from training SVMs using different feature sets are shown in Figure 5.3. Figure 5.3 plots the F1 score on the first row of graphs, precision on the second row of graphs, and recall on the third row of graphs for the 'fault' class on the y-axes. The first column of graphs is the scores on the training set and the second column of graphs is the scores on the testing set. The x-axis is the number of features used in the SVM.

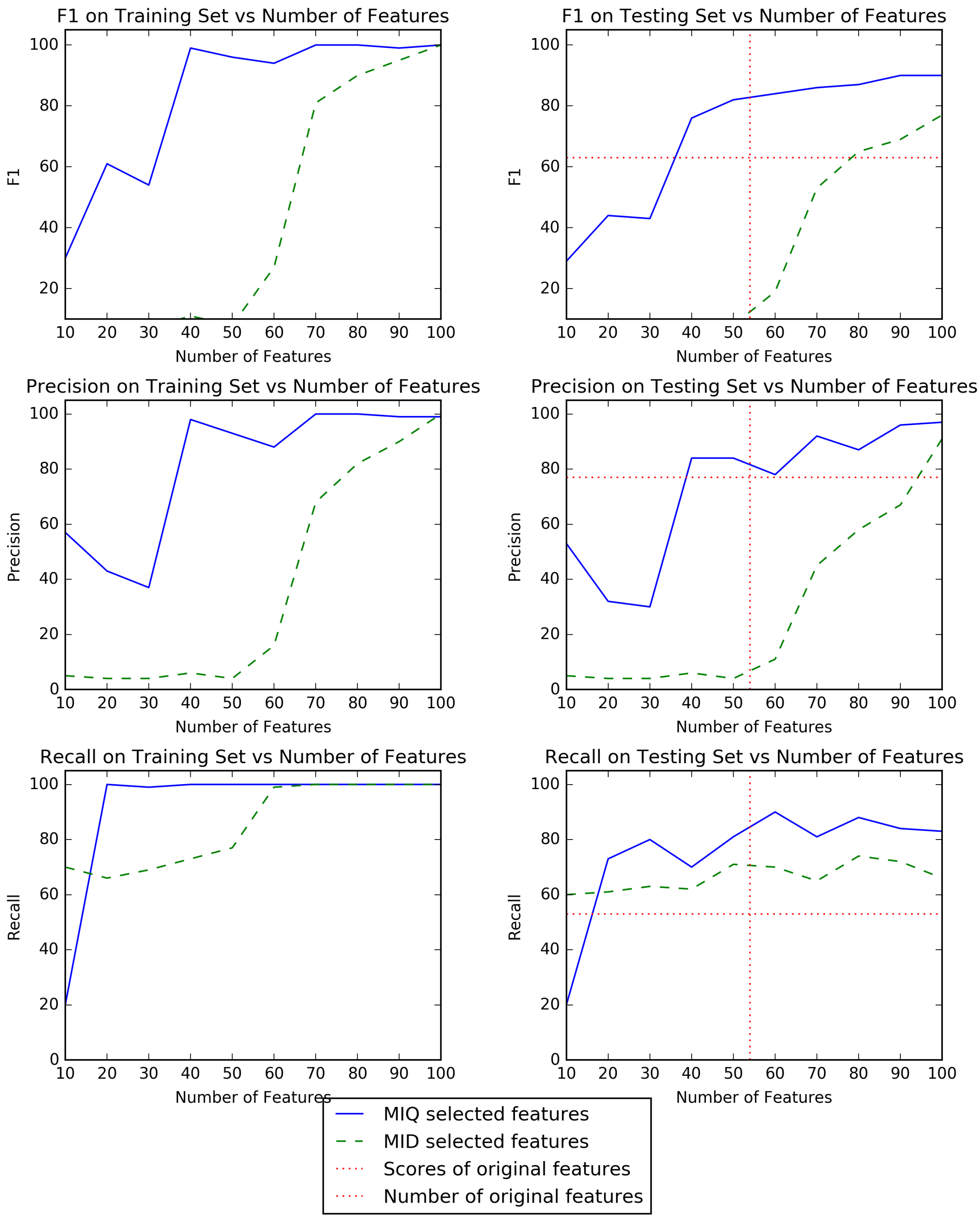


Figure 5.3: Scores on training and testing sets under the MIQ and MID scheme with reference scores and number of features of the original feature set

Our ultimate goal is to save money and reduce computation time with fewer sensors and fewer features. From Figure 5.3, it is apparent that most of the benefit to prediction performance is already realized around 40 features, when the performance metrics plateaus, and thus the x-axis is not extended further.

The horizontal dotted red lines sketch out the classification scores for an SVM trained on the original features in the "operational" data set: an F1 score of 63%, precision of 77%, and recall of 53%. Values above the horizontal dotted red reference lines show classification performance better than performance achieved using the original features.

The vertical dotted red line traces the number of features in the original feature set; there are 54 features in the original feature set. Values to the left of the dotted red line indicate the use of fewer features than in the original feature set.

The solid blue line and dashed green lines plot the classification scores of SVMs trained using features from the new feature set as chosen by mRMR under the MIQ scheme and the MID scheme respectively. Features selected using the MIQ scheme result in better classification scores than features selected using the MID scheme. This is observed in how the MIQ lines are higher than the MID lines.

Using new derived features selected by mRMR under the MIQ scheme, the same F1 score and precision as the original feature set is achieved using approximately two-thirds of the original number of features. Furthermore, a better recall score is attained using as few as 20% of the original number of features.

When using the same number of features as the original feature set, the F1 score improves by 19%, the precision by 7% and the recall by 27% on the test set. Even with 75% of the number of features, the F1 score is higher by 13%. The F1 score reaches 90% when 90 features from the new feature set are used, which is an improvement in F1 score of 27%. The flattening and irregularity of the line after 50 features are possibly due to the correlation of the features, as there are only initially 54 features from which additional features are derived. Furthermore, the selected features are likely to be dependent upon the wind turbine faults that are present in the dataset, and the selected features could be expected to vary with specific wind turbine faults.

These classification performance results demonstrate that the proposed procedure for incorporating expert domain knowledge can not only improve fault detection performance but can do so using fewer features. The use of fewer features enables savings from decreased data collection needs — such as from sensors, instrumentation, installation, networking, data quality, and data transfer needs — as well as avoiding problems that arise from machine learning on high dimensional data, such as large computation times and resource needs. The number of features used in an implementation of this fault detection system will thus be based on the desired trade-off between improved fault detection performance and data collection and computation costs.

5.5 Conclusion

The methodology proposed in this paper to incorporate expert knowledge by creating new features from existing sensor data enables higher classification scores and improved detection of faults while using fewer features — an improvement in the F1 score of almost 20% while using the same number of features, up to 27% with more features, and even an increase of 13% is possible while using only 75% as many features as in the original feature set. These improvements are seen using only basic, general knowledge of wind turbines that is not specific to the particular wind turbine installation.

The use of fewer features helps address problems that arise from a large number of dimensions, while saving data collection costs. Another benefit is that this approach allows for flexibility in the choice of machine learning algorithm because the data sets are in a format suitable for input into a wide variety of machine learning algorithms. Future work can explore the use of this methodology of incorporating domain knowledge using other machine learning algorithms.

Another avenue of investigation is to try other feature selection methods, particularly feature selection methods that take into account the correlation between features. Future research is also needed to explore how features selected by feature selection methods may be dependent on the faults present in the data. To investigate this behaviour, feature selection can be conducted on individual types of faults to investigate which features are good predictors for which faults. Additional research is also needed using more faults and different faults.

The methodology presented in this paper can also be applied to fault detection in other applications, such as other turbines, solar panels, and buildings. It would be interesting to compare and contrast performance and adapt this methodology across applications.

Chapter 6

Building Energy Systems

6.1 Chapter Overview

This chapter demonstrates the proposed procedure on chiller plants. First, the processing of the data and computations ran on the data are explained. Then, results are presented and analysed.

6.2 Data

Data Collection

The specific building energy system under investigation is a chiller plant. These plants are used in the commercial and industrial buildings sectors. A central chiller plant is a large facility that provides centralized chilled water to cool large buildings and multi-building campuses.

Data for normal operations and for operations under fault conditions was generated using a physics-based model of a commercial building chiller plant. It is undesirable to introduce faults in the real plant to collect measured data because introducing faults would interfere with normal operations of the facility and damage the expensive equipment. The physics -based model was created and calibrated by a research partner using the Modelica modeling language within the Dymola modeling and simulation environment.

The specific facility is the chiller plant at the Molecular Foundry at the Lawrence Berkeley National Laboratory. The Molecular Foundry is a 6-floor nanoscience research facility cooled by a chiller plant that is operational 24 hours a day, seven days a week. The chiller plant has three chillers, two cooling towers, three primary chilled water pumps, three condenser water pumps, and one secondary chilled water pump. A photo of the facility is shown in 6.1.

The physics-based models were created using the Modelica modeling language within the Dymola modeling and simulation environment. A graphical overview of the model for the Molecular Foundry is show in Figure 6.3.

Figure 6.1: Molecular Foundry case study facility

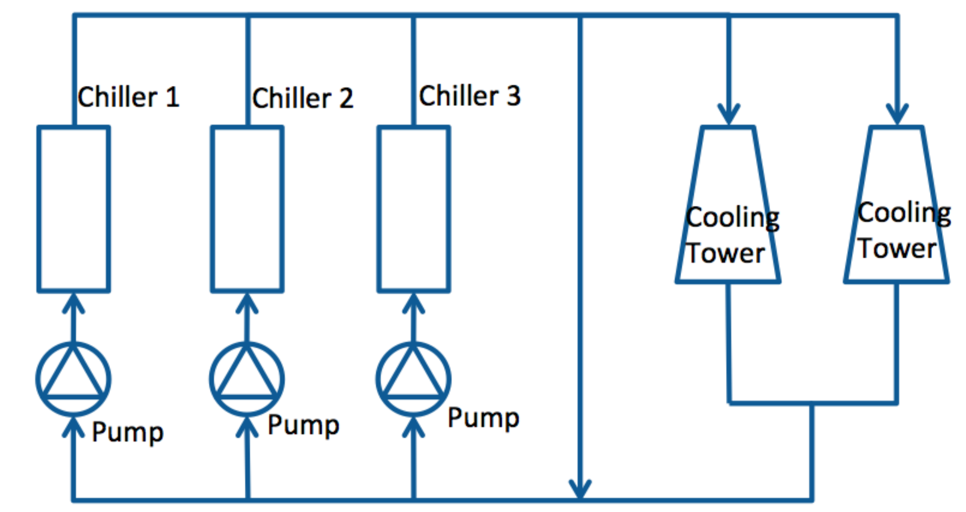


Figure 6.2: Schematic imaged of the condenser water loop of the case study facility

Simulations were run for two months under no fault conditions and for a year for each fault that is investigated. The chiller plant faults included in this data set are: *cooling tower fan failure*, *abnormal chiller cycling*, *primary pump failure*, and *secondary pump degradation*. While the types of faults present in the data affect the features that are selected in the feature selection process, the methodology presented here can be applied to other faults. The number of data points for each of the faults is summarized in Table 6.1.

From the simulation, measurements that correspond to available measurements common in real world chiller plant control systems were extracted at a commonly available sampling

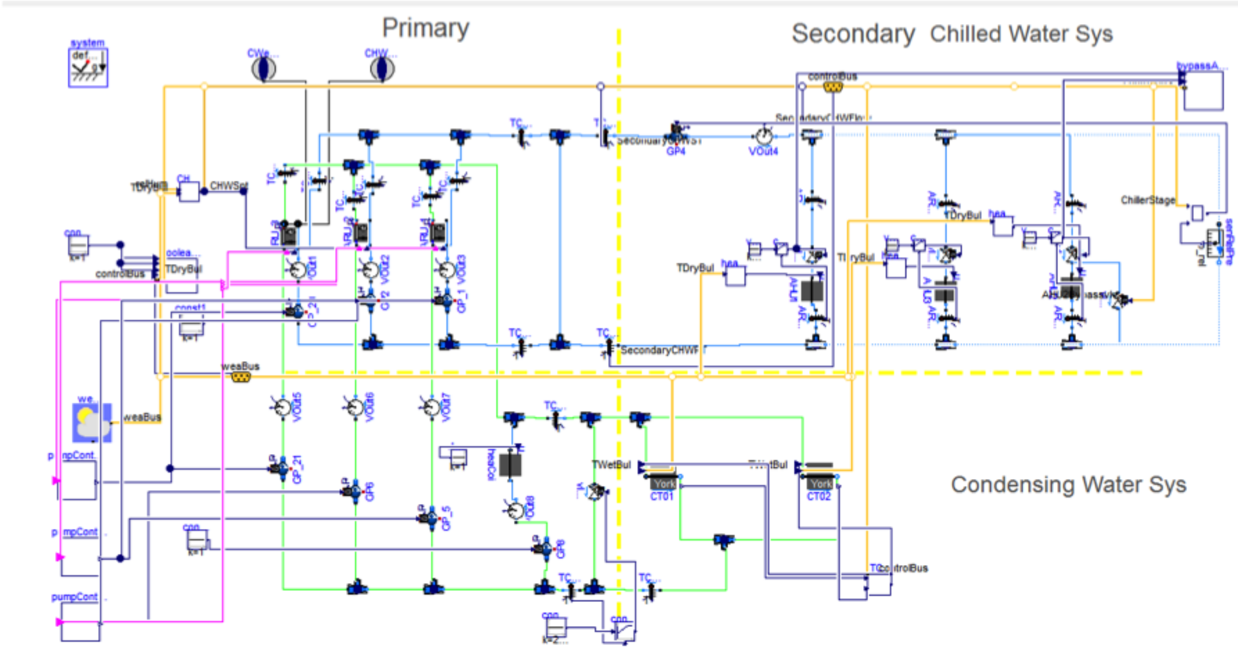


Figure 6.3: Visual overview of the physics-based model created using Modelica of the Molecular Foundry, which was one of the chiller plant models used to generate training data under various fault conditions

Table 6.1: Included Faults and Corresponding Frequencies

Fault	Number of Data Points
Cooling Tower Fan Failure	1254
Abnormal Chiller Cycling	17447
Secondary Pump Degradation	17518

rate. This results in 54 features with a sampling resolution of every 5 minutes. These features include control signals, temperatures, power, and On/Off status of equipment in the chiller plant, such as the cooling tower, chillers, and pumps.

Data Cleaning and Pre-Processing

To clean the data, samples that are in between the 5 minute sampling times are removed. Samples with the same timestamp are averaged.

While the simulation provides continuous values, in the building control system, some variables are only binary On/Off signals. To make the simulation reflect the available binary data in the building control system, for those variables, continuous values are converted to binary values.

Creation of Classification Labels

The labels for the samples are determined based on when equipment with a fault is operating. Samples obtained from running the simulation with no faults is labelled with the 'no faults' label. When the simulation is run with a fault, the resulting samples are labelled with the corresponding 'fault' class label when the equipment with the fault is On. Otherwise, samples are labelled as 'no fault'. For example, when a simulation is run with a mechanical failure fault in Pump 1, samples during which Pump 1 is On is labelled with the 'mechanical fault' label.

6.3 Methodology

Application Domain Knowledge Features

The original features in the BAS — Building Automation System — data are supplemented with new features created from the original features using knowledge of chiller plants, namely, knowledge of the quantities that the original features correspond to; the location of where data for those features is collected; and an understanding of the operation of the chiller plant. For example, the cooling tower range temperature is an important metric to track the performance of the cooling tower. Thus, a new feature is the difference between the temperature of the water approaching a cooling tower and the temperature of the water leaving a cooling tower. This and other instances of domain knowledge were acquired through interviewing building scientists and building managers, chiller plant operations and design training manuals, and basic science and engineering knowledge.

A summary of the new features derived from the original features using domain knowledge is shown in Table 6.2, which represents 67 additional derived features.

Time Series Features

The chiller plant is a physical system and it is known that the original features and derived features are physical quantities that form time series, hence timestamped data samples are not independent from one another. It is desirable to have the machine learning model work across time and to include correlation across sample columns. To that end, the data samples are converted to sliding time series representations of one hour using delay features.

To represent the data as time series, delay features for each original feature are created to include the data from time $t-n$ to t. For example, if the original feature "Chiller 1 Power" is at time t and the sampling resolution is 5 minutes, one new feature is "Chiller 1 Power at time $t-5min$", another new feature is "Chiller 1 Power at time $t-10min$", etc. all the way to "Chiller 1 Power at time $t-(n*5)min$". This results in 706 time-delay features.

The n, which is called the order of the lag, is selected using AIC and BIC. For each feature, the max order between AIC and BIC was selected. AIC consistently selected equal or higher orders than BIC. The maximum lag was selected for most of the features, When

Table 6.2: Example Features from Knowledge of Chiller Plants

Cooling Output	Total Plant Tonnage
	Cooling Tonnage for Each Chiller
Efficiencies	Kilowatt per Ton for Each Chiller
	Efficiency for Each Cooling Tower
Differences	Temperature Range for Each Cooling Tower
	Chilled Water Delta T for Each Chiller
	Condensed Water Delta T for Each Chiller

Table 6.3: Selected lag orders for some features. For features not shown, the selected lag order is 12

Feature	**AIC Order**	**BIC Order**
Total Tons	11	11
Chiller 1 Temp Setpoint	10	2
Chiller 1 Condenser Water Flow	12	11
Chiller 1 Leaving Condenser Temp	12	11
Chiller 2 Temp Setpoint	10	2
Chiller 3 Temp Setpoint	10	2
Chiller 3 Power	10	8
Chiller 3 On Off Status	12	7
Chiller 3 Condenser Water Delta Temp	6	12
Chiller 3 Leaving Chilled Water Temp	12	8
Cooling Tower 1 Fan Power	11	11
Primary Loop Chilled Water Supply Temp	12	6
Secondary Loop Pressure Difference	12	11
Secondary Loop Pump On Off Status	0	0
Secondary Loop Pump Control	12	11
Chiller 3 Condenser Water Pump Control	12	1
Chiller 3 Chilled Water Pump On Off	12	1
Relative Humidity	12	12

the max lag is not selected the resulting orders are close to the max. The variables for which the selected lags are not equal to the max lag are shown in Table 6.3, along with the selected lags by AIC and BIC.

Statistical Features

For each feature, the sliding mean and standard deviation is calculated. The window size for the sliding mean and standard deviation is the order as selected by Akaike information criterion (AIC) and Bayesian information criterion (BIC) in the previous step. Where the AIC and BIC selected an order of 0, a window size of 12 is used.

Data Cleaning and Standardization

Features that are constant are removed. Next, the data is standardized; The data is centered by removing the mean value of each feature, then scaled by dividing features by their standard deviation. This results in the data having a mean of zero and unit variance.

Feature Selection

Given that the number of features are greatly increased, the mutual information based minimal-redundancy-maximal-relevance criterion (mRMR) feature selection method is used to find a subset of features useful for prediction of the faults. mRMR reduces redundancy in the features and selects those most relevant to prediction [55]. Both Mutual Information Difference (MID) and Quotient (MIQ) schemes are used.

Sample Weights

The sample weights are the products between class weights and time weights. The most recent data is weighted twice as heavily as the oldest data, with weights for the other samples in linear proportion with time. Next, the weights are multiplied by balanced class weights to give sample weights.

Subsets for Training and Testing

When testing on time series, care must be taken to not use future data to predict past data. To this end, the following procedure is used to select training and testing subsets of the data.

The entire data set is sorted by ascending time and divided into twelve continuous time blocks of almost equal size. Twelve is selected because then each time block approximately corresponds to a month.

Then, five machine learning models are trained on the first m time blocks using 10-fold cross validation. From among the five machine learning models, the best performing model based on F1 score is tested on the $m+1$ time block. This training and testing process is done from $m = 1$ to $m = 11$ and the averages of the training and testing scores are calculated. An illustration of this process is shown below:

Train on [Jan] test on [Feb]

Train on [Jan, Feb] test on [March]

Train on [Jan, Feb, March] test on [April

etc.

Machine Learning

To detect a fault, the fault detection problem is framed as a two class classification problem, where the two classes are “fault” and “no fault”. The different faults are grouped together in the “fault” class.

To see if the new derived features improve fault detection performance, support vector machines (SVMs) are trained using the new feature set and the original feature set for different numbers of features. To train each SVM, a randomised grid search is performed over hyperparameters using 10-fold cross validation to find the hyperparameters which yielded the highest F1 classification score. The hyperparameters that are searched over are C, γ, and the kernel.

6.4 Results and Discussion

Selected Features

Table 6.4 lists the top ten features as selected by the MIQ and MID schemes to classify no-fault operations and fault operations. All the faults are grouped together. Because the two feature selection schemes use different measures of distance, some variation in rankings between the MIQ and MID schemes is expected. There is some overlap between the features selected under the MIQ and MID schemes.

In the feature selection process, the new derived features are favoured more than the original features. Under the MIQ scheme, all of the top ten features are new features, with one exception — Secondary loop pump power, which is ranked #8. All the top ten features under the MID scheme are all new features except for one feature — Cooling tower 1 control. The top ten features contain representatives from all the different types of new features: application domain knowledge features, time series features, and statistical features.

Figure 6.4 plots the number of new derived features among the features selected by mRMR under the MIQ scheme, along with their respective rankings. The dotted red line traces the expected curve if all the top features are new derived features.The deviations of the curves from the dotted reference line are the number of original features among the top selected features. Up to approximately the first 20 features of MIQ, the actual curves follow this reference line closely. This indicates that the top 20 selected features are almost all new features. In the top 50 selected features, there are 42 derived features and 8 original measurement features. This suggests that among the top features selected to maximize relevance while minimizing redundancy, most of the selected features are new

Table 6.4: Top 10 features selected for fault vs no fault by mRMR under the MIQ and MID scheme with feature rankings

Feature	**MIQ**	**MID**
Chiller 3 kilowatt per ton at t-40min	1	NA
Secondary loop pump power at t-10min	2	2
Condenser water pump 1 OnOff status at t-15 min	3	NA
Chiller 3 kilowatt per ton	4	1
Secondary loop pump power standard deviation	5	4
Chiller 3 kilowatt per ton at t-60min	6	5
Chiller 3 kilowatt per ton at t-10min	7	NA
Secondary loop pump power	8	NA
Chiller 3 kilowatt per ton at t-55min	9	NA
Chiller 3 kilowatt per ton at t-5min	10	NA
Chiller 1 OnOff status at t-60min	NA	3
Relative humidity t-20min	NA	6
Cooling tower 1 control	NA	7
Chiller 3 kilowatt per ton at t-30min	NA	8
Chiller 1 OnOff status at t-5min	NA	9
Secondary loop pump power at t-50min	NA	10

features. Under the MID scheme, more original measurement features are selected than under the MIQ scheme.

The top selected features are probabilistically related to the faults. However, the top selected features are not all directly at the fault location at time t. Many of the top selected features are delay features. The high ranking of expert engineering features agrees with field knowledge. For example, chiller kilowatt per ton is frequently recommended to monitor chiller performance. Chiller kilowatt per ton likewise ranks highly on this list, given that there are chiller degradation faults in the data set. However, it is interesting to observe that the delays of chiller kilowatt per ton are selected. This may be due to transients in the system causing a delay in time between the start of the fault and the sensors registering symptoms of the fault. For example, in the case of the chiller, temperatures in the chiller may raise for a few minutes before temperature sensors in the chiller register, the temperature raises above the deadband, the chiller control system responds, compressors start working harder, and power consumption increases.

The top selected features are engineering features that are calculated from several sensor measurements. For example, chiller killowatt per ton requires four input sensors. Thus, the use of engineering features fuses information from multiple sensors.

The selection of features is dependent on the faults in the data set. If a different set of faults is to be detected, the choice of features will very likely be different. Nonetheless, the results are promising in that nine faults representing four types of faults show fault detection

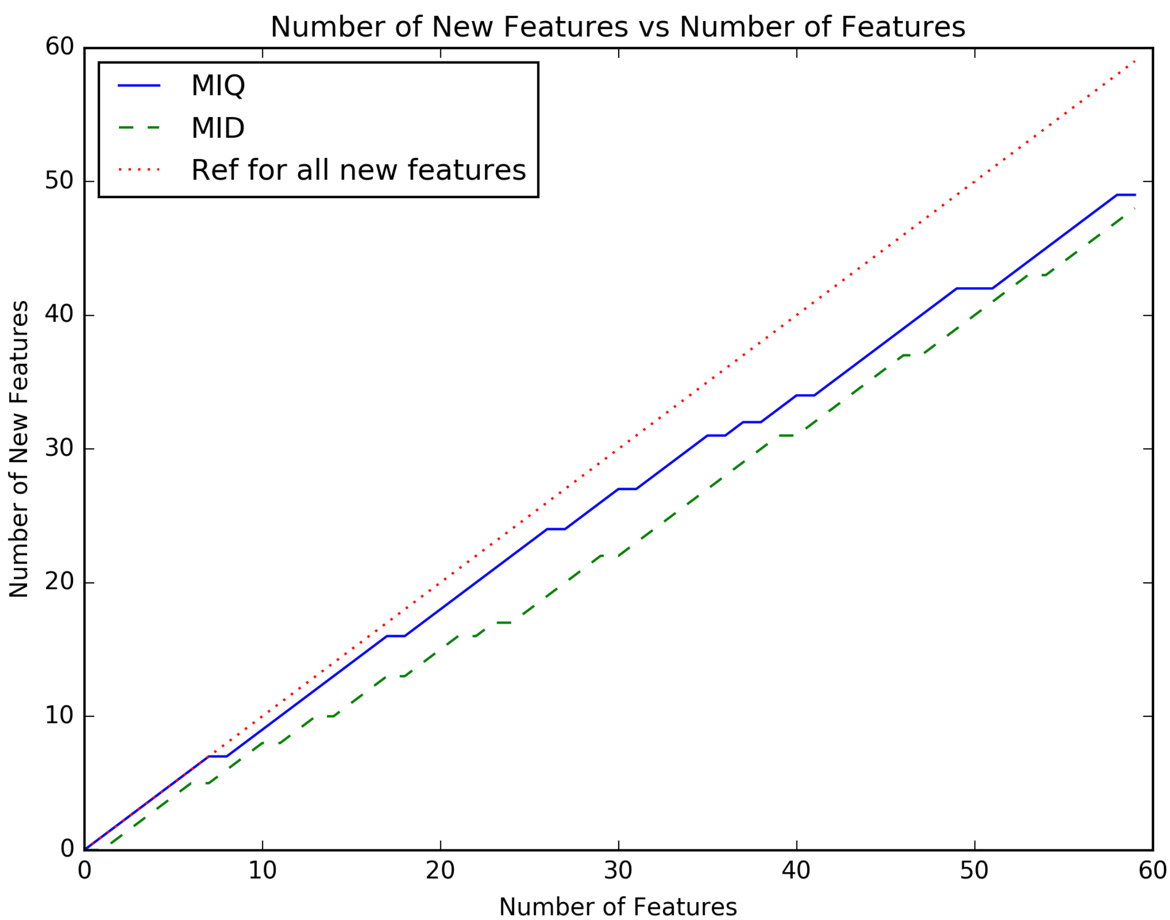


Figure 6.4: Number of new features among the top features with a dotted reference line indicating the expected line if all the features were new features

performance reaching high asymptotic values at four features, as will be shown in the next section.

Model Prediction Results

The use of new features and feature selection resulted in fault detection with high detection performance while only using a surprisingly small number of features.

Figure 6.5 plots accuracy metrics against the number of features. This is for detection of any of the faults in the data set, which is framed as a two class clasification problem, where the two classes are "fault" and "no fault". Figure 6.5 plots the F1 score on the first row of graphs, precision on the second row of graphs, and recall on the third row of graphs for the 'fault' class on the y-axes. The first column of graphs is the scores on the training set and the second column of graphs is the scores on the testing set. The x-axis is the number of features. Based on the higher scores under the MIQ scheme compared with the MID scheme, experiments were only run under the MIQ scheme.

With as few as four features, classification scores exceed 90% on the training set and 80% on the testing set. The original measurement feature set has 54 features, which is significantly more than the number of features required under this methodology. Recall scores are consistently above 90% starting at one feature. The precision increases with increasing number of features, which in turn improves the F1 score. This is because the F1 score is a weighted sum of precision and recall.

All scores asymptote after four features. The asymptotes on the training data is approximately 95% for all the performance metrics. The asymptotes on the testing data is approximately 80% for precision, 90% for recall, and 85% for F1 score. The difference between performance on the training set versus the testing set suggests that the fault detection model is slightly over-fitting to the training data.

Misclassification error rate on the training set can be traded-off with misclassification error rate on the testing set. That is, the model can be trained less to the training data and a higher misclassification rate on the training data can be tolerated in exchange for hopefully lower misclassification rate on the test set. The training set was chosen to be historical data in relation to the testing set. Thus, the testing set likely deviates from the training set due to changing weather conditions impacting cooling load demand and chiller plant performance.

To detect a larger number of different faults, however, the number of features needed for fault detection may be higher than as shown in these results. Different faults at different locations within the chiller plant will likely require features at the different locations. Different faults within the same location may also need additional features to distinguish between the faults. Nonetheless, the proposed methodology allows for user choice in the types of faults to focus on. Also, this methodology enables a trade off between detection accuracy and number of types of faults with data collection and computation costs. Real sensor data points can be supplemented with virtual sensors data points and be processed in workflows used to create fault detection models and machine learning models.

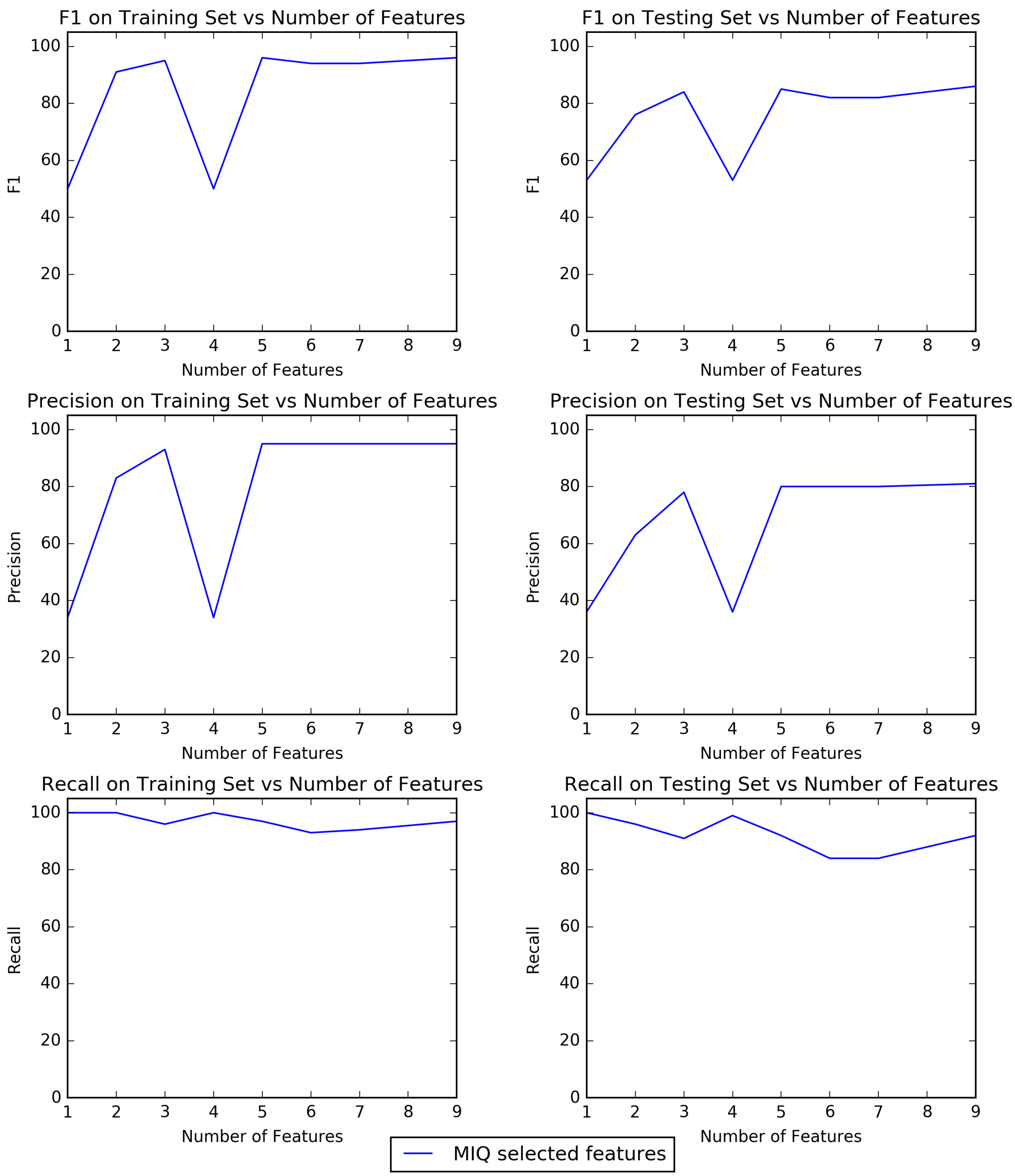


Figure 6.5: Scores on training and testing sets under the MIQ scheme

6.5 Conclusion

The large number of sensor points, dozens and hundreds, can be a dauntingly large number to track for performance monitoring. However, the above analysis shows that faults can be detected with high accuracy scores using a few select sensors. performance plateaus with as few as four features. The top selected data points, while they are probabilistically connected to the location of the fault, are not the data points that are directly associated with the fault. For example, the top few features are delay variables, or are from data points on pieces of equipment connected to the equipment with the fault. This may be due to the selected data points being binary variables instead of continuous variables, and thus binary variables will have a more distinct boundary.

The above analysis demonstrates how data from real sensors can be merged with expert knowledge 'virtual sensors' that contain understanding of the behaviour of the system and the relationships between the real sensors. Use of the virtual sensors can increase the performance of fault detection, as shown in this case study.

Additional research is needed to examine how the number of features required to reach asymptotic fault detection performance scales with more faults.

This chapter and the previous chapter demonstrated machine learning and applied statistics to detect faults with application to wind turbines and commercial building chiller plants to improve energy performance. The next chapter applies mathematics to residential buildings to save energy. To optimize energy use, mathematical optimization schedules energy consumption in residential buildings. The results of the analyses can inform decision-making and policy.

Chapter 7

Energy Scheduling

7.1 Chapter Overview

This chapter describes the background and modeling. Then, the model components are exercised and the results analysed to aid decision-making and policy-making. This chapter is based on [39]. Previous publications by the author are [38] and [63].

7.2 Background

The smart grid is a cyber-physical system for which there is an enormous opportunity to improve energy efficiency and reliability through big-data.

In the field of energy management, the current trend of increasing integration of Information and Communication Technology has enabled market actors to develop technologies that has engendered the electric grid as a cyberphysical system — the smart grid. This enables the development and use of applications such as the smart meter, bidirectional communication, advanced metering infrastructure (AMI), home automation, and home area networks [44]. Measurement and recording of electricity consumption and two-way communication between the meter and the utility's system can help adjust energy consumption patterns, and thereby achieve economic benefits.

Innovations in "Internet of Things" (IoT) devices have further led to connected power meters, lights, occupancy sensors, electric vehicles, appliances, and household electric battery storage that are capable of data collection and communication. The realization of monetary and energy benefits also necessitates the efficient performance of algorithms on these large data sets with numerous decision variables. This is of particular importance in order to scale from electric grid technologies from one user, such as a residential energy customer, to multiple customers, commercial customers, communities, and beyond.

The electric grid as a cyber-physical system results in the creation of large data sets, including for example, electric meter readings, electricity end-use consumption, electricity price signals, consumer preferences, and device control signals, among others. Using this data, big data methods such as mathematical programming can manage electricity demand more efficiently and assess the likely sources of value to be derived from smart grid technologies under electricity programs such as demand response.

To meet energy needs on the grid, demand response is a valuable resource because it can help reduce the volatility of electricity prices, mitigate market power of generators, and enhance grid reliability. This data includes electricity prices, electric meters, electricity end-use power consumption, consumer preferences, and control signals, among others. A valuable resource to meet energy needs is demand response,

Demand response achieves these objectives by lowering the peak demand for energy, which reduces the need to construct new and expensive generation units, and by providing ancillary grid services such as regulation and reserves to reliably integrate variable resources such as renewable generation [23]. In 2013, the potential contribution of demand response resources in the United States was 28,798 MW in Regional Transmission Organization (RTO), Independent System Operator (ISO), and Electric Reliability Council of Texas (ERCOT) markets [22], which represents an increase of 5.9% since 2009 [23]. Note that these values have fluctuated recently, due to economic impacts on electricity consumption.

Demand response can be defined as "changes in electric use by demand side resources from their normal consumption patterns in response to changes in the price of electricity, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized" [21]. Demand response technologies can be and are considered as components within a broad range of energy supply scenarios [23]. Correspondingly, there are significant policy implications associated with the development of demand response technologies (DR technologies).

While DR technologies are already prevalent in the commercial and industrial sectors of the U.S. economy, the market for DR technologies in the residential sector is currently in a nascent stage [23]. Accordingly, this paper examines the potential economic benefits associated with the application of DR technologies in the residential sector. Specifically, this paper offers (1) an optimal load shifting (OLS) model that can manage electricity demand more effectively; and (2) an assessment of the most likely sources of value to be derived from DR technologies.

The OLS model is formulated to be utilized by a Home Energy Management System (HEMS), which operates as a central command system for electricity usage inside a home, for components such as household appliances and air conditioning systems. The purpose of the OLS model is to obtain greater efficiency and savings, both in terms of economics and energy usage. Specifically, the model achieves savings by shifting electric loads to flatten the electricity load shapes of a particular household. Therefore, the adoption of such a model by multiple households can improve the efficiency of managing energy demand that may result in significant cost savings for consumers.

Our approach has been guided by recent academic literature, which found substantial economic benefits derived from demand response and other smart grid technologies. Chen, Wei, and Hu create an energy efficient scheduling algorithm that takes into account uncertainties in household appliance operation time and intermittent renewable generation, variable frequency drives, and capacity-limited energy storage, which may decrease monetary costs by up to 45% [8]. Mohsenian-Rad, Wong, and Jatskevich applied game theory to demand-side energy management between a utility company and its customers, showing a potential reduction in peak-to-average ratio of the total energy demand, total energy costs, and individual users' daily electricity charges [52]. In one case study, Malik and Bouzguenda [50] conducted a cost-benefit analysis and concluded that the long-term load management benefits of the smart grid outweighs the upgrade costs needed to create a more intelligent grid. Additionally, Faruqui, Harris, and Hledik wrote that the maximization of economic savings requires the use of both smart meters and dynamic pricing. They estimated that the present value of savings in peaking infrastructure can be as high as 67 billion [19].

This paper addresses several principles from the development of the model and from observations of current trends in the sector. First, we propose an intelligent algorithm that can effectively optimize energy consumption. In addition to reducing peak energy usage, we identified an economic interest that is advanced by implementing DR programs: the utility may be able to manage an electricity supply shortage in a more expeditious manner, with lower costs and less delay time.

Looking forward, the current business landscape indicates that HEMS technologies are set to expand with the prospect of both mature and start-up companies poised to enter the market. Importantly, it is predicted that the electric vehicle (EV) industry may play a critical role in expanding the HEMS market in the future. Recently, load shifting for EVs has become increasingly popular to respond to the significant energy demand created when multiple EVs charge during peak times. Through the use of an optimal decentralized protocol for EV charging, utilities can alleviate the heavy burden on the grid during peak hours by shifting the EV charging loads [2, 26, 66]. Similarly, the model aims to demonstrate the benefits of such load shifting with respect to a smart grid and the use of appliances inside the home.

Although the wide adoption of HEMS technologies is promising, the current academic literature has investigated the benefits of HEMS technologies as it is now, and the results are not entirely positive. For instance, Dam, Bakker, and Buiter [12] performed an overall life cycle impact on three HEMS that shows net energy savings over five years, but a negative return on investment in terms of monetary cost. However, their study was based on existing HEMS that did not use optimization processes and are less intelligent. Further, Hargreaves, Nye, and Burgess [35] reported on a year-long, in-home smart energy monitor trial, where results leave a homeowner more knowledgeable about reducing consumption, but not necessarily more motivated to do so. Given these results, the creation of a more intelligent HEMS helps address the negative findings of these works [11, 34]. Our model leverages algorithms that support optimization, which can result in a positive return on investment in terms of monetary cost. In addition, the intelligent HEMS could run automatically, without any reliance on the personal motivation of homeowners to reduce consumption.

Essentially, this paper offers three main contributions. First, the implementation of the model shows that the consumer can accrue economic benefits by shifting consumption away from higher-priced periods. Importantly, the model has a modular design that is flexible and can be configured to accommodate added capabilities. The model consists of the following four modules: load shifting, thermal control, battery electricity storage system, and automated windows. It is concluded that potential economic savings exist; however, further research is needed to apply the model to additional data sets that represent greater diversity in terms of pricing schemes, climate conditions, and consumer preferences.

Second, the findings are promising for the continued development of more intelligent electricity management in the residential sector. For example, the model is capable of performing quickly, which can be a highly preferred attribute of any algorithm that is designed to be implemented by a HEMS for real-time dynamic management. Specifically, the flexibility of the model is demonstrated in the results section.

Third, we develop a smart grid valuation framework that can be used to interpret the results produced by the model with respect to the efficiency of smart appliances and their respective prices. The framework is centered on four key questions: (1) which smart appliance provides the greatest overall savings?; (2) which smart appliance provides the greatest incremental savings?; (3) which smart appliance has the highest benefit/cost ratio?; (4) what incentives do smart appliances provide for behavioral changes?

Results show that appliances can save between 36-69% of the energy cost compared to a non-energy scheduled appliance. Also, payback periods for smart energy appliances can range from 2 years to 60 years depending on the appliance.

Finally, this paper concludes with a discussion about the model's benefits, the major concerns associated with its use in the real world, and possible future areas of investigation.

7.3 Optimal Load Shifting Model Formulation

This section describes the Optimal Load Shifting (OLS) model, which is a conglomeration of modules that are linked by the need to minimize the total cost and to manage the total household electricity use. The modules represent various appliances in terms of their physical behavior, end-user preferences, and automated controls. The automated controls are optimized to minimize costs of energy consumption, while obeying the physical characteristics of the appliances, the end-user preferences, and a limit on the total household electrical load.

The OLS model minimizes electricity consumption costs with perfect foresight over a limited time horizon (e.g. one week). While the nominal objective function is to minimize the total cost of electricity, potential additional terms in the objective could reflect end-user convenience in the following ways:

- coordination of stakeholder benefits (e.g. with the electric utility)
- aspects of an aggregator providing services (if present), and
- a time-sensitive mix of preferences for multiple household residents.

Because of its modular formulation, an agent can chose which modules to include or dismiss when running the OLS model. The five modules of the OLS model are:

- LS – Load Shifting
- TC – Thermal Control
- BESS – Battery Electricity Storage System
- AW – Automated Windows
- NS – Non-Shiftable

The above acronyms will be used in the following as a short form to refer to individual modules.

The modules are mathematical programming optimization problems. The Load Shifting module is a linear program (LP), while the other modules are mixed 0-1 linear programs. The theoretical computational complexity of solving a continuous LP is polynomial, and LP algorithms are fully implemented into mature commercial software, e.g., Cplex. GUROBI, MOSEK, etc.. Today, a practical linear program with millions of decision variables and constraints can be solved in minutes with a PC. Many open source codes are available with slightly worse performance. The mixed 0-1 LP is theoretically proven to be NP-hard in the worst-case, but in practice many of them can be solved efficiently. For example, the 0-1 linear programs described in our paper with thousands variables and constraints were all solved on a single machine in less than a second. Typically, the mixed 0-1 LP solvers (also developed inCplex. GUROBI, MOSEK, etc.) use LP solvers as subroutines, and intelligently solve a sequence of LPs to lead to an optimal solution.

The following sections describe each module's control/decision variables, fixed data parameters, constraints, objective function, and its mathematical optimization formulation.

Indices

The OLS model has indices for end-uses and time. The end-use loads are categorized according to the way they can be shifted or not. The shiftable end-use load categories are *work storage* and *energy storage*. The end-uses for work storage and energy storage have decision variables for shifting load. Non-shiftable loads are aggregated and do not require an index.

The notation in Table 7.1 describes the general OLS model indices for shiftable loads and time. It gives the index name, the valid range of values, and a short description. These indices are utilized in all of the modules.

Table 7.1: Descriptions of OLS model indices

Index	Range	Description
i	$1 \leq i \leq m$	m end-uses indexed by i that are work storage, e.g. clothes washer, dishwasher, and clothes dryer.
j	$1 \leq j \leq n$	n end-uses indexed by j that are energy storage, e.g. heating, AC, refrigerator, and freezer.
t	$1 \leq t \leq h$	h intervals indexed by t, e.g. hours, 5 minutes, minutes.

Load Shifting Module

The load shifting (LS) module shifts the loads of work storage end-use units. Work storage end-use units have a preferred duration and time window for energy use as determined by user preferences, thereby allowing the optimization to advance or delay the work. Examples of work storage end-uses are clothes washers and dishwashers.

Decision Variables

The control/decision variable x_i is the time indicator of the operating schedule for appliance i. It is a vector of dimension h where its tth entry equals 0 when the appliance does not operate at time interval t and 1 when the appliance operates; it is described in Table 7.2. It is assumed that the appliance will finish its task without interruption. An explanation of enforcing running without interruption is located in the appendix.

Table 7.2: Descriptions of LS Module variables

Control Variable	Units	Description
$x_{i,t}$	*none*	Schedule for work storage end-use i for interval t

Data Parameters

The LS module parameters appear in the major categories of user preferences, environment data, and end-use characteristics. Raw data sources must be processed into the formats of these parameters for proper use in the OLS model. Table 7.3 provides the specific details.

Table 7.3: Descriptions of LS Module parameters

User Preference	Units	Description
$D_{i,t}$	{0, 1}	Indicates allowed operation of end-use i in interval t
Environment Data	**Units**	**Description**
P_t	\$/kWh	Electricity price in interval t
L_t	kWh	Fixed load in interval t
A_t	kWh	Energy Cap for interval t
End-UseCharacteristic	**Units**	**Description**
Z_i	*none*	Number of intervals of consumption for end-use i
l_i	kWh	Average energy use for end-use i, when active

To aid in the understanding of this module's formulation, a work storage end-use units is introduced: a clothes washer. The washer is defined as end-use $i = 1$. Let the interval t be hours with $h = 24$ hours. Throughout this section, assumptions will be added to this example to explain the corresponding parts of the module.

Work Storage End-Use Illustrations

Assume that the washer consumes 0.3 kWh. Then, the average electricity use, l_i, is defined as a scalar for the energy consumption for end-use i if it were to run for an interval. It is assumed that all appliances, i, have constant energy use, in order to simplify the exposition. Therefore, $l_1 = 0.3$.

The indicator of allowable operation, D_i, is set by the homeowner. Assume the washer is allowed to operate during the time period between 12 a.m. to 8 a.m., and the model's first interval begins at 12 a.m., then the vector of dimension-24 for allowable operations is defined as

$$D_1 = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ \ldots\ 0].$$

The control/decision variable x_i, for end-use unit i, is a decision vector for the model. If the optimal time for the washer to run is to start at 3:00 a.m. and finish before 4:00 a.m., given $Z_1 = 1$, and the model starts at 12 a.m., then the entries of the vector will take values:

$$x_1 = [0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ \ldots\ 0],$$

where the location of the value '1' corresponds to interval 3:00 a.m to 4:00 a.m. and all other hourly values have zero value.

Work storage job completion

Below the constraint for job completion in the module is explained. Assume the washer runs for a full hour. If the washer (end-use unit $i = 1$) must operate one time from 12 a.m. to 8 a.m., then a condition is imposed that the sum of all hourly products of schedule operations and allowable operations equals Z_1, as a linear equation

$$(x_1 \cdot D_1) = Z_1, \text{ that is, } \sum_{t=1}^{h} D_{1,t} x_{1,t} = Z_1.$$

The right-hand side corresponds to the required number of operating intervals (hours) over all of the allowable intervals specified by user preference.

LS Module Formulation

The cost of electricity for end-use unit i can be expressed as

$$l_i(P \cdot x_i) = l_i \left(\sum_{t=1}^{h} P_t x_{i,t} \right),$$

where the dot operation $(\cdot)$ represents the inner product of two vectors.

Combining the above constraints and cost, the LS module takes the form:

$$\min_{x_i} \quad \sum_i l_i(P \cdot x_i) \tag{7.3.1}$$

$$x_i \cdot D_i = Z_i \;\forall i, \tag{7.3.2}$$

$$\sum_i l_i x_{i,t} + \Omega_t \leq A_t \;\forall t,$$

$$x_{i,t} \in \{0, 1\} \;\forall (i, t),$$

where Ω_t will represents other module loads in interval t.

Thermal Control Module

The Thermal Control (TC) module acts as a type of energy storage that can consume energy earlier or later, thereby allowing for the consumption of energy at less-costly times. Examples of thermal energy storage end-uses are heaters, air conditioners, combined heater and air conditioners (HVAC), and refrigerators. These appliances must maintain internal temperatures within user-specified ranges.

Typically, the control/decision variable for an energy storage appliance is the energy consumption per time period over the optimization horizon, and it affects how much heating or cooling is supplied by the appliance. At the same time, some tasks can be interrupted and then continued with little impact. This latter source of flexibility is not currently present in the OLS model but could be added at a later time.

Table 7.4: Descriptions of TC Module variables

Control Variable	**Units**	**Description**
$q_{j,t}$	kWh	Energy consumption of the thermal control end-use unit j for interval t
State Variable		
$T^{in}_{j,t}$	° F	Inside temperature of the space that is heated or cooled by end-use j for interval t

Decision Variables

An HVAC is used to illustrate the TC module formulation. The control/decision variable is the energy $q_{j,t}$ supplied to the HVAC unit j in interval t, and the state variable is the indoor temperature $T^{in}_{j,t}$. They are summarized in Table 7.4.

Data Parameters

The TC module parameters appear in the major categories of user preferences, environment data, end-use characteristics, and appliance limitations. Raw data sources must be processed into the formats of these parameters for proper use in the OLS model. Table 7.5 provides the details.

Table 7.5: Descriptions of TC Module parameters

User Preference	Units	Description
$T_{j,t}^{min}$	° F	Minimum temperature for end-use j in interval t
$T_{j,t}^{max}$	° F	Maximum temperature for end-use j in interval t
Environment Data	**Units**	**Description**
P_t	\$/kWh	Electricity price in interval t
$T_{j,t}^{ext}$	° F	External temperature for end-use j in interval t
A_t	kWh	Energy Cap for interval t (restated)
End-UseCharacteristic	**Units**	**Description**
$q_{j,t}$	kWh	Electricity consumption for end-use j and interval t
C_j	kWh/° F	Heat capacity for end-use j
K_j	kWh/° F	Thermal conductivity for end-use j
e_j	*none*	Energy conversion efficiency for end-use j. For thermodynamic direction of heat flow, this value is positive for heating and negative for cooling.
Appliance Limitations	**Units**	**Description**
$q_{j,t}^{max}$	kWh	Energy consumption limit for end-use j for interval t

Note that because of the treatment of the efficiency e_j being positive or negative for heating and cooling, respectively, an HVAC requires two end-use indices to discern this difference and that optimality conditions ensure that only one use is active in each period, t.

Indoor Temperature Range and Thermal Energy Control Limit

It is assumed that the electric heating end-use j, must maintain temperatures between a maximum and minimum. One can think of this as a resident's preferred comfort zone.

$$T_{j,t}^{min} \leq T_{j,t}^{in} \leq T_{j,t}^{max}, \ \forall (j,t)$$

Assume t represents hours, and that the indoor temperature $T_{j,t}^{in}$ is dependent upon the energy control variable ($q_{j,t}$) for the HVAC, which is defined as end-use unit $j = 1$, and this energy quantity must be less than a maximum value, $q_{j,t}^{max}$, for each end-use unit j for all intervals t. Then, the following relation is obtained for limiting the range of the thermal energy control:

$$0 \leq q_{j,t} \leq q_{j,t}^{max}, \ \forall (j,t).$$

Thermal Storage State Equation

The actual heating (or cooling) supplied by the HVAC is typically less than $q_{j,t}$ due to losses, and it would be $e_j q_{j,t}$ for all t, where efficiency parameter e_j always has a value between -1 and 1.

Further, the thermal equilibrium is also influenced by heat flow from the surroundings, $q_{j,t}^{sur}$, through a thermal insulation model for unit j. Its relation with the given exterior temperature, $T_{j,t}^{ext}$, and state decision variable, $T_{j,1}^{in}$, is given as

$$q_{j,t}^{sur} = K_j (T_{j,t}^{ext} - T_{j,t}^{in}), \ \forall t \neq h,$$

where K_j is the thermal conductivity coefficient of unit j.

Thus, by conservation of energy and thermodynamics, the temperature difference from one interval to the next must be proportional to the energy exchanged with the HVAC and the surroundings, that is,

$$\begin{aligned} C_j (T_{j,t}^{in} - T_{j,t-1}^{in}) &= e_j q_{j,t-1} + q_{j,t-1}^{sur} \\ &= e_j q_{j,t-1} + K_j (T_{j,t-1}^{ext} - T_{j,t-1}^{in}), \ \forall (j,t) \end{aligned}$$

where C_j is the thermal heat capacity coefficient of unit j, and $T_{j,0}^{in}$, together with exogenous data parameter $T_{j,0}^{ext}$, are part of given data parameters.

TC Module Formulation

The corresponding cost of energy for the HVAC, defined as end-use unit j, is

$$P \cdot q_j = \sum_{t=1}^{h} P_t q_{j,t}.$$

Combining the above constraints and cost, the TC optimization module takes the form:

$$\min_{(T_j^{in},\ q_j)} \quad \sum_j (P \cdot q_j) \tag{7.3.3}$$

$$C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) = e_j q_{j,t-1} \tag{7.3.4}$$

$$+K_j(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \ \forall (j,t), \tag{7.3.5}$$

$$T_j^{min} \le T_j^{in} \le T_j^{max} \ \forall j,$$

$$0 \le q_j \le q_j^{max} \ \forall j,$$

$$\sum_j q_{j,t} + Omega_t \le A_t \ \forall t.$$

where T_j^{min} and T_j^{max} are given, and Ω_t will represents other module loads in interval t.

Battery Electricity Storage System Module

The Battery Electricity Storage System (BESS) module represents how a battery can store energy during intervals when electricity prices are low and provide electricity for household appliances when electricity prices are high. This may be beneficial for those non-shiftable loads that occur in peak hours.

Decision Variables

The decision variables for the module are b_t, the charging rate at interval t, and b_t^{dis}, the discharging rate at interval t. The variables indicate the amount of that energy the battery consumes during charging and the amount of energy the battery supplies during discharging. The state variable, s_t, is the amount of energy stored in the battery at the end of interval t. They are summarized in Table 7.6.

Table 7.6: Descriptions of BESS module variables

Control Variable	**Units**	**Description**
b_t	kWh	energy consumption for the battery for interval t
b_t^{dis}	kWh	energy provided from battery for interval t
State Variable		
s_t	kWh	amount of energy stored in battery for interval t

Data Parameters

The BESS Module parameters appear in the major categories of end-use characteristics and appliance limitations. Raw data sources must be processed into the formats of these parameters for proper use in the OLS model. Table 7.7 provides additional details.

Table 7.7: Descriptions of BESS module parameters

Environment Data	**Units**	**Description**
A_t	kWh	Energy Cap for interval t (restated)
End-UseCharacteristic	**Units**	**Description**
b^{eff}	*none*	battery efficiency between 0 and 1
Appliance Limitations	**Units**	**Description**
b^{lim}	kWh	maximum storage capacity of the battery
b_{max}	kWh	maximum charge rate of the battery
b^{dis}_{max}	kWh	maximum discharge rate of the battery

Typically, the physical battery that the homeowner purchases will provide the battery size, maximum charge rate, and discharge rate.

Battery Storage

The amount of energy stored in the battery for interval t is defined by the amount of energy from the previous interval plus or minus the amount of energy it consumes during charging or provides during discharging, between the previous interval and the current one. Due to inefficiencies, the actual energy consumed or provided by the battery is less than the amount of energy that the battery consumes or provides. This inefficiency is represented by the state equation constraint:

$$s_t = s_{t-1} + b^{eff} b_t - \frac{1}{b^{eff}} b^{dis}_t \ \forall t.$$

Furthermore, s_0 is a given parameter, representing the starting state of charge.

BESS Module Formulation

Note that the corresponding *net* cost of electricity for the battery is

$$P \cdot (b - b^{dis}) = \sum_{t=1}^{h} P_t (b_t - b_t^{dis}).$$

Combining the above constraints and cost, the BESS Module takes the form:

$$\begin{aligned} \min_{(s,b,b^{dis})} \quad & P \cdot (b - b^{dis}) \\ & s_t = s_{t-1} + b^{eff} b_t - \tfrac{1}{b^{eff}} b_t^{dis} \ \forall t, \\ & 0 \le s_t \le b^{lim} \ \forall t, \\ & 0 \le b_t \le b_{max} \ \forall t, \\ & 0 \le b_t^{dis} \le b_{max}^{dis} \ \forall t, \\ & (b_t - b_t^{dis}) + \Omega_t \le A_t \ \forall t, \end{aligned} \tag{7.3.6}$$

where Ω_t will represents other module loads in interval t.

Automated Windows Module

The previous modules have unique formulations, while the Automated Windows (AW) Module breaks that trend, because it is a variation of the TC Module. This section describes the treatment of automated windows that can be opened and closed in order to supplement the HVAC system with fresh air, especially when the external temperature is within the comfort range. The model treats the difference between opened and closed windows simply as a change in the thermal conductivity K_j defined earlier.

Modifying Thermal Control Module

The easiest extension to the TC Module formulation (7.3.3) would be making K_j into a decision variable where it could take on continuous values between a lower bound (closed window value) and upper bound (open window value). Although this modification is logically straightforward, it creates a difficulty for the integer progamming solver by making constraint (7.3.5), reproduced below, quadratic:

$$C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) = e_j q_{j,t-1} + K_j(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \ \forall (j,t),$$

since both K_j and $T_{j,t-1}^{in}$ are now decision variables.

Linear Relaxation

In order to alleviate this unwanted quadratic formulation, two modifications are made. The first is to assume that the window can only be completely opened or completely closed.

Therefore, this allows data parameter K_j to be either K_j^{cw}, representing the thermal conductivity when windows are closed, or K_j^{ow}, representing the thermal conductivity when windows are fully open. Note that if end-use j has no controllable windows, then $K_j^{cw} = K_j^{ow}$.

The second modification is to perform a linear relaxation on the quadratic formulation with the aid of the first modification. This is realized by creating a new binary variable w_t, which acts as an indicator of whether the windows are open or not, during interval t.

The first step in the relaxation is turning the equality constraint (7.3.5) into an equivalent set of two linear inequality constraints.

$$C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \leq e_j q_{j,t-1} + K_j(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \ \forall (j,t),$$

$$C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \geq e_j q_{j,t-1} + K_j(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \ \forall (j,t).$$

Then create a duplicate set of the two inequality constraints, and replace k_j in the first set with K_j^{cw} and in the second set with K_j^{ow}.

$$C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \leq e_j q_{j,t-1} + K_j^{cw}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \ \forall (j,t),$$

$$C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \geq e_j q_{j,t-1} + K_j^{cw}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \ \forall (j,t);$$

and

$$C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \leq e_j q_{j,t-1} + K_j^{ow}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \ \forall (j,t),$$
$$C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \geq e_j q_{j,t-1} + K_j^{ow}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \ \forall (j,t).$$

Only one set of the two inequality constraints should hold during each interval t. To ensure this, the model introduces the binary variable w_t along with a large scalar value S to the four inequalities above in a way that will tighten one set of inequalities (indicating that the corresponding thermal conductivity is being used) and relax the other set (indicating that the other thermal conductivity is inactive).

$$\begin{aligned}
C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \ &\leq e_j q_{j,t-1} + K_j^{cw}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \\
&\qquad + S w_{t-1} \ \forall (j,t), \\
C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \ &\geq e_j q_{j,t-1} + K_j^{cw}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \\
&\qquad - S w_{t-1} \ \forall (j,t);
\end{aligned}$$

and

$$\begin{aligned}
C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \ &\leq e_j q_{j,t-1} + K_j^{ow}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \\
&\qquad + S(1 - w_{t-1}) \ \forall (j,t), \\
C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \ &\geq e_j q_{j,t-1} + K_j^{ow}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \\
&\qquad - S(1 - w_{t-1}) \ \forall (j,t).
\end{aligned}$$

(The run times for scenarios of this formulation are described later.)

TC/AW Module Formulation

Thus, the combined TC and AW module formulation takes the form:

$$
\begin{aligned}
\min_{(T_j^{in},\ q_j,\ w)} \quad & \textstyle\sum_j (P \cdot q_j) && (7.3.7)\\
C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \quad & \le e_j q_{j,t-1} \\
& + K_j^{cw}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \\
& + S w_{t-1}\ \forall (j,t), \\
C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \quad & \ge e_j q_{j,t-1} \\
& + K_j^{cw}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \\
& - S w_{t-1}\ \forall (j,t), \\
C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \quad & \le e_j q_{j,t-1} \\
& + K_j^{ow}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \\
& + S(1 - w_{t-1})\ \forall (j,t), \\
C_j(T_{j,t}^{in} - T_{j,t-1}^{in}) \quad & \ge e_j q_{j,t-1} \\
& + K_j^{ow}(T_{j,t-1}^{ext} - T_{j,t-1}^{in}) \\
& - S(1 - w_{t-1})\ \forall (j,t), \\
T_j^{min} \quad & \le T_j^{in} \le T_j^{max}\ \forall j, \\
0 \le q_j \quad & \le q_j^{max}\ \forall j, \\
w_t \quad & \in \{0,1\}\ \forall t, \\
\textstyle\sum_j q_{j,t} + \Omega_t \quad & \le A_t\ \forall t.
\end{aligned}
$$

where again $T_{j,0}^{in}$, together with exogenous data parameter $T_{j,t}^{ext}$, are given parameters, and where Ω_t will represents other module loads in interval t.

Non-Shiftable Loads Module

The Non-Shiftable Loads (NS) Module implements the system's power constraint on the household appliances that are fixed and can not be shifted. It also reports the fixed cost of the non-shiftable loads. The module has no decision variables, and its purpose is strictly for accounting.

These fixed loads can be denoted by L_t, a fixed load in interval t. Thus, the (fixed) electricity cost from the NS load would be

$$P \cdot L = \sum_{t=1}^{h} P_t L_t.$$

The impact on other module formulations is that A_t, the energy cap for interval t, is simply replaced by $(A_t - L_t)$. It is assumed that the non-shiftable load does not exceed the energy cap, or that an internal energy source can keep this constraint feasible.

7.4 Data

The total size of the data depends on the number of energy end-uses, the data sampling rate, and the desired time resolution of the energy consumption schedule. For all the modules, the scheduling horizon length and the period length can be specified by two integers. The electricity price and energy capacity values are represented by two floats per time sample. The size of data for the modules is summarized in Table 7.8. The Automated Windows module requires the Thermal Control module as well. For the Load Shifting module, the non-shiftable load may require several data series of floats. This is because the non-shiftable end-uses likely require more than one meter. As a result, the non-shiftable load likely needs to be calculated as the sum of power consumption across the non-shiftable loads, which gives one float per non-shiftable end-use per timestamp.

Table 7.8: Data size growth rate per module

Module	Data per End-Use	Data Per End-Use Per Timestamp
Load Shifting	-	1 float, 2 binary
Thermal Control	3 float	3 float
Automated Windows	TC + 2 float	TC + 1 binary
Battery Electricity Storage	3 float, 1 int	1 float

One can see how the size of the data grows quickly when accounting for numerous end-uses across numerous houses.

7.5 Results

This section exercises the OLS model for five cases that vary by modules, end-use data sourcing, and form of electricity prices. Table 7.9 highlights the salient aspects of the data for each case. The appendix contains a section called "Data" that explains fully the data used to create each case.

The five cases vary across three locations. The Modules column indicates which modules of the OLS model are utilized in each case. The acronyms are defined in the Appendix. The Data column indicates whether the data is Real, from actual measurements, or Simulated, based on assumed appliance loads and end-use timings. The outside temperature and electricity prices are actual data. The column for Electricity Pricing indicates the type of prices used for that location. Time of Use (TOU) can have two or three pricing periods per day. Day Ahead prices are specified hourly on the day before use takes place.

Table 7.9: Descriptions of Cases

Case	Location	Modules	End-Use Data	Electricity Pricing	Time Step
1	Boston, MA	LS, NS	Real	Time of Use	60 min.
2	Springfield, IL	LS, TC, NS	Simulated	Day Ahead	5 min.
3	Springfield, IL	LS, TC, BESS, NS	Simulated	Day Ahead	5 min.
4	Springfield, IL	LS, TC, AW, NS	Simulated	Day Ahead	5 min.
5	Austin, TX	TC, NS	Simulated	Time of Use	5 min.

Case 1: Boston Load Shift

Although the data is based on actual measures of end-use, it lacks a substantial amount of information and points of interest that are desirable for testing the OLS model. The following list highlights its major limitations:

- It does not have the end-user's temperature preference.
- It does not indicate whether the house has an electric, gas, or hybrid HVAC system.
- Outside temperatures are close to the typical comfort range.
- It is not clear whether the available data indicates the home's entire electric usage.
- The Boston area implements only a peak and off-peak time-of-use pricing scheme.

Given these limitations, this data is used merely to test the model's ability to shift loads. The shiftable loads are HVAC, clothes washer/dryer, and dishwasher. In order to test the full model, assumptions are needed for the HVAC data, which are explained shortly. Nevertheless, the results are still promising.

The model is run once a day starting at 12 a.m., for seven days, without lookahead, which is unnecessary. Lookahead is unnecessary because conditions more than a day in advance do not influence the decisions made to schedule energy consumption that day. This is because constraints are defined within the day. Both the start time and the duration used in the OLS model are generally configurable and can be changed to meet user preferences. During the week tested, the OLS model shifted loads on four out of the seven days. The daily Original and Shifted electricity costs are listed in Table 7.10, with the total savings from load shifting being 4.63%, which appears in the second row of the last column.

Comparing results between the Original row and the Shifted row, observe that no shifting takes place on days 3, 5, and 7. The remaining load shifting occurred as follows:

- washer, dryer, and dishwasher on days 1 & 6,
- dishwasher on days 2 & 4.

Table 7.10: Case 1: Boston Load Shift. Daily electricity costs for 5/15/11-5/21/11

	Day 1	Day 2	Day 3	Day	Day 5	Day 6	Day 7	Total	**Savings**
Original	$1.41	$1.56	$1.10	$0.83	$0.83	$1.20	$0.84	$7.78	–
Shifted	$1.31	$1.47	$1.10	$0.77	$0.83	$1.10	$0.84	$7.41	4.63%

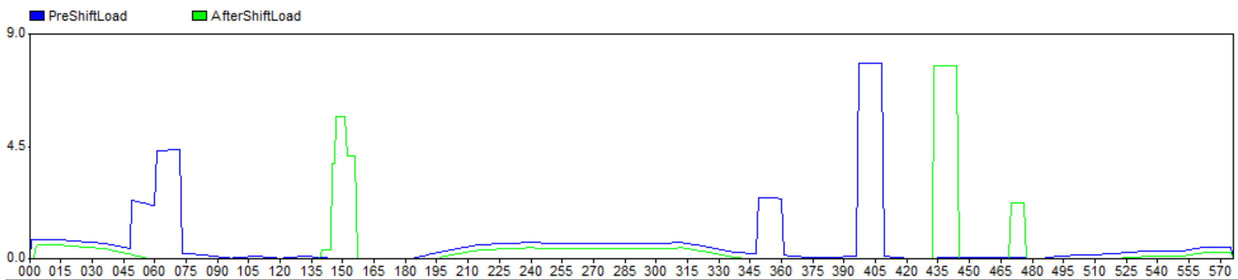


Figure 7.1: Case 2: Springfield without Battery. Original and shifted loads for 8/11 - 8/12/13

In the TOU pricing scheme, savings are limited. Loads are shifted from afternoon and evening times (peak hours) to early morning times (off peak hours). Because some appliances were already running in off-peak hours, savings are limited. In this data set, there is no indication of an HVAC system, which could be a major controllable load affecting energy costs and benefits. To overcome this, the other cases include HVAC use with thermal control.

To make up for the limitations of the Boston dataset, the OLS model is run on the fuller, simulated data for Springfield. Case 2 has only the load shifting and thermal controls; Case 3 adds a battery to the home; and Case 4 adds automated windows.

Solution times for the Boston Load Shift case are very fast. Because the OLS model is so simple, all CPU times are less than one second.

Case 2: Springfield without Battery

For the Springfield cases, the OLS model is run every two days, starting at 3 p.m., for a total of six days (three runs). Each run has no lookahead, because the storage is so small that it cycles each day.. Further, the end-user's preferred comfort range for the thermal controls is between 60 °F and 70 °F.

The shiftable loads in Cases 2, 3, and 4 are: HVAC, dishwasher, clothes washer/dryer, and PEV charger.

Figures 7.1 to 7.3 depict a time series for the original and shifted loads in kW for 5 minute intervals. The 0 interval is 3:00 p.m.

In Day 1 (8/11), the washer, dryer, dishwasher, and plug-in electric vehicle loads shift to the cheapest electricity price intervals, which is in the early morning. (The PEV was only driven a short distance on this Sunday, therefore the load for the PEV is very low compared

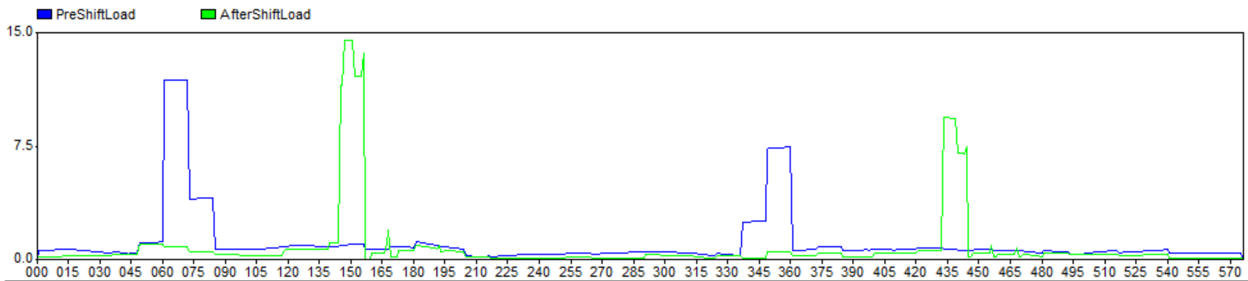


Figure 7.2: Case 2: Springfield without battery. Original and shifted loads for 8/13 - 8/14/13

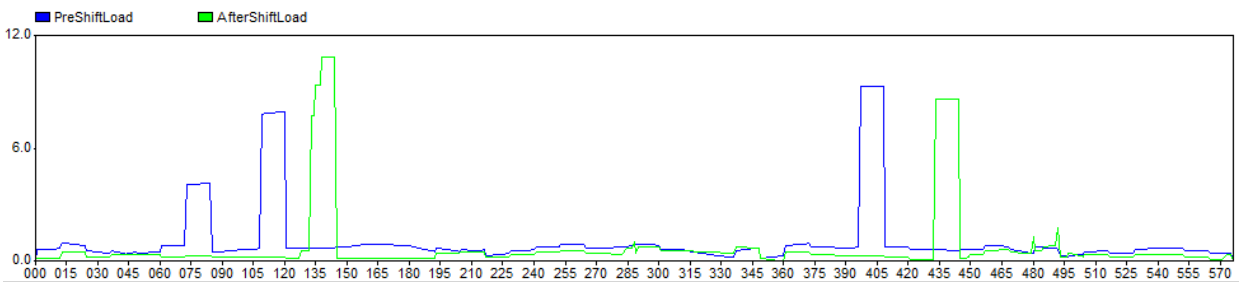


Figure 7.3: Case 2: Springfield without battery. Original and shifted loads for 8/15 - 8/16/13

to a weekday). In Day 2 (8/12), only the dishwasher and PEV loads shift. The PEV has a smaller preference window than the dishwasher; both items are shifted to the cheapest electricity price intervals for their respective preference windows.

In Day 3 (8/13), the washer, dryer, dishwasher, and plug-in electric vehicle loads shift to the cheapest electricity price interval in the early morning. In Day 4 (8/14), only the dishwasher and PEV loads shift. They are shifted to the cheapest electricity price intervals for their preference window.

In Day 5 (8/15), the washer, dryer, dishwasher, and plug-in electric vehicle loads shift to the cheapest electricity price time in the early morning. In Day 6 (8/16), only the PEV load shifts to the cheapest electricity price intervals for its preference window.

Note that end-uses are shifted to early morning hours, which is when electricity prices are lowest. Also, the persistent, small gap between the original and shifted loads stems from the energy saved due to the flexibility in the end-user's comfort range as opposed to having a single temperature setting.

Over the 6-day period, this simulated end-user realizes almost 40% monetary savings by utilizing load shifting and smart thermal controls, as summarized in Table 7.11.

Computation times for each 2-day sequence are listed in Table 7.12. The solution time is that spent on solving the optimization problem, while the total time includes setup and reporting.

Table 7.11: Case 2: Springfield without Battery. Savings for 8/11/13 – 8/16/13

Days	Days 1 & 2	Days 3 & 4	Total 5 & 6	Savings	
Original	\$0.95	\$1.39	\$1.21	\$3.55	
Shifted	\$0.55	\$0.80	\$0.80	\$2.15	39.44%

Table 7.12: Case 2: Springfield without Battery. Solution times in CPU seconds.

	Days 1 & 2	Days 3 & 4	Days 5 & 6
Solution Time (sec.)	0.80 s	1.00 s	0.97 s
Total time (sec.)	1.11 s	1.41 s	1.06 s

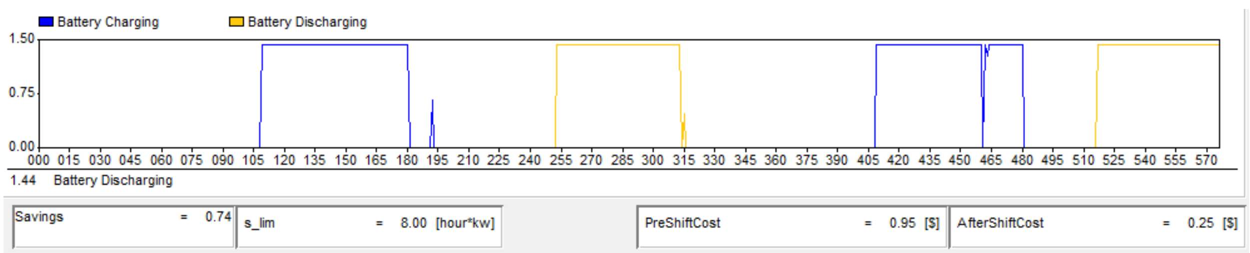


Figure 7.4: Case 3: Springfield with Battery. Original and shifted loads for 8/11 - 8/12/13

Case 3: Springfield with Battery

The Springfield results can be further improved by adding battery storage. A battery has the flexibility to charge when electricity prices are low and to let the end-user use the stored energy during high price periods. This effectively allows an end-user to shift their non-shiftable loads.

Further, the end-user may sell the electricity back to the utility during high-price times if the utility permits, as is possible in this case.

Figure 7.4 depicts battery charging and discharging (kW) in 5-minute intervals, where interval 000 is 3:00 pm. The battery has an 8.0 kWh capacity and a maximum charge and discharge rate of 1.44 kW. When adding this battery to the Springfield case for Days 1 and 2, the savings rises to 74%, which occurs because the battery charges during the low price periods (late night) and discharges during the peak periods (late afternoon).

The original electricity cost for Days 1 and 2 is \$0.95. After load shifting without the battery, the cost reduced to \$0.55, and after load shifting with the battery, the cost is \$0.25. Thus, with the battery, the end-user realizes over 70% savings for the two-day testing period.

While this is a simple example, use of the battery becomes more important when the household may reach its total electricity limit. In which case, energy in the battery can be utilized to reduce the cost of non-shiftable loads in high-priced hours, while keeping the total household load below the household limit.

The solution times are on the order of one second for the first two-day horizon. Therefore, adding the OLS battery storage module has not added much burden to the computation.

Case 4: Springfield Automated Windows

Automated windows could make thermal controls even smarter, because they add an ability to open and close the windows as a supplement to the HVAC system, but may be problematic because of security concerns. When utilizing the AW module, the thermal controls receive an extra operation to intelligently control the home temperature, but the Springfield case did not yield obvious economic savings. Days 1 and 2 resulted in the windows being opened for a single fifteen-minute period. This translated into a negligible increase in savings of 0.01% compared to only utilizing the LS and TC modules.

Alternative climate scenarios and locations must be tested to determine whether automated windows can make a meaningful impact on electricity costs.

Table 7.13 compares the solution and total CPU times across cases 3, 4, and 5 for the first two days of the horizon.

Table 7.13: Comparison of computational effort for days 1 and 2 in CPU seconds

Case	Solution time	Total time
2: Springfield without Battery	0.8 s	1.1 s
3: Springfield with Battery	1.0 s	1.2 s
4: Springfield Automated Windows	118.8 s	119.2 s

It is clear that adding the OLS automated windows module significantly increases the computational burden. This leads to the conclusion that there is a future avenue of investigation to explore alternate modeling and solution approaches in order to more efficiently determine how to automatically control windows.

Case 5: Austin Pre-Cooling

The Austin Pre-Cooling case demonstrates pre-cooling the household for a somewhat artificial case of allowing the indoor temperature range to vary more widely. It includes only the Thermal Control module of the OLS model and therefore the model minimizes only the costs related to the HVAC. The two time-series charts in Figure 7.5 show related temperatures and the operation of the HVAC over the model horizon (Tues - Thurs, 8/6/2013 3:00 PM to 8/8/2013 3:00 PM).

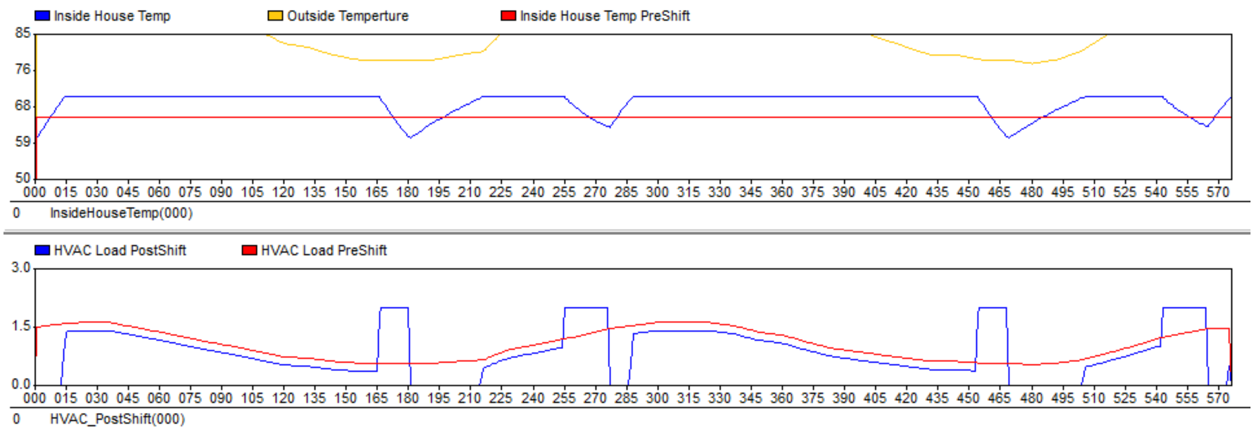


Figure 7.5: Case 5: Austin Pre-Cooling. Home temperature and HVAC loads for 8/6/13 - 8/8/13

These series are shown over identical time periods, and the inside temperatures and HVAC loads are shown before (pre-shift in red) and after (post-shift blue) load shifting. The yellow temperature series is the outside temperature, which is mostly synchronous with the HVAC loads, as expected.

The pre-shift indoor temperature is maintained within a narrow band at an almost constant level. The post-shift indoor temperature is maintained over a wider band in order to better demonstrate the operation of pre-cooling the household, in anticipation of higher electricity prices. In this case, pre-cooling takes place every time the TOU scheme changes from a lower price to a higher price. Over this two-day window, pre-cooling is exercised four times. There is no pre-heating, because the outside temperature is always higher than the indoor temperature range.

The Austin HVAC system has a 2 kW capacity, and for the periods directly prior to the electricity price change, the HVAC will work at this capacity to lower the indoor temperature to the lower bound of the preference range. The indoor temperature change is timed to hit the lower bound just as the TOU price rises. Instead of using energy shortly after this higher-price period begins, the OLS model turns off the HVAC and lets the indoor temperature rise until it meets the upper bound on the preference range. Then, it turns on the HVAC and keeps the temperature at the upper bound, which is the most efficient level, because it has the smallest difference with the outside temperature, which governs heat transfer through the building insulation.

Allowing a wider indoor temperature range obviously decreases energy costs by itself, because the indoor temperature is allowed to be higher than the nominal level for high outside temperatures. However, further savings are realized from the pre-cooling, within the wider allowable indoor temperature range. For the two-day period, the end-user realizes a 27% savings, where the HVAC cost without use of the TC module is $2.88 and with its use is \$2.10.

Since this version of the OLS model tests only the HVAC in order to demonstrate pre-cooling, the model formulation is a linear program, which leads to a solution time of less than one second.

7.6 Smart Grid Analytical Framework

The five sets of case study results provide useful insights about the valid operation of the OLS model and how overall optimization of the HEMS can provide end-user savings. This section describes a framework to further aid in interpreting the OLS model results. The OLS model is a tool that provides a way to study the fundamental behaviors of a HEMS before and after adding smart modules for various types of functionality, like load shifting, thermal controls, and battery storage. This Smart Grid Analytical Framework builds on the types of observations made possible by the OLS model to allow investigators to answer questions like:

- Which smart appliance provides the greatest overall savings?
- Which smart appliance provides the greatest incremental savings?
- Which smart appliance has the highest benefit/cost ratio?
- What incentives do smart appliances provide for enabling behavioral changes?

This section describes four analytical frameworks that can help answer these questions.

- *Smart Appliance Cost Analysis* - Compares actual appliance costs with and without communication and control abilities. This is useful for understanding the cost structure of smart appliances for investment purposes.
- *Smart Appliance Benefits Analysis* - Shows how to use the OLS model and cases to investigate appliance-level benefits. This is useful for understanding individual end-user decision-making.
- *Smart Appliance Marginal Benefit Analysis* - Shows how to use the OLS model and cases to investigate the marginal benefits of individual appliances. This is useful for understanding the benefits of scale and the potential decisions of an aggregator or electric utility.
- *End-User Preference Sensitivity Analysis* - This is useful for understanding the value of various rate designs and their potential benefits to end-users and other stakeholders.

Smart Appliance Cost Analysis

This Smart Appliance Cost Analysis compares the costs of actual appliances for which there are comparable smart and non-smart versions. A smart appliance is one that has the ability to be remotely controlled, and companies have been introducing recently these kinds of smart household appliances. However, their offerings are still few in number. In the following, a company's smart appliance was compared in price to a company's most-similar model without the smart characteristic. Due to the current nature of the market for smart appliances, they are all higher-end models, and it was not possible to locate lower-tier models with smart characteristics at this time. Further, not all appliance providers have smart appliances in each category. For instance, there appears to be only one smart dishwasher as of April 2014.

The scatter plots in Figure 7.6 show appliance base prices versus the absolute difference (left plot) and percentage difference (right plot) between the smart and non-smart versions.

At this time it is difficult to form a meaningful conclusion about the cost of something being *smart*, because of the infancy of the market and the dearth of smart household appliances, infrastructure, and incentives. Yet, the right plot does indicate a trend for smart appliance prices to have about a 20% markup. One refrigerator does have a lower (10%) markup, and one washer and one dryer have much higher markups of 50% and 80%, respectively.

The above examined the incremental cost associated with adding smarts to household appliances, but the analysis did not investigate smart thermostats. It makes less sense to compare the price of a normal thermostat to a smart one, because almost every household already has a thermostat. When deciding whether to upgrade to a smart one, an end-user may be more likely to look for the one with the lowest cost. As such, Table 7.14 contains the costs of the smart thermostats on the market, as of April 2014.

Table 7.14: Smart thermostat costs, as of April 2014

Smart Thermostat	**C**ost
Nest 2nd Generation	$250
Honeywell WiFi	$202
ecobee Smart Si 01	$200
Homewerks CT-30-H-K2	$100
Allure EverSense	$284

Smart Appliance Benefit Analysis

The Appliance Benefit Analysis investigates the average benefits of each appliance. This type of analysis can indicate incremental benefits by type of appliance, which can assist an end-user to understand the returns on investment from adding communication and controls to various types of appliances.

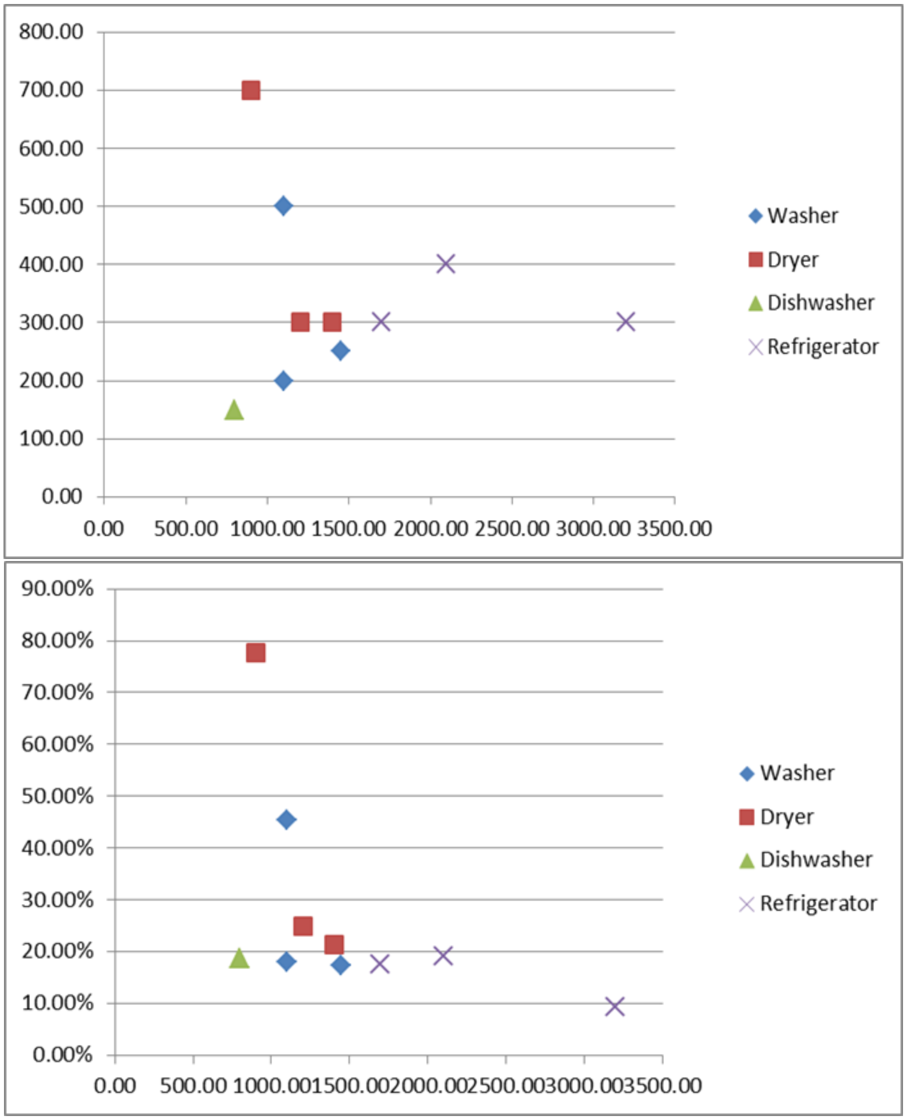


Figure 7.6: Base Price v. Abs. Difference and Base Price v. Per. Difference

The benefit is calculated as the savings obtained from allowing each appliance's load to be shifted. Table 7.15 contains various measures of appliance benefits over 7 days, based on a variation of Case 2: Springfield without Battery. The prices are much lower in the off-peak period, with a Day Ahead price range of [0.019, 0.04136] ¢/kWh.

Table 7.15: Smart and Absolute Benefit of an Appliance

Appliance	**N**on-Smart Cost ($/day)	**S**mart Cost ($/day)	**S**avings ($/day)	**S**avings (%)
Washer	0.0069	0.0042	0.0026	38.35
Dryer	0.0608	0.0385	0.0223	36.67
Dishwasher	0.0347	0.0222	0.0124	35.84
PEV	0.2194	0.1263	0.0931	42.45
HVAC	0.4253	0.1383	0.2870	67.47

Average costs ($/day) are for each appliance over the 7-day model horizon. The Non-Smart column gives the average cost for the nominal load profile of each appliance. The Smart Cost column gives the average cost of operating the given appliance when its load is optimally shifted within the limits of the end-user's preferences. The Savings ($) benefit is the monetary value that was obtained by letting the given appliance load to shift. The Savings (%) is the percent change in Smart Cost relative to the Non-Smart Cost.

By examining these values for a given time frame, an end-user can determine whether purchasing an appliance with load shifting ability would be worthwhile.

Assuming that the Savings (%) values are typical, or that typical values can be somehow computed with the OLS model, they can be used to compute the expected annual savings attributable to each appliance, given Average Annual Costs for each appliance from energy.gov [18]. Table 7.16 contains the elements of such a calculation.

Table 7.16: Annual Savings for an appliance

Applinace	**Savings (%)**	**Average Annual Cost**	**Annual Savings**
Washer	38.35	$ 5.36	$ 2.05
Dryer	36.67	$ 71.40	$ 26.18
Dishwasher	35.84	$ 18.90	$ 6.77
PEV	42.45	$ 232.51	$ 98.69
HVAC	67.47	$ 108.00	$ 72.87

The Savings (%) values are repeated from Table 7.15. The product of the Savings (%) values and the costs for each appliance are in the Annual Savings column. With this annual savings value, an end-user can compare this annual benefit with the annual amortized cost associated with adding communications and control (smarts) to the appliances and determine whether it makes economic sense to spend the extra money.

Combining the Annual Savings values in Table 7.16 with a few assumptions about the costs of smart appliances and thermostats yields the results in the analysis in Table 7.17. The following is assumed:

- The cost of making an appliance smart is an extra 20% on the purchase price.
- The nominal costs of non-smart versions of a washer, dryer, and dishwasher are given in the Table 7.17.
- The incremental cost of a smart charger is on the order of a smart meter, namely $200.
- The incremental cost of a smart thermostat is about $200, assuming that a homeowner would replace a good thermostat having no resale value.
- The payback period does not include a discount rate and is only for ranking purposes.

Table 7.17: Benefit-Cost Analysis for each appliance

Applinace	**Non-Smart Cost**	**Incremental Cost**	**Annual Benefit**	**Payback Period**
Washer	$ 600	$ 120	$ 2	60 years
Dryer	$ 300	$ 60	$ 26	2.3 years
Dishwasher	$ 300	$ 60	$ 7	8.6 years
PEV Charger	$ 2,000	$ 200	$ 99	2.0 years
HVAC	$ 200	$ 200	$ 73	2.7 years

It can be concluded from this table that adding smarts to the more energy intensive appliances, like the clothes dryer, the PEV charger and the HVAC system should provide a more robust benefit to the end user than a washer or dishwasher.

Smart Appliance Marginal Benefit Analysis

The following marginal benefit analysis is possible to implement for appliances that do not have defined load shapes, but are instead modeled as interactive with their environment and as having a limit on their capability. For this reason, the HVAC is used as the example appliance, and we test the sensitivity of its benefits to changing kW capacity values.

Normally an end-user decision about the appropriate size of an HVAC system is specifically based on two factors: climate and house size, but with a HEMS, one can investigate the benefits derived from the ability of a larger HVAC system to pre-cool/pre-heat faster. This analysis leverages the Austin Pre-Cooling case, which increases the HVAC capacity from 2 to 12 kW, as shown in Table 7.18.

The benefit analysis in Table 7.18 shows that increasing the HVAC capacity has diminishing returns. Nevertheless, this framework shows how varying HVAC capacity can help to properly size the system.

Table 7.18: Case 5: Austin Pre-Cooling. Benefit of Increasing HVAC size

HVAC Capacity (kW)	**2**	**3**	**4**	**6**	**8**	**10**	**12**
Savings (%)	27.10	28.07	28.34	28.53	28.60	28.64	28.67
Savings ($)	0.78	0.81	0.82	0.82	0.82	0.82	0.83
Incremental Benefit ($)	n/a	0.0279	0.0078	0.0055	0.0020	0.0012	0.0009
Marginal Benefit ($/kW)	n/a	0.0279	0.0078	0.0028	0.0010	0.0006	0.0005

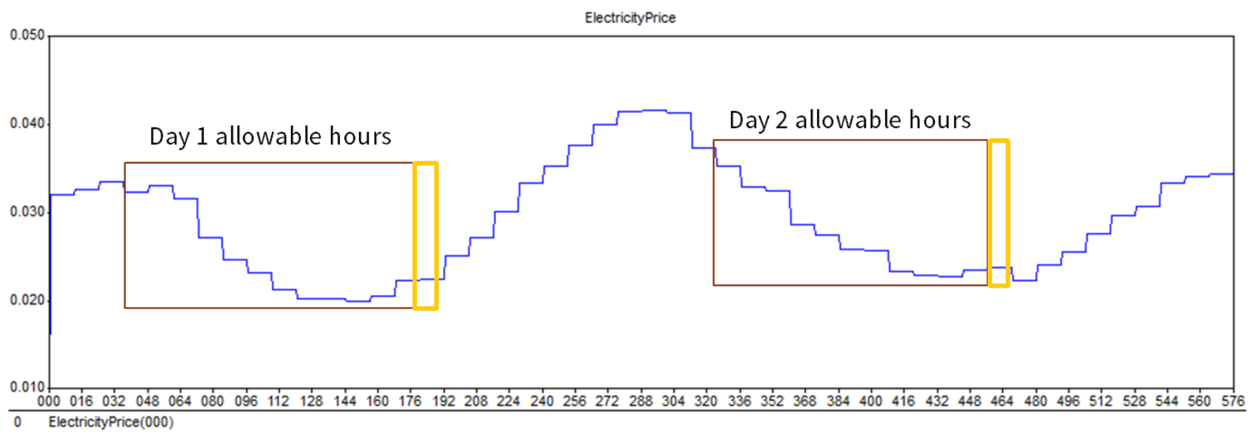


Figure 7.7: Extending Resident's Allowable Hours

Sensitivity on Resident's Behavior

The HEMS offers the end-user the ability to choose the allowable periods of operation to which an appliance can be shifted. One question that logically arises is, "What is the cost or benefit associated with extending or curtailing the window of allowable operation?"

Testing this condition on the Springfield cases demonstrates that no benefit is obtained from lengthening the allowable period by an hour. This is because the allowable window of operation for the set of appliances that can be shifted already encompass the lowest price period, as seen in Figure 7.7. But this analysis is case-specific and may show interesting results for people with particular work schedules, and alternative cases must be tested to conduct such a cost-benefit analysis.

A similar sensitivity analysis on the settings of the allowable temperature range could also be done to determine the marginal benefit of expanding the range.

7.7 Conclusions

It has been shown that the proposed OLS model can be used to accrue economic benefits for each of the five cases through the application of DR technologies in the residential sector. This model can effectively manage electricity demand in the residential sector and the findings are promising for the continued development of smarter electricity management in the future. For instance, the model is capable of performing quickly. More importantly, the model has demonstrated its flexibility to configure its modules to various circumstances. Therefore, existing modules can be improved as needed to represent further details of appliance use and additional modules can be added that better utilize user and electric utility preferences in order to enhance their mutually beneficial relationship.

It has been explained that the end-user can accrue economic benefits by shifting the consumption loads away from higher-priced periods. In the results section, the precise savings are shown in addition to the graphs of shifted load. As stated above, the model has a modular design that is flexible and can be configured to accommodate added capabilities. It has been demonstrated that batteries and automated windows are examples of such added capabilities that further improve end-user benefits. The battery components can give extraordinary savings, but must be compared to installation and maintenance costs and a utility's willingness to buy back electricity. Automated windows may provide benefits, but further testing is needed to identify the optimal conditions for achieving economic savings.

In conclusion, the smart grid analysis framework provides guidance in determining whether current smart appliances are economically priced. For example, by installing a smart thermostat or dryer, the cost might be recouped within five years (based on the Springfield data set). However, a 20% markup on washers and dishwashers might not be recouped within five years, based on the simulated savings. A PEV, which is a load-intensive appliance, may see significant savings from smart charging, but it is difficult to put a cost on that ability, because of the significant variations of installations and the rapidly changing business landscape.

Possible Future Investigation

At the time of developing the algorithm, there were few load shifting algorithms. We not only show a mathematical formulation, but also demonstrate its potential economic benefit. Because of this, we chose simplifying assumptions when creating the algorithm. Future research on optimization algorithms can enable mathematical formulations of cyber-physical systems to be solved with fewer simplifying assumptions.

One simplifying assumption is for the automated windows module. We implement a linear relaxation on the quadratic constraint that is created by allowing the thermal conductivity within a range of values, when windows are opened or closed. By implementing a linear relaxation and only allowing for opened or closed windows, we are left with a much easier problem to solve. This leads to only one set of the two inequality constraints being tight and the other being slack. Further developments in optimization algorithms can solve the more difficult original problem without this relaxation.

New optimization algorithms can also be developed that more efficiently solve problems with cascading dependencies between variables and constraints to model physical phenomena in cyber-physical systems in finer detail. For example, in the thermal control module, we modeled the single residential home as one room. This assumption reduces the model complexity that would be born out of by including heat transfer between rooms and imposing individual temperature ranges for each one. While this simplification provides a sound approach that can calculate economic benefits, future optimization algorithms can more efficiently solve more complex models. Given the trend towards smarter systems, such as homes with multiple temperature zones, such advances in algorithms may lead to additional economic benefits.

To achieve computation efficiency, special care has to be taken to formulate carefully the optimization problem. For example, one old formulation of the Automatic Windows module had a solution time in hours; but with a new formulation to deal with 0-1 variables, the solution time became seconds. Therefore, certain "know-how" in the formulation of 0-1 mixed LP achieves time efficiency that is necessary to make our models practical, and leads as well to future research for even more improvements. In general, we are confident that our OLS approaches can be fully developed into a mature and effective technology in terms of computation resources.

The OLS model's modular design provides the ability to make additional enhancements simpler to design and implement. First, it can be manipulated to handle both non-concurrent and concurrent load shifting. This idea is particularly interesting for charging PEVs. It would be possible to accrue benefits from charging a fleet with overall supply capped for a limited time. In addition, a household might mitigate costs at a time when prices are volatile by not charging concurrently.

Furthermore, the OLS model can be modified to minimize or maximize other costs and benefits regarding the operation of a household or a utility — possibly a utility's ability to weigh a demand response option versus the cost of turning on an emergency generator.

Regarding the formulation, two avenues of future investigation easily arise. The first is to extend the current formulation, which is essentially a single temperature control zone, to multiple zones. This change allows one to explore added controls and more detailed modeling of the building. For example, heat exchanges between rooms can be modeled using fluid and thermal models. Furthermore, the behavior of the building's zones can be modeled using simulation programs and languages like EnergyPlus and Modelica. By turning off or rather utilizing less energy in certain zones within a building, the model can explore possible added value to consumers, utilities, and third parties.

The second avenue of future investigation for the formulation relates to the formulation of the automatic windows. Due to the slower run time associated with the automatic windows as performed now, one could create a two-level approach to address when windows should open and close. This approach can be benchmarked with the current approach with the expectation of delivering a quicker solution.

Lastly, as an analysis tool, the OLS model can provide insight into certain aspects of residential energy consumers and demand response. A few ideas are:

- *Appliance Behavior* – Determine the most important factors affecting appliance behavior for purposes of forecasting sales and market penetration.

- *End-User Behavior* – Determine the most important factors affecting end-user behavior for purposes of characterizing how they can best benefit from the smart grid.
- *Demand-Side Pricing* – A significant application of this model is to determine prices for operating demand-side resources within a microgrid or other structure that is not large enough for wholesale pricing.
- *Demand Response Forecasting* – Calibrate a model for controlled and voluntary demand response to various factors like weather, prices, and energy conservation announcements to be used in short-term load forecasting.
- *Long-Range Planning* – Show how aggregate load shapes change as a result of external factors like weather, prices, and incentives.
- *Energy Policy* – Determine how to represent smart grid benefits for end-users, aggregators, utilities, and the general public.

These insights may be achievable from the foundation provided by the OLS model. The OLS model represents fundamental appliance behaviors and may be scalable to larger systems to help answer key stakeholder questions.

The techniques and Smart Grid Analytical Framework presented here can be extended to other application domains. For example, the energy end-uses in the OLS model can be modified to handle commercial loads. The energy behaviour and physical constraints of commercial loads can be described as optimization constraints for end uses such as automated lighting systems, fleets of commercial vehicles, and commercial heating, ventilation, and cooling systems. The Smart Grid Analytical Framework can be adapted to answer questions, such as those presented in the preceding list, for commercial and industrial buildings and facilities, or networks of connected devices in the internet of things. The proposed OLS model and Smart Grid Analytical Framework represent flexible approaches with a large potential for broader applications.

Chapter 8

Conclusions and Future Research

8.1 Conclusions from Fault Detection

When trying to turn existing controls data or deployed sensors into meaningful information, one may be faced with an overwhelming number of data points. However, the previous results show that only a few sensors may be needed to detect faults with high accuracy rates. Furthermore, these few sensors can be automatically selected. Feature selection methods from machine learning and statistics can calculate and recommend data points that are strong predictors of a fault, are highly relevant, and minimally redundant.

Application domain knowledge can be encoded to create virtual sensors. These virtual sensors, in the sense that they do not correspond to values taken directly off a physical sensor, can be more predictive of a fault than direct sensor data.

An algorithmic approach recommends the most effective features. The top selected features have a probabilistic relationship to the physical location of a fault, but may not be the sensor directly at the fault location at the time of the fault.

When using the expanded data set with feature selection, analysis on the wind turbine data shows that accuracy metrics exceed those of using the original SCADA data alone in the wind turbine. The same accuracy metrics is realized with only two-thirds as many data points. The same accuracy measures as the original feature set is achieved using approximately two-thirds of the original number of features. When using the same number of data points – that is, without additional investment in new communications and data storage – the accuracy metrics show an improvement of 19% over the original data set.

The ability to detect faults with high accuracy with few sensors is encouraging for widespread deployment of such a fault detection system. The few numbers of features shows that across a type of system, for data collection purposes, individual instances of a system need not share a large number of sensors. To install the data collection part of such a fault detection system, only a subset of all the sensors need to be identified.

To detect a larger number of different faults in the possible set of faults, however, the number of features needed for fault detection may be higher than as shown in these results.

Nonetheless, the proposed methodology allows for user choice in the types of faults to focus on and the ability to trade off detection accuracy and number of types of faults with data collection and computation costs. Real sensor data points can be supplemented with virtual sensors data points and be processed in workflows used to create fault detection models and machine learning models.

The flexibility in the proposed methodology eases adoption of the methodology for different types of systems. This method was first demonstrated to wind turbines, then commercial building chiller plants, and can be adopted to, for example, commercial building HVAC systems, solar panels, natural gas turbines, and computer hardware.

To apply this methodology to another system only requires 1) the identification of common existing data points and 2) basic application knowledge, gathered from experts, in the form of a list of metrics. Then, the proposed methodology enables increased fault detection accuracy with fewer features without vendors and experts to install and configure the system. The resulting software can also be readily deployed to other instances of the same type of system, due to the lack of reliance on individual specific information. Deployment costs are also lowered due to this approach's demonstrated effectiveness when using of off-the-shelf, open source algorithms, and publicly published algorithms.

The resulting analysis can be both implemented in real systems (ie. real buildings and real wind turbines) or as a policy making tool to answer stakeholder questions

8.2 Conclusions from Energy Scheduling

Economic benefits can be accrued from demand side management through optimal scheduling of energy consumption. The proposed OLS model can effectively manage electricity demand in the residential sector and the findings are promising for the continued development of smarter electricity management in the future.

It has been explained that the end-user can accrue economic benefits by shifting the consumption loads away from higher-priced periods. The model has a modular design that is flexible and can be configured to accommodate added capabilities. It has been demonstrated that batteries and automated windows are examples of such added capabilities that further improve end-user benefits. The battery components can give extraordinary savings, but must be compared to installation and maintenance costs and a utility's willingness to buy back electricity. Automated windows may provide benefits, but further testing is needed to identify the optimal conditions for achieving economic savings.

In conclusion, the smart grid analysis framework provides guidance in determining whether current smart appliances are economically priced. For example, by installing a smart thermostat or dryer, the cost might be recouped within five years (based on the Springfield data set). However, a 20% markup on washers and dishwashers might not be recouped within five years, based on the simulated savings. A PEV, which is a load-intensive appliance, may see significant savings from smart charging, but it is difficult to put a cost on that ability, because of the significant variations of installations and the rapidly changing business landscape.

Furthermore, the OLS model can be modified to minimize or maximize other costs and benefits regarding the operation of a household or a utility — possibly a utility's ability to weigh a demand response option versus the cost of turning on an emergency generator. The OLS model represents fundamental appliance behaviors and may be scalable to larger systems to help answer key stakeholder questions. As an analysis tool, the OLS model can provide insight into certain aspects of residential energy consumers and demand response.

8.3 Future Work

The proposed fault detection methodology can be extended to include faults that evolve, become worse, and also can be used to account for when faults are fixed, the physical system is changed or upgraded, or when maintenance occurs, as happens in real world systems. The fault detection capabilities can be expanded to include fault diagnosis, identification, and prognosis.

The modularity of the proposed methodology allows for future research on more effective choices of parameters. For example, further analysis can be done on the choice of a machine learning model, feature selection process, sample weights, time decay function, and class weights. From a time series analysis perspective, other cross-validation for time series schemes can be used to assess model performance. Different choices can also be made for the selection and weighting of the sliding window for the mean and standard deviation. The maximum order of the AIC and BIC can be increased. Other time series characteristics can also be calculated and included as features. Instead of using machine learning feature selection techniques, techniques from statistics can be used to choose features. This includes calculating descriptive techniques from modelling the sensor data as time series.

Data for fault detection can also be framed as point process data, which is a set of ordered numbers that represent the times of events that occurred in some time interval. Then, techniques for point processes can be applied to the fault detection problem. For point processes, a progression of techniques can be used, from descriptive, to moments and correlation, to regression, and likelihood, as described in [5].

Additional research is needed to collect more sensor data and fault labels under normal operations and different faults. Sensor measurements under faulty operations are particularly difficulty to obtain, due to low probability faults and the undesirability of introducing faults to expensive systems for the purposes of data collection.

How to fuse expert knowledge about the application with data driven approaches is also a direction of research. The informal interview process can be formalized and improved. The balance of generic system knowledge and knowledge of the specific deployment also needs to be balanced. Expert knowledge can also be combined with data driven approaches in other, more effective ways.

New methods to combine data from different sources and across different systems is also needed. A new deployment may have limited, if any, historical data, and it is desirable to use data from other deployments and other systems, both real measurements and simulated data.

Given differences between individual instances, for example between individual buildings or simulated and real systems, procedures are needed to fuse data from different sources to take advantage of available data. The accuracy of predictions from machine learning models is highly dependent on the quantity and quality of training data.

Another avenue for investigation is in addressing computational challenges of using data-driven approaches, and leveraging cloud computing. At the time of this thesis, personal computers and computer servers were able to conduct the calculations for the proposed methodology. However, extending the methodology to data sets that include more faults, long time periods of data, and more systems requires computing resources on the scale of supercomputers and cloud computing. Thus, research is needed to effectively scale and formulate methodologies. Examples include parallelizing processes, securing data, and sharing data. Alternate formulations of the problem can also be developed to simplify the computation requirement.

Future research can also improve the ease of installation and the user experience. User interfaces should be easy to use, accessible, and encourage consistent response to faults and energy improvement opportunities. This may be accomplished through research in integration of fault detection systems with operations and maintenance, design of the control and automation system, and design of user interfaces. Fault detection could be as simple as verbally speaking. More advanced decision-analytic queries, such as where to invest \$ X for energy improvement, may also be possible and as easy as asking a digital assistant such as Apple's Siri, Amazon's Alexa, or Microsoft's Cortana.

The OLS and Smart Grid Analytical Framework presented here can be extended to other application domains. For example, the energy end-uses in the OLS model can be modified to handle commercial loads. The energy behaviour and physical constraints of commercial loads can be described as optimization constraints for end uses such as automated lighting systems, fleets of commercial vehicles, and commercial heating, ventilation, and cooling systems. The Smart Grid Analytical Framework can be adapted to answer questions, such as those presented in the preceding list, for commercial and industrial buildings and facilities, or networks of connected devices in the internet of things. The proposed OLS model and Smart Grid Analytical Framework represent flexible approaches with a large potential for broader applications.

The trend towards increased computation resources, connectivity, controls, data, and the cloud is exciting and encouraging for developments in these research directions.

Appendix A

Appendices for Chapter 7

A.1 List of Acronyms

AW	Automated Windows
AMI	Advanced Metering Infrastructure
BESS	Battery Electricity Storage System
BEV	Battery Electric Vehicle
DR	Demand Response
EV	Electronic Vehicle
FCEV	Fuel Cel Electric Vehicle
HEMS	Home Energy Management System
HVAC	Heating, Ventilation, & Air Conditioning
ICT	Information and Communication Technology
LS	Load Shifting
NS	Non-Shiftable
OLS	Optimal Load Shifting
PEV	Plug-In Electronic Vehicle
REDD	Reference Energy Disaggregation Data Set
TC	Thermal Control
TOU	Time of Use
UTC	Coordinated Universal Time

A.2 Data

This section describes the data needs for using the OLS model and data sources for these case studies. Some data pertains to appliance use, and other data is used for consumer preferences and electricity prices. These sources are combined to create cases for later analysis. The cases along with their salient characteristics are located in Table 7.9, which is repro-

Table A.1: Descriptions of Cases

Case	Location	Modules	End-Use Data	Electricity Pricing
1	Boston, MA	LS, NS	Real	Time of Use
2	Springfield, IL	LS, TC, NS	Simulated	Day Ahead
3	Springfield, IL	LS, TC, BESS, NS	Simulated	Day Ahead
4	Springfield, IL	LS, TC, AW, NS	Simulated	Day Ahead
5	Austin, TX	TC, NS	Simulated	Time of Use

duced below. This data description is useful as a reference to publicly available data and because it demonstrates an assemblage of data needed to conduct such analysis.

Desired Data Properties

Optimizing electricity load shapes depends on the time-varying behavior of user preferences, end-use appliances, and electricity prices. For this reason, the model requires nominal time series load data at the appliance level. The time resolution should preferably be as fine as possible for the potential decision cycle. In the following cases, the time resolution is based on bulk electricity system (BES) [20] needs, and it ranges from seconds to minutes. Modern Advanced Meter Infrastructures make measurements on the order of 5-minute intervals, which are related to the wholesale electricity market having typically the same interval lengths for Real Time energy dispatch.

The time series data should extend from the present for about a week, and it should correspond to the end-user's flexibility to schedule their loads over the immediate future. The OLS model is a bottom-up representation of electricity use and can be concentrated on the electricity use within a single building. But it can also be used to represent coordinated behavior between multiple interconnected structures.

Case 1: Boston Load Shift

A highly useful initial source of end-use data is the Reference Energy Disaggregation Data Set (REDD) [46], because it is publicly available and possesses many of the desired traits described above. In addition, REDD is intended for research purposes.

The REDD data set includes energy-use data for six residential houses for about two months in 2011. However, there are long periods of sparse data, and each house has different metered components. The house known as *house_1* is described here as an example, because it had a large number of sub-meters, particularly for appliances that are candidates for load shifting. Such appliances include a dishwasher, clothes dryer, and clothes washer. The week of May 15-21, 2011 is selected for a case study, because its data is consistently available over the duration of the time period.

Description of *house_1* Data

The *house_1* data is divided into twenty data channels and a *labels* file. The labels file contains a mapping from the numbers of each channel file to a metered circuit or appliance.

Pre-Processing Data

The REDD data is preprocessed to make it available in a form that fits well with the OLS model. First, the data channels are selected, and then the channel data is converted from power readings to energy use over the desired interval length. These processes are as follows.

The candidate shiftable end-uses are:

- Dishwasher
- Washer-dryer
- Heating
- Air conditioning
- Refrigerator
- Freezer
- Plug-In Electronic vehicle

All other loads in the metered circuits are treated as non-shiftable.

The data channel files have a format with two columns of data: the first column is a time stamp and the second is a power flow reading. The time stamp is in coordinated universal time (UTC) format, and these readings occur approximately every three seconds. The power flow reading is in watts. There is no header row.

To use this data stream within the OLS model, the power flow readings are first converted to units of average hourly or 5-minute energy, depending on the scenario. A MATLAB function (named `integrate_per_hour`) computes these values. To automate the process, a script (named `power2energy_per_hour`) iterates through all the data channels and converts them to streams of average hourly energy. The units of watts are thus converted to kWh.

Since *house_1* has three data channels labeled `washer_dryer`, the energy for those three channels is aggregated (Note: that the model includes an assumption that the clothes dryer, when used, is used immediately following the clothes washer). Thus, the model assumes that the three channels represent different parts of the circuit to which a washer and a dryer are connected rather than the fact that there are three washers and dryers.

Data Ready for Model

The advantage of the REDD data is that it is real world data. However, it is limited in two major respects. The first limitation is the periods of missing data, which leads to questions pertaining to the use of this information as a case, particularly when coupled with the lack of specific homeowner preferences. The second limitation is the impossibility of identifying whether the given home has an electric HVAC system, because its temperate climate over the given period and/or the lack of insights about the homeowner's indoor temperature range preferences.

Cases 2, 3, 4: Springfield General Data

Because of the stated limitations of the REDD data, the project team prepared a separate set of simulated data for one house with the goal being to fully exercise the OLS model features. The simulated data set is based on a combination of design choices about the location, outside temperature range, and available electricity prices.

In particular, the team picked a 6-day period during August, for which weather and electricity prices could be collected directly. A Day-Ahead price scheme was chosen over a TOU scheme, because such prices have more resolution than just peak and off-peak values. Lastly, the assumed end-user has a plug-in electric vehicle, works a 9am-5pm job, and has a fairly basic set of end-use preferences.

Location

The location for the simulated data set was chosen to be the city of Springfield, Illinois for two main reasons. First, in the course of the chosen week (8/11/13-8/17/13), the climate requires both heating and cooling. Second, the local utility, called *Ameren*, provides its consumers with the option of day ahead pricing for electricity use. This feature is particularly attractive, because it makes the prices more varied than merely a time-of-use pricing scheme.

Climate

The Springfield temperatures are the actual temperatures during the modeling horizon of 8/11/13-8/17/13. They were acquired from the Weather Underground, a commercial weather service that provides real-time weather information [77]. Their site collects most of its data from the National Weather Service.

Electricity Price

The prices are the day ahead prices for Springfield during 8/11/13-8/17/13. They were acquired from Ameren's website [3]. It is also important to note that a person can acquire the day ahead prices one day in advance, which is a highly beneficial feature for the algorithm.

Simulating the Appliances

Washer The U. S. Department of Energy reports that a washer's wattage ranges from 350 to 500 watts [18]. This information was used to produce a simple algorithm to produce the kW usage for the washer when used. It is assumed the Springfield household will use its washer 4 times a week. The washer runs on Sunday, Tuesday, Thursday, and Saturday. The times were set to times of the day when the end-user would be at home: 7 pm, 8 pm, 8 pm, and 12 pm, for each day, respectively.

Next, the phantom loads for the washer, which are extremely small, were produced with a simple algorithm. The algorithm was designed by looking at the actual phantom loads from the REDD data set. From that, it appears that a washer randomly draws phantom loads at fractions of a watt throughout the day or draws no loads at all.

Dryer The U. S. Department of Energy reports that an electric dryer's wattage ranges from 1800 to 5000 watts [18]. This information was used to produce a simple algorithm to produce the kW usage for the dryer when used. It is assumed that the Springfield household will use its dryer 4 times a week. It operated on Sunday, Tuesday, Thursday, and Saturday. The times were set to the times of the day when the end-user would be at home: 8 pm, 9 pm, 9 pm, and 1 pm.

Next, the phantom loads for the dryer, which are extremely small, were produced with a simple algorithm. The algorithm was designed by looking at the actual phantom loads from the REDD data set. From that, it appears that a dryer randomly draws phantom loads at fractions of a watt throughout the day or draws no loads at all.

Dishwasher The U. S. Department of Energy reports that a dishwasher's wattage ranges from 200 to 2400 watts and even greater if the dishwasher's drying option is enacted [18].

This model does not use the drying option. This information was used to produce a simple algorithm to produce the kW usage for the dishwasher when used. It is assumed that the Springfield household will use its dishwasher 6 times a week. It operated on Sunday, Monday, Tuesday, Wednesday, Thursday, and Saturday. The times were set to those when the end-user would be at home: 7 pm, 8 pm, 8 pm, 7 pm, 9 pm and 6 pm, for each day, respectively.

Next, the phantom loads for the dishwasher, which are extremely small,were produced with a simple algorithm. The algorithm was designed by looking at the actual phantom loads from the REDD data set. From that, it appears that a dishwasher randomly draws phantom loads at fractions of a watt throughout the day.

Electric Car The Global EV Outlook predicts 20 million electronic vehicles will be used worldwide by 2020, and U.S. consumers already represent 38% of that number. [68] Electric vehicles are defined in this report as passenger car plug-in hybrid electric vehicles (PHEV), battery electric vehicles (BEV), and fuel cell electric vehicles (FCEV). In addition, because a large part of an EV owner's electricity cost may be attributed to the EV, and due to the efficacy of shifting such a load, the OLS model incorporates the PEV in the Springfield simulation.

The Mini E was selected to be used for the model. The Mini E gets 0.22 kWh/mile [69]. It is assumed that the Mini E must refill after 100 miles, however, in the simulation, the automobile never reaches this level of mileage. If the Mini E's battery is completely empty and its owner seeks to recharge it fully, it would take approximately 3 hours to do so. However, in the simulation, it will never get to that point because of daily recharging and the specific homeowner's driving profile. The homeowner is assumed to be using a level 2 PEV charger [75], with maximum load of 7.68 kW [76].

Assumptions

- Driver works 15 miles from his or her home.
- Driver drives 30 work miles plus a uniform random number from 0 to 10 miles on a workday
- Driver drives uniform random number from 0 to 10 miles on a weekend.
- The PEV charges every day, and the times were 4 p.m., 6 p.m., 6 p.m., 6 p.m., 6 p.m., 6 p.m., and 8 p.m. for each day, respectively.
- No phantom loads were generated for the PEV charging station.

Non-Shiftable Loads Each day, a house has loads that cannot be shifted, including but not limited to televisions, stereos, laptops, lights, and tablet/phone charging. In addition, for this simulation, the refrigerator is treated as a non-shiftable load, although it could potentially be treated as an energy storage end-use. The non-shiftable load quantity was generated as a random multiple between 1.0 and 1.5 of the measured non-shiftable load from the REDD data set.

Case 4: Springfield Automated Windows

Mild climates can benefit from opening windows to allow temperate air to enter a building in order to maintain a preferred temperature range. This feature of the OLS model was run only on the first two days of the Springfield, IL data.

Time Horizon

The time period for which we test this portion is short in comparison with the one- or two-day periods we use in other cases. This is to reduce the computation time. The time period here is 4 hours of 5-minute intervals on Tuesday evening (8/13/13) from 8 pm - 12 am.

Location

The location is Springfield, IL.

Climate

The exact time horizon was chosen in order to emphasize when benefits from opening and closing windows would be apparent. The temperature begins at 60.1 degrees Fahrenheit and falls into the low 50's over the time horizon.

Electricity Prices

To avoid any pre-cooling or pre-heating effects, a TOU pricing scheme was chosen instead of the day ahead scheme of the other Springfield, IL cases. The modeling horizon is a peak price period, and the price is 0.12 $/kWh.

Case 5: Austin Pre-Cooling

The Austin, Texas location was chosen in order to further emphasize the specific pre-cooling aspect of the model. Its character is that it has an external temperature range that is always above the homeowner's preferred temperature range. In this way, the model exercises only its cooling feature. For this location, only the HVAC portion of the model is tested, and according, only real data on temperature and prices are needed. The loads in this case are also synthesized for a single house.

Location

The city of Austin, Texas was chosen for its hot climate. The model was implemented for the period of August 6, 3:00PM to August 8, 3:00 PM (Tues-Thurs). The location provides a climate is beneficial for investigating and testing comfort preference zones, HVAC power levels, and economic impacts on locations where HVAC systems must run continuously.

Climate

The temperatures used are the actual temperatures in Austin, Texas during 8/6/13-8/8/13, and they range from 78.1 to 104 degrees Fahrenheit. They were acquired from Weather Underground, a commercial weather service that provides real-time weather information on the Internet [77]. The site collects most of its data from the National Weather Service.

Electricity Price

The prices are the TOU prices for Austin during 8/6/13-8/8/13, and they were acquired from Austin Energy's website [16]. It is also important to note that an Austin consumer can acquire TOU prices in advance, which is a highly beneficial feature for the algorithm.

A.3 Continuous Operation

The LS module formulation needs a few added constraints to represent the fact that some end-uses i run continuously during its duty cycle (can switch on and off at most twice). This can be achieved by adding linear constraints:

$$y_{i,t} = x_{i,t} - x_{i,t-1} \tag{A.3.1}$$

$$\sum_t |y_{i,t}| \leq 2 \tag{A.3.2}$$

$$y_{i,1} + y_{i,h} \leq 1 \tag{A.3.3}$$

The new variable y defined in (A.3.1) is introduced to indicate whether end-use i switched between on and off states in interval t. The constraint (A.3.2) guarantees that $y_{i\cdot}$ can only switch states twice over the model's horizon. With this constraint alone, there are two possibilities for $x_{i\cdot}$ to switch twice. The first is that end-use i will run continuously, as in the following example definitions:

$$x_i = [0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ \ldots]$$

$$y_i = [0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ \ldots]$$

The second is that end-use i will run continuously at the beginning of the model for some period, stop, and then continue to run continuously at the end of the model, as in

$$x_i = [1\ 1\ 0\ 0\ \ldots\ 0\ 0\ 1\ 1\ 1]$$

$$y_i = [0\ 0\ 1\ 0\ \ldots\ 0\ 0\ 1\ 0\ 0]$$

To avoid the second possibility, enforce (A.3.3).

Constraints can be constructed to ensure that the washer runs prior to the dryer and that the dryer runs directly after the washer.

Bibliography

[1] U.S. Energy Information Administration. *Annual Energy Outlook 2013*. U.S. Energy Information Administration, 2013.

[2] Changsun Ahn, Chiao-Ting Li, and Huei Peng. "Optimal decentralized charging control algorithm for electrified vehicles connected to smart grid". In: *Journal of Power Sources* 196 (23 Dec. 2011), pp. 10369–10379.

[3] Ameren. Ameren Website. URL: `https://www2.ameren.com/RetailEnergy/realtimeprices.aspx`.

[4] Modelica Association. *Modelica and the Modelica Association*. Modelica Association . Accessed: 2016-11-22. URL: `https://www.modelica.org/`.

[5] David R. Brillinger. "Nerve Cell Spike Train Data Analysis: A Progression of Technique". In: *Journal of the American Statistical Association* 87.418 (1992), pp. 260–271. DOI: `10.1080/01621459.1992.10475205`. eprint: `http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1992.10475205`. URL: `http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1992.10475205`.

[6] Shane Butler, John Ringwood, and Frank O'Connor. "Exploiting SCADA system data for wind turbine performance monitoring". In: *2013 Conference on Control and Fault-Tolerant Systems (SysTol)*. IEEE, Oct. 2013, pp. 389–394. ISBN: 978-1-4799-2855-2. DOI: `10.1109/SysTol.2013.6693951`. URL: `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6693951`.

[7] Brigitte Chebel-Morello, Simon Malinowski, and Hafida Senoussi. "Feature selection for fault detection systems: application to the Tennessee Eastman process". In: *Applied Intelligence* 44.1 (2016), pp. 111–122.

[8] X. Chen, T. Wei, and S. Hu. "Uncertainty-Aware Household Appliance Scheduling Considering Dynamic Electricity Pricing in Smart Home". In: *IEEE Transactions on Smart Grid* 4.2 (June 2013), pp. 932–941. ISSN: 1949-3053. DOI: `10.1109/TSG.2012.2226065`.

[9] M. C. Comstock, J. E. Braun, and E. A. Groll. "The Sensitivity of Chiller Performance to Common Faults". In: *HVAC&R Research* 7.3 (2001), pp. 263–279. DOI: `10.1080/10789669.2001.10391274`. eprint: `http://www.tandfonline.com/doi/pdf/10.1080/10789669.2001.10391274`. URL: `http://www.tandfonline.com/doi/abs/10.1080/10789669.2001.10391274`.

[10] Jingtan Cui and Shengwei Wang. "A model-based online fault detection and diagnosis strategy for centrifugal chiller systems". In: *International Journal of Thermal Sciences* 44.10 (2005), pp. 986–999. ISSN: 1290-0729. DOI: `http://dx.doi.org/10.1016/j.ijthermalsci.2005.03.004`. URL: `http://www.sciencedirect.com/science/article/pii/S1290072905000918`.

[11] S. S. van Dam, C. A. Bakker, and J. C. Buiter. "Do home energy management systems make sense? Assessing their overall lifecycle impact". In: *Energy Policy* 63 (Dec. 2013), pp. 398–407.

[12] S.S. van Dam, C.A. Bakker, and J.C. Buiter. "Do home energy management systems make sense? Assessing their overall lifecycle impact". In: *Energy Policy* 63 (2013), pp. 398–407. ISSN: 0301-4215. DOI: `http://dx.doi.org/10.1016/j.enpol.2013.09.041`. URL: `http://www.sciencedirect.com/science/article/pii/S0301421513009658`.

[13] Thomas G. Dietterich. "Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops SSPR 2002 and SPR 2002 Windsor, Ontario, Canada, August 6–9, 2002 Proceedings". In: ed. by Terry Caelli, Adnan Amin, Robert P. W. Duin, Dick de Ridder, and Mohamed Kamel. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. Chap. Machine Learning for Sequential Data: A Review, pp. 15–30. ISBN: 978-3-540-70659-5. DOI: `10.1007/3-540-70659-3_2`. URL: `http://dx.doi.org/10.1007/3-540-70659-3_2`.

[14] Chris Ding and Hanchuan Peng. "Minimum redundancy feature selection from microarray gene expression data". In: *Journal of bioinformatics and computational biology* 3.02 (2005), pp. 185–205.

[15] Zhimin Du, Bo Fan, Xinqiao Jin, and Jinlei Chi. "Fault detection and diagnosis for buildings and {HVAC} systems using combined neural networks and subtractive clustering analysis". In: *Building and Environment* 73 (2014), pp. 1–11. ISSN: 0360-1323. DOI: `http://dx.doi.org/10.1016/j.buildenv.2013.11.021`. URL: `http://www.sciencedirect.com/science/article/pii/S0360132313003417`.

[16] Austin Energy. Austin Energy Website . URL: `http://www.austinenergy.com/wps/wcm/connect/4f5bf55e-f9a5-4997-8b00-c531065bd279/ResidentialAustin.pdf?MOD=AJPERES`.

[17] U.S. Department of Energy. U.S. Department of Energy Website. 2016. URL: `http://energy.gov/eere/energybasics/articles/glossary-energy-related-terms`.

[18] U.S. Department of Energy. *Estimating Appliance and Home Electronic Energy Use*. Energy.gov. Nov. 2011. URL: `http://www.energy.gov/energysaver/articles/estimating-appliance-and-home-electronic-energy-use`.

[19] Ahmad Faruqui, Dan Harris, and Ryan Hledik. "Unlocking the €53 billion savings from smart meters in the EU: How increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment". In: *Energy Policy* 38 (10 Oct. 2010), pp. 6222–6231.

[20] Van Ness Feldman. *FERC Approves New Definition of Bulk Electric System, Reserves Determination of Which Facilities Are "Used in Local Distribution"*. Van Ness Feldman Website. Jan. 2013. URL: `http://www.vnf.com/1126`.

[21] FERC. *2012 Assessment of Demand Response and Advanced Metering: Staff Report*. Tech. rep. Federal Energy Regulatory Commission, Dec. 2012. URL: `https://www.ferc.gov/legal/staff-reports/12-20-12-demand-response.pdf`.

[22] FERC. *2014 Assessment of Demand Response and Advanced Metering: Staff Report*. Tech. rep. Federal Energy Regulatory Commission, Dec. 2014. URL: `https://www.ferc.gov/legal/staff-reports/2014/demand-response.pdf`.

[23] FERC. *National Action Plan on Demand Response*. Tech. rep. Federal Energy Regulatory Commission, June 2010. URL: `https://www.ferc.gov/legal/staff-reports/2010-dr-report.pdf`.

[24] Stephen Frank, Michael Heaney, Xin Jin, Joseph Robertson, Howard Cheung, Ryan Elmore, and Gregor Henze. "Hybrid Model-based and Data-driven Fault Detection and Diagnostics for Commercial Buildings". In: *ACEEE American Council for an Energy-Efficient Economy*. 2016.

[25] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Vol. 1. Springer series in statistics Springer, Berlin, 2001.

[26] Lingwen Gan, Ufuk Topcu, and Steven H. Low. "Optimal Decentralized Protocol for Electric Vehicle Charging". In: *IEEE TRANSACTIONS ON POWER SYSTEMS* 28 (2 May 2013), pp. 940–951.

[27] Simon Gill, Bruce Stephen, and Stuart Galloway. "Wind turbine condition assessment through power curve copula modeling". In: *IEEE Transactions on Sustainable Energy* 3.1 (2012), pp. 94–101. ISSN: 19493029. DOI: `10.1109/TSTE.2011.2167164`.

[28] Jamie L Godwin and Peter Matthews. "Classification and Detection of Wind Turbine Pitch Faults Through SCADA Data Analysis". In: *International Journal of Prognostics and Health Management* 4 (2013), p. 11. ISSN: 21532648. URL: `http://www.phmsociety.org/node/982`.

[29] Jessica Granderson, Mary Ann Piette, Ben Rosenblum, Rong Hu, and et al. *Energy Information Handbook: Applications for Energy-Efficient Buildings*. Lawrence Berkeley National Laboratory, LBNL-5272E, 2011.

[30] Ying Guo, Josh Wall, Jiaming Li, and Sam West. "A machine learning approach for fault detection in multi-variable systems". In: *ATES in conjunction with Tenth Conference on Autonomous Agents and Multi-Agent Systems (AAMAS) AAMAS*. Vol. 2011. 2011.

[31] Fredrik Gustafsson. "Fault detection terminology". In: *Adaptive Filtering and Change Detection* (2001), pp. 475–476.

[32] Isabelle Guyon and André Elisseeff. "An Introduction to Variable and Feature Selection". In: *J. Mach. Learn. Res.* 3 (Mar. 2003), pp. 1157–1182. ISSN: 1532-4435. URL: `http://dl.acm.org/citation.cfm?id=944919.944968`.

[33] Z. Hameed, Y.S. Hong, Y.M. Cho, S.H. Ahn, and C.K. Song. "Condition monitoring and fault detection of wind turbines and related algorithms: A review". In: *Renewable and Sustainable Energy Reviews* 13.1 (2009), pp. 1–39. ISSN: 1364-0321. DOI: `http://dx.doi.org/10.1016/j.rser.2007.05.008`. URL: `http://www.sciencedirect.com/science/article/pii/S1364032107001098`.

[34] Tom Hargreaves, Michael Nye, and Jacquelin Burgess. "Keeping energy visible? Exploring how householders interact with feedback from smart energy monitors in the longer term". In: *Energy Policy* 52 (Apr. 2013), pp. 126–134.

[35] Tom Hargreaves, Michael Nye, and Jacquelin Burgess. "Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors". In: *Energy Policy* 38.10 (2010). The socio-economic transition towards a hydrogen economy - findings from European research, with regular papers, pp. 6111–6119. ISSN: 0301-4215. DOI: `http://dx.doi.org/10.1016/j.enpol.2010.05.068`. URL: `http://www.sciencedirect.com/science/article/pii/S030142151000460X`.

[36] John M House, Hossein Vaezi-Nejad, and J Michael Whitcomb. "An expert rule set for fault detection in air-handling units". In: *Transactions-American Society of Heating Refrigerating and Air Conditioning Engineers* 107.1 (2001), pp. 858–874.

[37] R. Lily Hu, Kevin Leahy, Ioannis C. Konstantakopoulos, David M Auslander, Costas J. Spanos, and Alice M. Agogino. "Using Domain Knowledge Features for Wind Turbine Diagnostics". In: *IEEE International Conference on Machine Learning and Applications* (2016).

[38] R. Lily Hu, Ryan Skorupski, Robert Entriken, and Yinyu Ye. "A Benefit-Cost Framework for Optimal Load Shifting". In: *Advanced Workshop in Regulation and Competition 28th Annual Western Conference* (2015).

[39] R. Lily Hu, Ryan Skorupski, Robert Entriken, and Yinyu Ye. "A Mathematical Programming Formulation for Optimal Load Shifting of Electricity Demand for the Smart Grid". In: *IEEE Transactions on Big Data* (2017).

[40] Rolf Isermann. "Model-based fault-detection and diagnosis – status and applications". In: *Annual Reviews in Control* 29.1 (2005), pp. 71–85. ISSN: 1367-5788. DOI: `http://dx.doi.org/10.1016/j.arcontrol.2004.12.002`. URL: `http://www.sciencedirect.com/science/article/pii/S1367578805000052`.

[41] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Vol. 6. Springer, 2013.

[42] Srinivas Katipamula and Michael R. Brambley. "Review Article: Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part I". In: *HVAC&R Research* 11.1 (2005), pp. 3–25. DOI: `10.1080/10789669.2005.10391123`. eprint: `http://www.tandfonline.com/doi/pdf/10.1080/10789669.2005.10391123`. URL: `http://www.tandfonline.com/doi/abs/10.1080/10789669.2005.10391123`.

[43] Srinivas Katipamula and Michael R. Brambley. "Review Article: Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part II". In: *HVAC&R Research* 11.2 (2005), pp. 169–187. DOI: `10.1080/10789669.2005.10391133`. eprint: `http://www.tandfonline.com/doi/pdf/10.1080/10789669.2005.10391133`. URL: `http://www.tandfonline.com/doi/abs/10.1080/10789669.2005.10391133`.

[44] I. Khan, A. Mahmood, Nadeem Javaid, S. Razzaq, R. D. Khan, and M. Ilahi. "Home Energy Management Systems in Future Smart Grids". In: *CoRR* abs/1306.1137 (2013).

[45] Imran Khan, Alfonso Capozzoli, Stefano Paolo Corgnati, and Tania Cerquitelli. "Fault Detection Analysis of Building Energy Consumption Using Data Mining Techniques". In: *Energy Procedia* 42 (2013), pp. 557–566. ISSN: 1876-6102. DOI: `http://dx.doi.org/10.1016/j.egypro.2013.11.057`. URL: `http://www.sciencedirect.com/science/article/pii/S1876610213017591`.

[46] J. Zico Kolter and M. J. Johnson. *REDD: A Public Data Set for Energy Disaggregation Research*. 2011.

[47] Kevin Leahy, R Lily Hu, Ioannis C Konstantakopoulos, Costas J Spanos, and Alice M Agogino. "Diagnosing Wind Turbine Faults Using Machine Learning Techniques Applied to Operational Data". In: *IEEE International Conference on Prognostics and Health Management* (2016).

[48] J. Liang and R. Du. "Model-based Fault Detection and Diagnosis of {HVAC} systems using Support Vector Machine method". In: *International Journal of Refrigeration* 30.6 (2007), pp. 1104–1114. ISSN: 0140-7007. DOI: `http://dx.doi.org/10.1016/j.ijrefrig.2006.12.012`. URL: `http://www.sciencedirect.com/science/article/pii/S0140700707000072`.

[49] Claytex Services Limited. *Dymola*. Claytex Services Limited . Accessed: 2016-11-22. URL: `http://www.claytex.com/products/dymola/`.

[50] Arif S. Malik and Mounir Bouzguenda. "Effects of smart grid technologies on capacity and energy savings – A case study of Oman". In: *Energy* 54 (June 2013), pp. 365–371.

[51] E. Mills. "Building commissioning: a golden opportunity for reducing energy costs and greenhouse gas emissions in the United States". In: *Energy Efficiency* 4.2 (2011), pp. 145–173. URL: `http://link.springer.com/article/10.1007%2Fs12053-011-9116-8`.

[52] A. H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia. "Autonomous Demand-Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid". In: *IEEE Transactions on Smart Grid* 1.3 (Dec. 2010), pp. 320–331. ISSN: 1949-3053. DOI: `10.1109/TSG.2010.2089069`.

[53] Mengyan Nie and Ling Wang. "Review of Condition Monitoring and Fault Diagnosis Technologies for Wind Turbine Gearbox". In: *Procedia {CIRP}* 11 (2013). 2nd International Through-life Engineering Services Conference, pp. 287–290. ISSN: 2212-8271. DOI: `http://dx.doi.org/10.1016/j.procir.2013.07.018`. URL: `http://www.sciencedirect.com/science/article/pii/S2212827113004915`.

[54] Joon-young Park, Jae-kyung Lee, Ki-yong Oh, and Jun-shin Lee. "Development of a Novel Power Curve Monitoring Method for Wind Turbines and Its Field Tests". In: *IEEE Transactions on Energy Conversion* 29.1 (2014), pp. 119–128. ISSN: 0885-8969. DOI: `10.1109/TEC.2013.2294893`.

[55] Hanchuan Peng, Fuhui Long, and Chris Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy". In: *IEEE Transactions on pattern analysis and machine intelligence* 27.8 (2005), pp. 1226–1238.

[56] Jianying Qin and Shengwei Wang. "A fault detection and diagnosis strategy of {VAV} air-conditioning systems for improved energy and control performances". In: *Energy and Buildings* 37.10 (2005), pp. 1035–1048. ISSN: 0378-7788. DOI: `http://dx.doi.org/10.1016/j.enbuild.2004.12.011`. URL: `http://www.sciencedirect.com/science/article/pii/S0378778805000204`.

[57] S. Joe Qin. "Process data analytics in the era of big data". In: *AIChE Journal* 60.9 (2014), pp. 3092–3100. ISSN: 1547-5905. DOI: `10.1002/aic.14523`. URL: `http://dx.doi.org/10.1002/aic.14523`.

[58] B. Samanta and K.R. Al-Balushi. "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features". In: *Mechanical Systems and Signal Processing* 17.2 (2003), pp. 317–328. ISSN: 0888-3270. DOI: `http://dx.doi.org/10.1006/mssp.2001.1462`. URL: `http://www.sciencedirect.com/science/article/pii/S0888327001914621`.

[59] Jeffrey Schein, Steven T. Bushby, Natascha S. Castro, and John M. House. “A rule-based fault detection method for air handling units”. In: *Energy and Buildings* 38.12 (2006), pp. 1485–1492. ISSN: 0378-7788. DOI: http://dx.doi.org/10.1016/j.enbuild.2006.04.014. URL: http://www.sciencedirect.com/science/article/pii/S0378778806001034.

[60] John E. Seem. “Using intelligent data analysis to detect abnormal energy consumption in buildings”. In: *Energy and Buildings* 39.1 (2007), pp. 52–58. ISSN: 0378-7788. DOI: http://dx.doi.org/10.1016/j.enbuild.2006.03.033. URL: http://www.sciencedirect.com/science/article/pii/S0378778806001514.

[61] Sachin Sharma and DG Mahto. “Condition monitoring of wind turbines: a review”. In: *Global Journal of Researches In Engineering* 13.6 (2013).

[62] Robert H Shumway and David S Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.

[63] Ryan Skorupski, R. Lily Hu, Robert Entriken, and Yinyu Ye. “A Benefit-Cost Framework for Optimal Load Shifting”. In: *CIGRÉ (The Council on Large Electric Systems) US National Committee: 2014 Grid of the Future Symposium* (2014).

[64] Georgios Alexandros Skrimpas, Christian Walsted Sweeney, Kun S. Marhadi, Bogi Bech Jensen, Nenad Mijatovic, and Joachim Holboll. “Employment of Kernel Methods on Wind Turbine Power Performance Assessment”. In: *IEEE Transactions on Sustainable Energy* 6.3 (2015), pp. 698–706. ISSN: 1949-3029. DOI: 10.1109/TSTE.2015.2405971. URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-84933043328%7B%5C&%7DpartnerID=tZOtx3y1.

[65] V. Sugumaran, V. Muralidharan, and K.I. Ramachandran. “Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing”. In: *Mechanical Systems and Signal Processing* 21.2 (2007), pp. 930–942. ISSN: 0888-3270. DOI: http://dx.doi.org/10.1016/j.ymssp.2006.05.004. URL: http://www.sciencedirect.com/science/article/pii/S0888327006001142.

[66] N. Taheri, R. Entriken, and Yinyu Ye. “A Dynamic Algorithm for Facilitated Charging of Plug-In Electric Vehicles”. In: *Smart Grid, IEEE Transactions on* 4.4 (Dec. 2013), pp. 1772–1779. ISSN: 1949-3053. DOI: 10.1109/TSG.2012.2233768.

[67] Houari Toubakh, Moamar Sayed-Mouchaweh, and Eric Duviella. “Advanced pattern recognition approach for fault diagnosis of wind turbines”. In: *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*. Vol. 2. IEEE. 2013, pp. 368–373.

[68] Tali Trigg, Paul Telleen, R Boyd, F Cuenot, D D’Ambrosio, R Gaghen, JF Gagné, A Hardcastle, D Houssin, AR Jones, et al. “Global EV outlook: understanding the electric vehicle landscape to 2020”. In: *Int. Energy Agency* (2013), pp. 1–40.

[69] Mini USA. Mini USA Website. URL: http://www.miniusa.com/minie-usa/.

[70] Venkat Venkatasubramanian, Raghunathan Rengaswamy, and Surya N Kavuri. “A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies”. In: *Computers I& Chemical Engineering* 27.3 (2003), pp. 313–326. ISSN: 0098-1354. DOI: http://dx.doi.org/10.1016/S0098-1354(02)00161-8. URL: http://www.sciencedirect.com/science/article/pii/S0098135402001618.

[71] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, and Kewen Yin. “A review of process fault detection and diagnosis: Part III: Process history based methods”. In: *Computers I& Chemical Engineering* 27.3 (2003), pp. 327–346. ISSN: 0098-1354. DOI: http://dx.doi.org/10.1016/S0098-1354(02)00162-X. URL: http://www.sciencedirect.com/science/article/pii/S009813540200162X.

[72] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, and Surya N. Kavuri. “A review of process fault detection and diagnosis: Part I: Quantitative model-based methods”. In: *Computers I& Chemical Engineering* 27.3 (2003), pp. 293–311. ISSN: 0098-1354. DOI: http://dx.doi.org/10.1016/S0098-1354(02)00160-6. URL: http://www.sciencedirect.com/science/article/pii/S0098135402001606.

[73] Shengwei Wang and Jingtan Cui. “Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method”. In: *Applied Energy* 82.3 (2005), pp. 197–213. ISSN: 0306-2619. DOI: http://dx.doi.org/10.1016/j.apenergy.2004.11.002. URL: http://www.sciencedirect.com/science/article/pii/S0306261904001953.

[74] Shengwei Wang, Qiang Zhou, and Fu Xiao. “A system-level fault detection and diagnosis strategy for {HVAC} systems involving sensor faults”. In: *Energy and Buildings* 42.4 (2010), pp. 477–490. ISSN: 0378-7788. DOI: http://dx.doi.org/10.1016/j.enbuild.2009.10.017. URL: http://www.sciencedirect.com/science/article/pii/S0378778809002552.

[75] Wikipedia. *SAE J1772 — Wikipedia, The Free Encyclopedia*. [Online; accessed 6-November-2015]. 2015. URL: %5Curl%7Bhttps://en.wikipedia.org/w/index.php?title=SAE_J1772&oldid=688600018%7D.

[76] Wikipedia. *SAE J1772 — Wikipedia, The Free Encyclopedia*. [Online; accessed 16-November-2015]. 2015. URL: https://en.wikipedia.org/w/index.php?title=SAE_J1772&oldid=688600018.

[77] Wunderground. Wunderground Website. URL: http://wunderground.com.

[78] Wenxian Yang, Richard Court, and Jiesheng Jiang. “Wind turbine condition monitoring by the approach of SCADA data analysis”. In: *Renewable Energy* 53 (May 2013), pp. 365–376. ISSN: 09601481. DOI: 10.1016/j.renene.2012.11.030. URL: http://www.sciencedirect.com/science/article/pii/S0960148112007653.

[79] Wenxian Yang, Peter J. Tavner, Christopher J. Crabtree, Y. Feng, and Y. Qiu. "Wind turbine condition monitoring: Technical and commercial challenges". In: *Wind Energy* 17.5 (2014), pp. 673–693. ISSN: 10991824. DOI: 10.1002/we.1508. URL: http://onlinelibrary.wiley.com/doi/10.1002/we.1608/full.

[80] Seongkyu Yoon and John F. MacGregor. "Fault diagnosis with multivariate statistical models part I: using steady state fault signatures". In: *Journal of Process Control* 11.4 (2001), pp. 387–400. ISSN: 0959-1524. DOI: http://dx.doi.org/10.1016/S0959-1524(00)00008-1. URL: http://www.sciencedirect.com/science/article/pii/S0959152400000081.

[81] Harunori Yoshida, Sanjay Kumar, and Yasunori Morita. "Online fault detection and diagnosis in {VAV} air handling unit by {RARX} modeling". In: *Energy and Buildings* 33.4 (2001). Special Issue: {BUILDING} SIMULATION'99, pp. 391–401. ISSN: 0378-7788. DOI: http://dx.doi.org/10.1016/S0378-7788(00)00121-3. URL: http://www.sciencedirect.com/science/article/pii/S0378778800001213.

[82] A. Zaher, S.D.J. McArthur, D.G. Infield, and Y. Patel. "Online wind turbine fault detection through automated SCADA data analysis". In: *Wind Energy* 12.6 (Sept. 2009), pp. 574–593. ISSN: 10954244. DOI: 10.1002/we.319. URL: http://doi.wiley.com/10.1002/we.319.