IMPACT METRICS FOR RESIDENTIAL HVAC SYSTEMS USING CLOUD-BASED SMART THERMOSTAT DATA

A Thesis

by

PHANI ARVIND VADALI

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Bryan Rasmussen
Committee Members,	Richard Malak
	Charles Culp
Head of Department,	Guillermo Aguilar

May 2022

Major Subject: Mechanical Engineering

ABSTRACT

The aggregation of data from connected smart thermostats installed in a huge number of residential buildings has expedited the remote detection and diagnosis of faults in Heating, Ventilation and Air-Conditioning (HVAC) systems. Upon identification of faults in air-conditioning systems, manufacturers and occupants are interested to know how severe the impact of the faults is on the energy consumption and the thermal comfort of the occupants. Several studies in literature have previously attempted to quantify an energy impact metric and a thermal comfort impact metric of faults in an HVAC system, but the metrics developed lack the ability to be used objectively to compare several systems at once. Furthermore, no study has yet tried to examine the coupled relationship between the energy consumption of the system and the thermal comfort of the occupants to estimate an aggregate fault severity index of a system.

The current study attempts to provide a paradigm shift in the calculation of the energy impact metric. The thesis, firstly, proposes a methodology to model the energy consumption of the average system in a dataset comprising of similarly sized system operating in the same climate region. The performance of each air-conditioning system is compared to the performance of the average system to estimate the amount of impact faults have on their energy consumption. Additionally, the current study also estimates the level of thermal discomfort felt by occupants of the house using the Predicted Mean Vote (PMV) of the indoor environment. The average level of discomfort felt by the occupants living in the house is then compared with a baseline to estimate impact on the thermal comfort of occupants.

The two impact metrics are then combined together into one index that represents the fault severity index of the system which can then be used to rank systems to prioritize them for repair. The severity index of the system is a representation of the relative energy consumption level of the system if it were to produce no thermal discomfort. Another metric that comes as a by-product of this derivation is the amount of change in energy consumed by the system in order to make the indoor environment comfortable. The coupled nature of the four metrics will be delineated so as to

gain an insight into the characteristics of air-conditioning systems. Causes for faulty behavior of systems are examined and systems with mechanical faults are segregated from systems operating under ineffective operating conditions.

DEDICATION

To everyone who believed in me even when I didn’t

"The woods are lovely, dark and deep,

But I have promises to keep,

And miles to go before I sleep,

And miles to go before I sleep."

-Robert Frost

ACKNOWLEDGMENTS

Firstly, thank you, Dr. Rasmussen. Your mentorship and guidance has made this thesis possible. Thank you for teaching me to be persistent and sincere. Your work ethic and approach to research will forever serve as an inspiration. Your constant faith in me was a huge motivation and a pillar of strength during the darkest of times.

Thank you to Dr. Richard Malak and Dr. Charles Culp for agreeing to be a part of my thesis committee.

I should acknowledge all the help Fangzhou gave me on this project. Thank you for being so patient, understanding and kind (and smart!). Majority of this thesis is a product of our meetings and discussions. I should also thank my labmates specifically Daniel and Taeksoo as well as Rafael, Saurabh and Hussein. All the time we spent "cleaning" the lab will remain amongst my most cherished memories.

Thanks to my friends and roommates in town and back home. I never could have made it without you all. Thank you for being there.

Finally, I owe everything to my parents and my brother, without whom I am not. Thank you for all your love and care. Most of all thank you for your unparalleled patience during the course of my study here. I will always be grateful for the innumerable sacrifices you've had to make on my part.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Dr. Bryan Rasmussen (advisor), Dr. Richard Malak of the Department of Mechanical Engineering, and Dr. Charles Culp of the Department of Architecture. Additionally, Dr. Xia Ben Hu of the Department of Computer Science at Rice University served as committee member for few months.

Funding Sources

Graduate study was supported by the J. Mike Walker '66 Graduate Fellowship from the Department of Mechanical Engineering at Texas A&M University, and by Trane Technologies Inc. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Trane Technologies.

NOMENCLATURE

List of Abbreviations

HVAC	Heating, Ventilation, and Air Conditioning
AC	Air-Conditioner
PMV	Predicted Mean Vote
IECC	International Energy Conservation Code
EPA	Environmental Protection Agency
ASHRAE	American Society of Heating, Refrigeration and Air-conditioning Engineers
KDE	Kernel Density Estimation
OLS	Ordinary Least Squares
NREL	National Renewable Energy Laboratory
MRT	Mean Radiant Temperature
MTS	Maximum t-statistic
EPC	Energy Performance Certificate
EPI	Energy Performance Index
EUI	Annual energy consumption per unit ares
EU	European Union
HPLC	Heating Power Loss Coefficient
HTC	Heat Transfer Coefficient
PRISM	PRInceton Scorekeeping Method
IEQ	Indoor Environment Quality
IAQ	Indoor Air Quality
PPD	Predicted Percentage of Dissatisfied

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

According to the Residential Energy Consumption Survey [3] conducted by the Energy Information Administration in 2015, nearly 85% of residents in the US use a thermostat to control their heating equipment and 64% to control their cooling equipment. Since nearly half of the total energy consumed by the residential sector is used for space conditioning [3] residential thermostats are estimated to control close to 10% of the nation's total energy use [4]. With growing number of people working from home there is growing incentive in building thermostats that can improve the energy efficiency of the house while also providing adequate thermal comfort.

The use of thermostats began with deployment of manual thermostats which allowed the user to set a target temperature and controlled the Heating, Ventilation and Air-Conditioning (HVAC) system of the hosue so that the indoor space reaches the target value. With the progress of technology, residential buildings have slowly adopted the use of programmable thermostats which in addition to the above also allow the occupant to setup schedule-based setbacks that help save energy at night and during times when the home is unoccupied. Although 60% of the homes in US are estimated to have installed programmable thermostats [3] field studies have shown that only half of them have been programmed [5]. The amount of savings expected from using programmable thermostats was found to have been exaggerated. As indicated in [4] several studies have found no significant savings upon installation of programmable thermostats, primarily because the occupants of the homes did not use the energy saving programming features [6]. Hence, there has been a greater push in the industry to make the adoption of energy saving measures more pervasive by building programmable thermostats that can leverage modern control techniques.

In order to provide good communication of data and automatic control of home environments, thermostat manufacturers have developed an advanced version of the programmable thermostat called the smart or connected thermostat. Not only do smart thermostat use intelligent algorithms to learn more about occupants' behavior, but also provide occupants feedback on their thermostat settings and energy usage. Based on the manufacturer, smart thermostats can come with a range of extra features in addition to the ones found in programmable thermostats of which the notable ones include: learning algorithms, occupancy sensors, geofencing and software for remote control [7]. However, a common defining feature of smart thermostat is their ability to upload thermostat usage data unto to a cloud over the course of several months, to be available to analysts and engineers to build algorithms for better control and analysis of indoor environments.

Analysis of connected smart thermostat data is increasingly becoming a popular area of study because of the wealth of information they provide. Few of the primary motivations for studying smart thermostat data include: studying occupants' behavior [8], studying potential energy saving opportunities [7, 9, 10, 11], studying thermal comfort levels of homes [12, 13, 14], building models to predict building energy characteristics [15, 16], building smarter control algorithms to improve energy efficiency [17] and thermal comfort [18], building effective demand response programs for demand peak reduction [19] and finally for building algorithms to perform fault detection and diagnostics including weeding out systems with inadequate capacity or gradual capacity degradation.

While fault detection and diagnostic algorithms help in identifying potentially faulty systems, manufacturers and occupants are often interested in determining how severe the fault actually is. The fault severity index of an air-conditioning system is a good indicator of which system to prioritize for repair. Additionally, the occupant will benefit in knowing how much of an effect a change in their usage of the air-conditioning system will have. The severity of a faulty residential HVAC system is proportional to the amount of impact it has on the energy consumption of the house and on the thermal comfort of its occupants. Therefore, in order to quantify the severity of a faulty system, firstly, the impact of a fault on the occupant in the two aforementioned areas must be quantified.

Studies on using the data from connected thermostats to simultaneously analyze the system's energy consumption and occupants' thermal comfort have primarily been in the area of developing smarter control strategies. Merabet et al. [17] provides a comprehensive review of literature on history of studies of control algorithms for smart thermostats. However, the objective of constructing an impact metric is not to build a better control strategy but to provide a means to interpret the data with a different perspective.

The following thesis first provides context by giving background on the work done by previous researchers of Dr. Ramussen's group who have worked on this project. Chapter 2 therefore, introduces the nature of the data, pre-processing procedures that were used to prepare the data and how the data could be used to develop statistics based fault detection and diagnosis procedures. Subsequently, existing literature is reviewed in chapter 3 to identify the inadequacies of existing energy and thermal comfort impact metrics and proposes few general guidelines to fill the gaps. Chapter 4 describes a construction of an energy model of a given system by examining its smart thermostat data and compares it to a baseline to understand its relative energy performance. Similarly, chapter 5 compares the Predicted Mean Vote (PMV) of thermal comfort level of the indoor environment at various instants in time to what the PMV would have been in the absence of an air-conditioning system to construct a thermal comfort impact metric. Subsequently, chapter 6 proposes a method for combining the two mutually dependent metrics previously developed to estimate a fault severity index for an air-conditioning system. This will be done by examining the difficulty in reducing the thermal discomfort to zero and its subsequent effect on system's relative energy performance. Chapter 7 delineates the relationship between the four metrics established in the three former chapters and examines how the metrics can be correlated with the behavior of the air -conditioning system. Finally, chapter 8 explores a few examples from each of the cases developed in chapter 7 and provides a quantitative understanding of the metrics. Chapter 9 is a summary of the methods discussed in the thesis and also explores possible avenues for future work on this topic.

2. OVERVIEW OF LARGE SCALE DATA ANALYTICS FOR RESIDENTIAL HVAC SYSTEMS

This project titled, "*Large Scale Analytics for Residential HVAC Systems*", was sponsored by Trane Technologies Inc., from April 2018 to December 2021. The following chapter is summary of the contributions made by two previous students, Fangzhou Guo [2] & Austin Rogers [1], who have worked on this project previously. The thesis that follows in the rest of the chapters is built on the fundamental ideas established by them. All materials reproduced in this chapter have been done so with permission from the original authors.

The proposed study was conducted on data acquired form smart thermostats installed in residential buildings provided by Trane Technologies Inc. The dataset contains event-based time series data from nearly 400,000 HVAC systems installed in residential buildings across the United States. The key objective of the project was to use the data to build fault detection and diagnosis (FDD) methods for predictive and preventive procedures. Traditionally, FDD of residential split systems warrants the installation of additional sensors on the equipment making it cost-ineffective to be implemented on a large scale especially for mass produced residential systems. This project provides an alternative FDD approach wherein a large number of systems will be analyzed using only the limited amount of sensor information available from smart thermostats from each system. Analyzing smart thermostat data has a number of benefits for not just the home occupants but the manufacturers as shown in fig. 2.1.

Although smart thermostat data does not contain a lot of diversity in terms of parameters, because of its large scale statistics-based or artificial intelligence based algorithms can be developed for fault detection and diagnosis. Interesting features extracted from data from large amount of systems are used to build the algorithms and faulty systems found thereby are ranked based on the severity of faults detected. A general overview of the FDD process developed as part of the project is given in fig. 2.2.

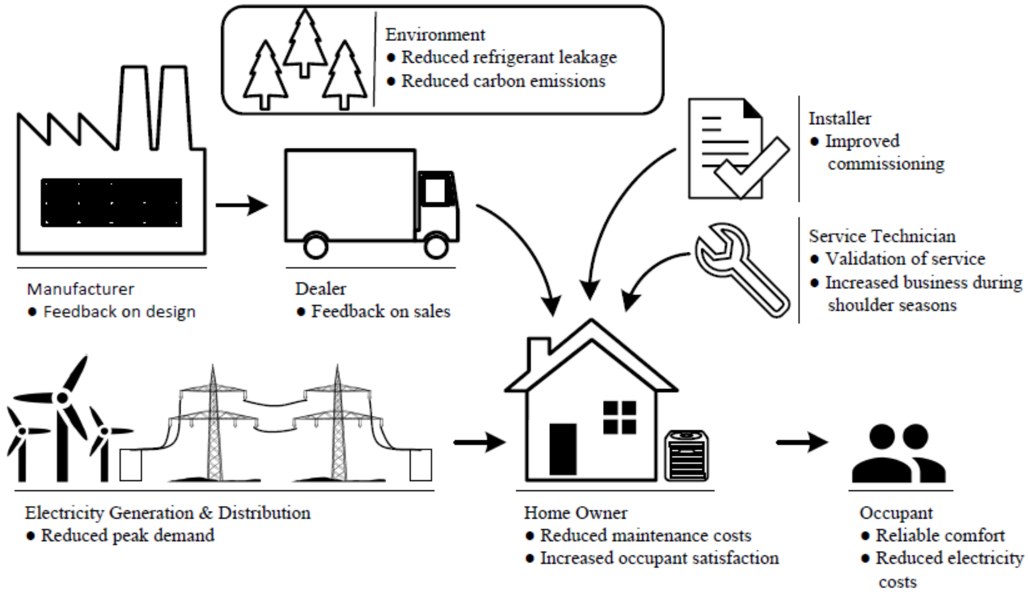


Figure 2.1: Benefits of using FDD for residential HVAC systems. Used with permission from [1].

2.1 An introduction to the data

The data queried from the cloud consists of two sets of parameters for each system which include configuration parameters and operation parameters. Configuration parameters provide information regarding the kind of HVAC system installed in the house while the operation parameters describe the state of the indoor environment and the system in time-series form with each data point being added as and when the state of the system changes. The table 2.1 illustrates a few of the key parameters of each type.

Table 2.1: Few key parameters in the smart thermostat database provided by Trane Technologies.

Configuration Parameters	**Operation Parameters**
Indoor/Outdoor unit type	Indoor/Outdoor temperature
Indoor/Outdoor unit stages	Indoor/Outdoor Humidity
Indoor/Outdoor capacity	Cooling/Heating Setpoint
System Location	Indoor/Outdoor unit operating conditions

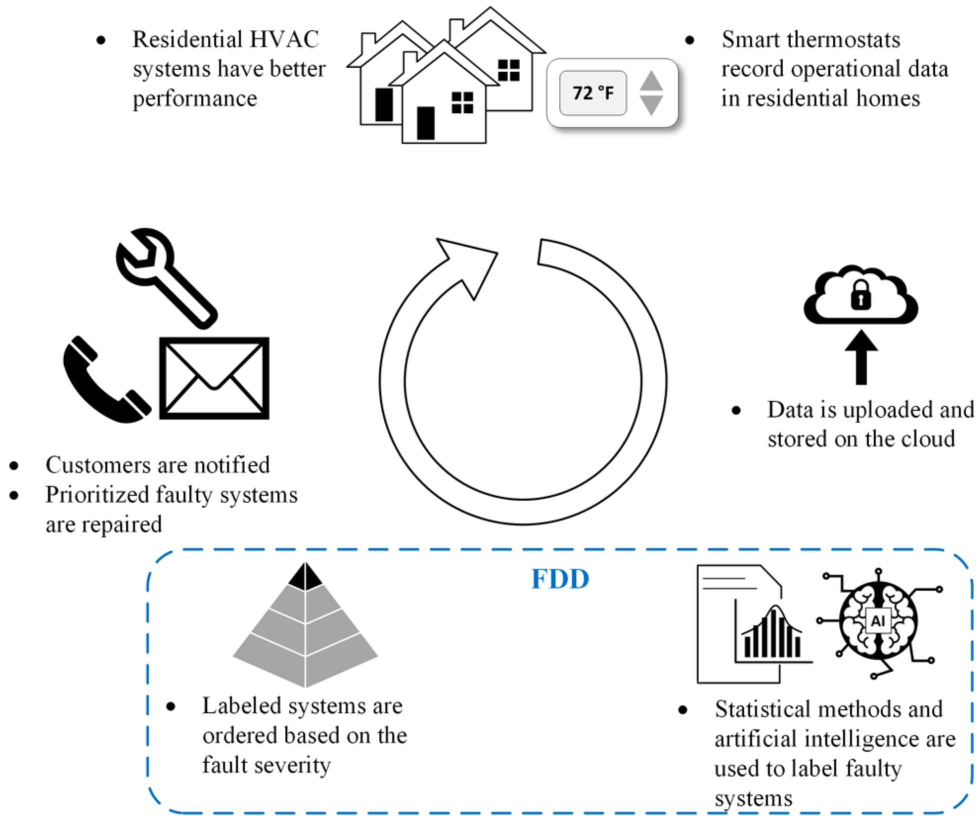


Figure 2.2: General overview of the FDD process. Used with permission from [2].

2.2 Data Selection and Cleansing

Since, data from thousands of system all across the United States is available for analysis an appropriate set of systems should be selected so the analysis is objective. Systems lying in the same climate zone must be grouped together because the local climate determines the size of the system. Comparing two systems that lie in different climate regions will lead to ambiguous results. For example, the cooling units lying in climate region 6A which comprises of states primarily in the north eastern part of the US will be of smaller size than systems lying in climate region 2A which belongs in the state of Florida because the systems in Florida will experience much severe summers in comparison.

Another factor to keep in mind while selecting data is the type of air-conditioning system. Since, the operational behavior changes drastically between single-stage systems, dual-stage systems and variable-speed systems, comparing systems of one type only will keep the comparison objective. Finally, an appropriate time scale should also be selected. Selecting a short time-scale will have significant drawbacks because the full range of ambient conditions will not be captured. Additionally, the data may be affected by irregular occupant behavior (leaving the window open for long periods of time, for example) or other irrelevant covariates. However, a large time-scale will lead to inaccurate results as well because the operating condition of the system may have changed by the end espcially due to degradation or repairs.

The data from all the systems is queried from the cloud and cleansed before analysis. Cleansing operations involved removing data points that raise sensor integrity issues and those that convey network communication errors. A rule-based filter was used to remove faulty datapoints, the criteria for few of which is given below (courtesy of [2]):

1. The indoor or outdoor temperature climbs above $130°F$ or goes below $0°F$ (occasionally the thermostat will list the temperature as $2^{15} = 32,768°F$ when the thermostat goes offline).

2. The temperature increases or decreases by more than $20°F$ between two consecutive updates. This is often caused by sensor integrity issues.

3. The indoor or outdoor temperature has not updated for more than four hours. If the sensor fails to update for a period of time and then suddenly resumes, then removing these communication failure periods is necessary.

4. The system maintains a constant temperature with little variance (tightly), but the setpoint error is large. In this case the reported setpoint is likely incorrect.

5. The system runs for more than 24 hours. In most cases this is caused when the system “off” signal is not received.

2.3 Mode Labelling

After querying and cleansing the data, the data from each system was divided into its modes of operation. These modes were constructed based on the states of the system: steady-state mode observed when the system is cycling on/off maintaining a constant indoor temperature (called the *regulating mode*), transient mode observed when the system is actively cooling to reach a new setpoint (called the *tracking mode*) and the *free response* mode which is when neither setpoint is active and the system is switched off. Figure 2.3 shows the data from the system partitioned into various modes as well as the time series variation of the state of the system.

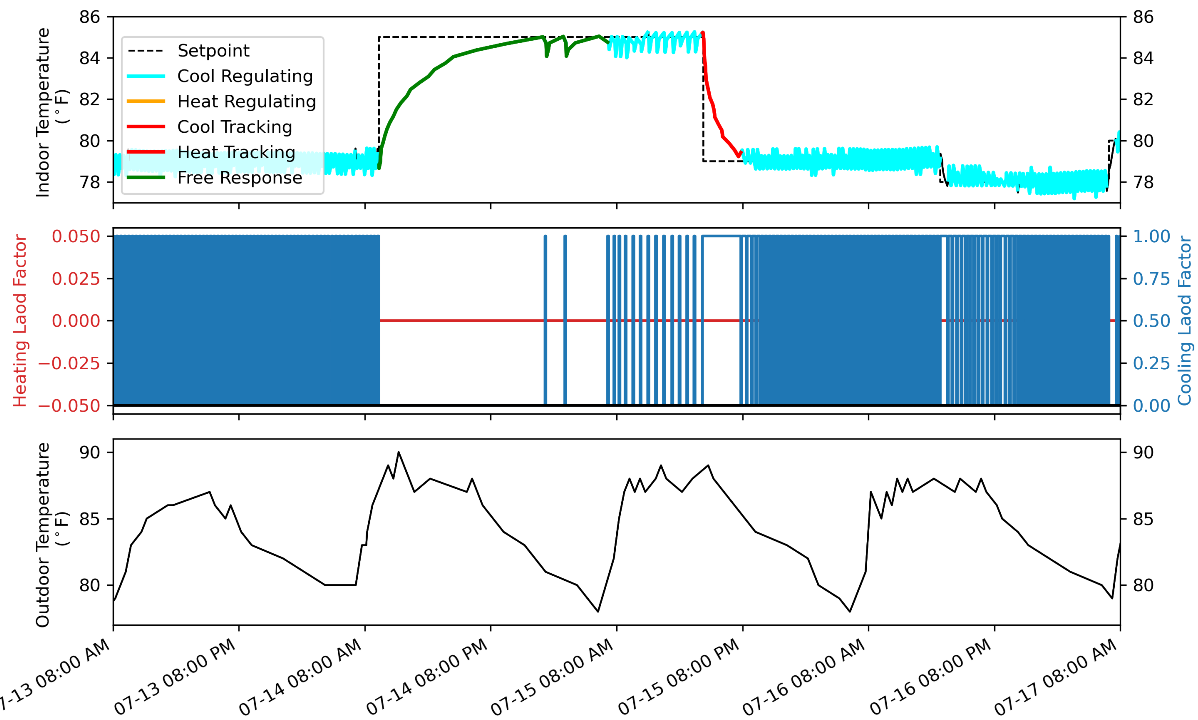


Figure 2.3: Time series data from a single-speed air-conditioning system in Florida along with modes of operation.

The modes labelling algorithm which divides data from each system into modes uses several rule-based filters and was developed by two former Ph.D. candidates who have worked on this project previously, Austin Rogers and Fangzhou Guo, and it can be found here [20].

2.4 Feature Extraction

Since data from each system has been partitioned into modes, features describing the properties of the systems can now be extracted. Fault detection is done on the basis of features that are extracted from each of the modes. The list of features is ever expanding because there is always a chance to discover feature that can better characterize the system. Few of the features that are extracted from each mode are given in tables 2.3, 2.2. A detailed list of features and definitions can be found in [2]. Two tables explaining the features are attached here below.

Table 2.2: Features extracted from behavioral modes. Used with permission from [2].

Symbol	Feature name	Behavioral modes			
		Tracking	Free Response	Regulating	All
ΔT_{oi}	Temperature difference	X	X	X	X
$\Delta\omega_{oi}$	Humidity difference			X	
ΔT_{ii}	Indoor temperature rise	X	X		
Δt	Cycle duration	X			
D	Duty factor			X	
L	Load factor	X		X	
E_c, E_h	Cooling/Heating effort			X	X
F	Cycle frequency			X	
T_{spe}	Setpoint error	X		X	
DH	Degree hour	X			
$T_{i,max}$	Max indoor temperature	X			

Table 2.3: Definition of Features. Used with permission from [2].

Symbol	Unit	Definition
ΔT_{oi}	[°F]	Temperature difference between home (indoor) and ambient (outdoor); defined as $\Delta T_{oi} = T_o - T_i$.
$\Delta\omega_{oi}$	[kg H_2O/ kg dry air]	Humidity difference between home and ambient; defined as $\Delta\omega_{oi} = \omega_o - \omega_i$.
ΔT_{ii}	[°F]	Indoor temperature rise (or drop) in a given time period: $\Delta T_{ii} = T_i(t_2) - T_i(t_1)$ in a time period from t_1 to t_2.
Δt	[hr]	Cycle duration. The amount of time that a system remains on.
D	[%]	The portion of time that the system is running, and this is typically defined for a given time period. For example, a duty factor of 40% in a 1-hour period denotes that the system ran for 24 minutes in that hour. However, a variable speed system may have a duty factor close to 100%.
L	[%]	The portion of full capacity at which the system runs. When defined over a given time period, this is the average portion of full capacity that the system runs at *when the system is on*. A single-stage system essentially always has a load factor of 100%.
E_c, E_h	[%]	Defined by the product of D and L, which denotes the overall portion of capacity for the given time period. Variable speed systems will typically have a high duty cycle and a low load factor, while single stage systems will have high load factor and low duty factor.
F	[cycle/hr]	The number of start/stop cycles per hour.
T_{spe}	[°F]	The error between indoor temperature and setpoint; defined as $T_{spe} = T_i - T_{spc}$ for cooling and $T_{spe} = T_{sph} - T_i$ for heating.
DH	[°F·hr]	The integral of setpoint error, during a time period from t_1 to t_2; defined as: $DH = \int_{t_1}^{t_2} \max(0, T_{spe})dt$.
$T_{i,max}$	[°F]	$T_{i,max}$ is defined as the maximum indoor temperature during a cooling tracking period, and correspondingly $T_{i,min}$ is defined as the minimum indoor temperature during a heating tracking period.

2.5 Statistics Based Fault Detection Methods

Data queried from the cloud is processed and partitioned into modes of operation after which features are extracted from each of the modes. Faulty systems were then identified by examining the statistical distributions of the features in each of the modes. Five detectors were developed as a part of this project by Dr. Fangzhou Guo [2] and Dr. Austin Rogers [1]. The detectors were firstly developed for cooling systems by the original authors upon which they were expanded to heating systems by the author of this thesis. A summary of their functionalities as explained by the original authors is given below:

2.5.1 Setpoint Tracking Failure Detector

This detector analyzes the transient behavior of the system. Feature of the system from the cooling tracking mode of the system are analyzed and a detailed description can be found in [21]. A single cooling tracking mode is continuous period of time wherein the system is actively cooling. Systems with normal transient behavior are ones where the cooling system switches on after a drop in setpoint and actively cools until the indoor temperature reaches the new setpoint. Periods of poor transient behavior include when the indoor temperature increases by over $10°F$ even though the system is cooling continuously for 24 hours. The plot in fig. 2.4 compares a normal tracking period and a poor tracking period. Features used to identify these poor tracking periods are the average degree hour over setpoint and average maximum indoor temperature. Average degree hour above setpoint is the average area between the indoor temperature and the setpoint temperature in all poor tracking periods of the system. Average maximum indoor temperature is an average of the ten highest maximum indoor temperatures recorded in poor tracking periods after removal of 5 most extreme instances. Normal system have a average degree hour above setpoint equal to 0 and average maximum indoor temperature lower than $80°F$. Systems that show abnormally high values for both of these features are identified as systems that could possibly be faulty.

2.5.2 Inadequate Capacity Detector

As the name suggests this detector identifies systems with inadequate capacity, specifically by examining the cooling regulating mode. During the cooling regulating period the system is cycling on and off maintaining a constant indoor temperature. Since the system is in pseudo steady state the amount of cooling provided by the cooling system is equal to the total heat load that the house experiences. The heat load comprises of the internal load generated by the occupant, the solar

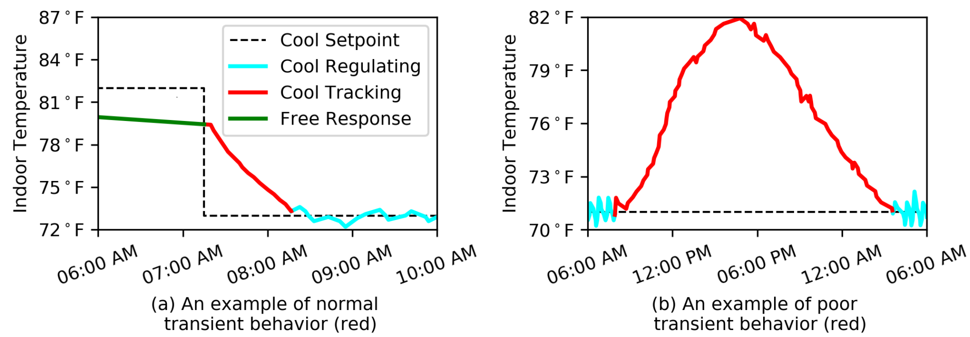


Figure 2.4: Illustration of (a) normal transient behavior (b) poor transient behavior in cooling tracking mode (in red). Used with permission from [2].

load and the heat load due to conduction of heat from outside. Under normal circumstances the systems that see a larger solar and internal load are assumed to have been fit with a system of larger capacity. However the heat load will change with changing outside temperature and so does the amount of cooling provided by the system. The amount of cooling provided by the system in a regulating period is proportional to the cooling effort of the system in that period. Cooling effort (E_c) is the product of duty factor and load factor as defined in table 2.3. In case of single speed systems that have a load of either 1 or 0, the cooling effort is essentially equal to the the duty factor. The cooling effort which is extracted from all the regulating periods of the system is the feature used to build a statistical distribution for comparing systems.

The detector uses the joint probability distribution of cooling effort and temperature difference as built using data aggregated from all the systems to construct what is called as a characteristic curve. The characteristic curve represents the average performance of all the systems or the expected cooling effort of the average system at every level of temperature difference. Subsequently, weighted average of distance of cooling effort values of a single system to the characteristic curve is calculated using the probability density function of temperature difference values as weight. Systems that have very high value of weighted average difference of cooling effort have cooling effort values that lie far away in comparison to how systems in the dataset perform on average.

Hence, they're cause of concern and possible cases of system with inadequate capacity to keep with the demands of the house. Detailed description of this detector can be found here [2].

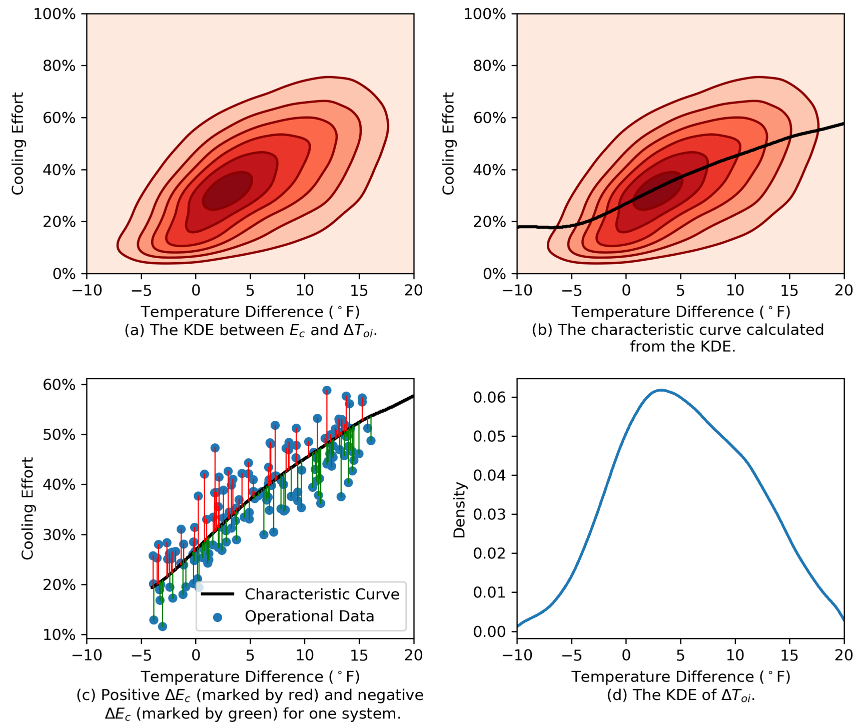


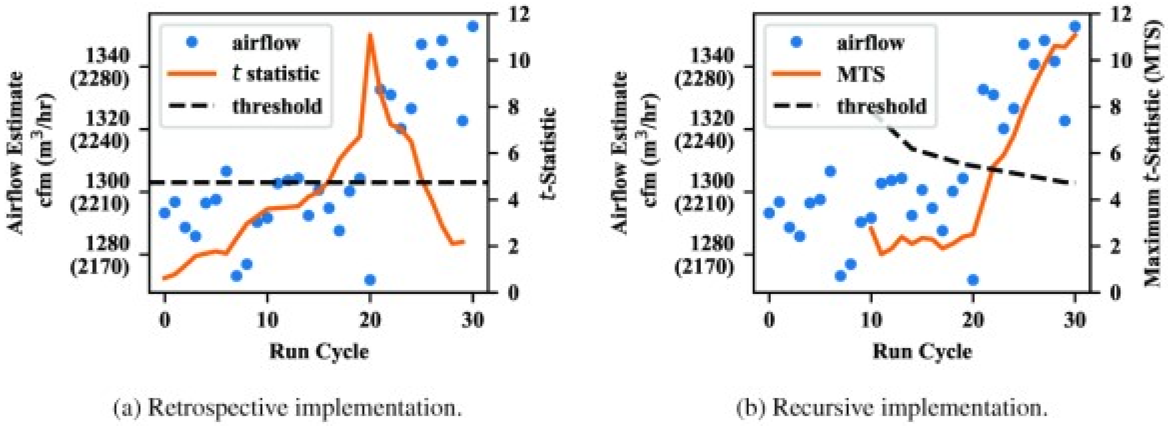
Figure 2.5: Kernel Density Estimation (KDE) was used to build the probability density function in (a). Weighted average distance is calculated from (c) and then examined to identify systems with inadequate capacity. Used with permission from [2].

2.5.3 Abrupt Change Detector

Once a system is detected with inadequate capacity, tracking back the historical data to investigate its past behavior becomes necessary. The inadequacy of capacity could be either caused by incorrect equipment model selection or improper commissioning, which exists before the actual usage of the system, or caused by equipment aging, refrigerant loss, air duct blockage, and even occupant faults, which are accumulated problems after actual usage. A system with only the former category of faults should show very little behavioral change throughout its operational data, while a system with either the latter category of faults or both should show significant behavioral change.

The Abrupt Change Detector is able to find sudden changes of behavior, followed by the Degradation Trend Detector will be introduced to find gradual changes of behavior. The Abrupt Change Detector could detect changes of a time series in its mean value and is applicable to both thermostat data features (e.g., the Cooling/Heating Effort) and diagnostic module features (e.g., the airflow rate). The method is based on the t-statistic which is commonly used to determine whether two samples are chosen from the same population. First of all, the t-statistics are calculated at every possible bisection of a given time series. The bisection that yields the maximum t-statistic (MTS) will be associated with the instant in time at which a change is most likely. Because the MTS will not follow the traditional Student's t distribution, secondly, Monte Carlo simulation is used with kernel density estimation to identify its probability density and appropriate thresholds for the MTS are then chosen using the probability density. Finally, a significant change is detected if the MTS of the time series is higher than the threshold.

The detector could be applied either retrospectively or recursively. In the retrospective implementation, a past period of data is analyzed in order to determine if a change point exists within that past period. In comparison, the recursive implementation is more suited to be applied to the most recent data after new data becomes available, and therefore can detect a new change as soon as possible.



(a) Retrospective implementation.

(b) Recursive implementation.

Figure 2.6: Application of the abrupt change detector to the time series of airflow rate. Both retrospective implementation and recursive implementation are shown. Used with permission from [1].

Moreover, the detector is capable of detecting multiple change points in a time series. After one change has been detected, the primary window will be selected from the location of the detected change point. To improve the localization of the change point, a secondary window that includes the change point is maintained until the left and right sample sizes are balanced or until another change is detected. Detailed information can be found here [22].

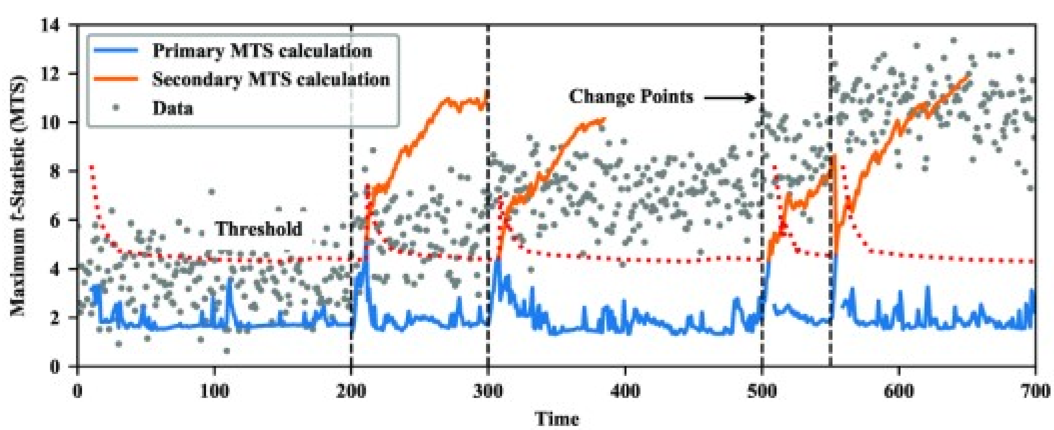


Figure 2.7: Handling of multiple change points. Used with permission from [1].

2.5.4 Degradation Trend Detector

The former fault detection method, the Abrupt Change Detector, is designed to find sudden changes in a time series of operational data. In contrast, the Degradation Trend Detector is designed to find gradual changes in a time series, especially the gradual degradation trend of system capacity.

According to the domain expertise, as the capacity of a system gradually degrades, the Cooling/Heating Effort and the Daily Cooling/Heating Hours at the same ambient conditions are supposed to have an increasing trend. Based on this, the Degradation Trend Detector applies the Mann-Kendall test, a robust non-parametric trend detection method, to find increasing trends of these features. In reality, the ambient conditions cannot be stable: the outdoor temperature is not constant, internal heat gain varies, and setpoint is often changed by occupants. Therefore, the Mann-Kendall tests are modified to account for those conditional variables that can affect the test features.

When the Cooling/Heating Effort is used in the test, the data is subdivided into several subsets according to the range of temperature difference, and tests are performed at the subset level. On top of that, modified versions of the Mann-Kendall test are applied to account for the serial correlation of each dataset. When the Daily Cooling/Heating Hours is used in the test, the Partial Mann-Kendall test is applied to eliminate the influence of the outdoor condition and setpoint variations and the serial correlation of features. In-depth discussion of this detector can be found here [23].

2.5.5 Control Problems Detector

Features from cooling regulating mode can also be used to identify systems control problems. Feature of the regulating period namely the cooling effort, cycle frequency, and cooling setpoint error data of a group of systems are extracted. The features are then subdivided into subsets as per the values of temperature difference and indoor humidity observed in each of the regulating periods. A two-part fault detection method was built to compare the features among systems to find outliers. In the first part, each feature is compared individually by estimating its total sample distribution using kernel density estimation, and the outliers are data points in the abnormally low

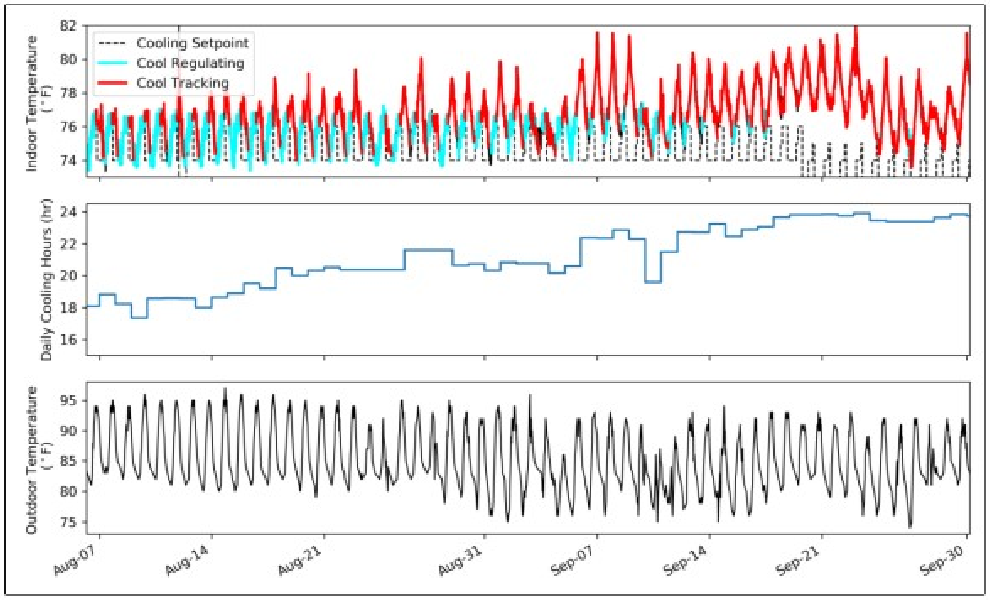


Figure 2.8: Daily operating condition of a flagged system. The Daily Cooling Hours is used as a feature. Used with permission from [2].

density region; and in the second part, features are combined together as a multivariate, and the Mahalanobis's distance is applied to isolate the outliers. Finally, the outliers with high cooling effort, high cycle frequency, and high cooling setpoint error are ranked, and given an adjustable quantile (i.e., top 0.5%), the top ranked systems within these outliers will be flagged as anomalies. Because the flagged systems have abnormally high compressor cycling frequency and regulating setpoint error, the problem could be related to the control system not working or thermostat misplacement. Plus, the cooling effort added into the detector helps a comprehensive analysis by also taking the capacity check into consideration. In depth explanation of the this detector can be found her [24].

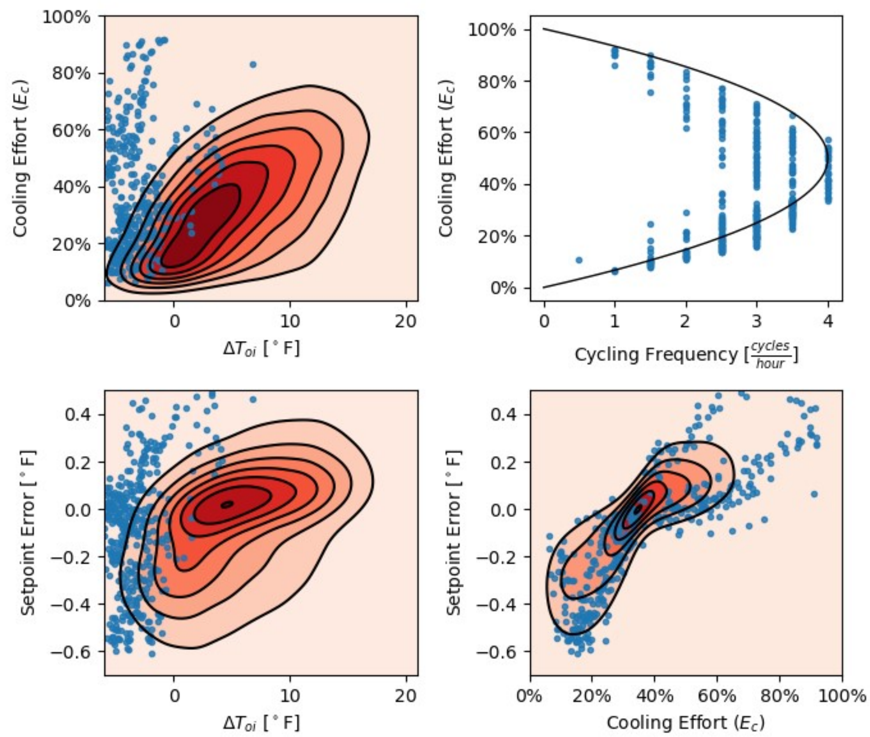


Figure 2.9: Features of an air-conditioning system flagged by the control problems detector. Used with permission from [2].

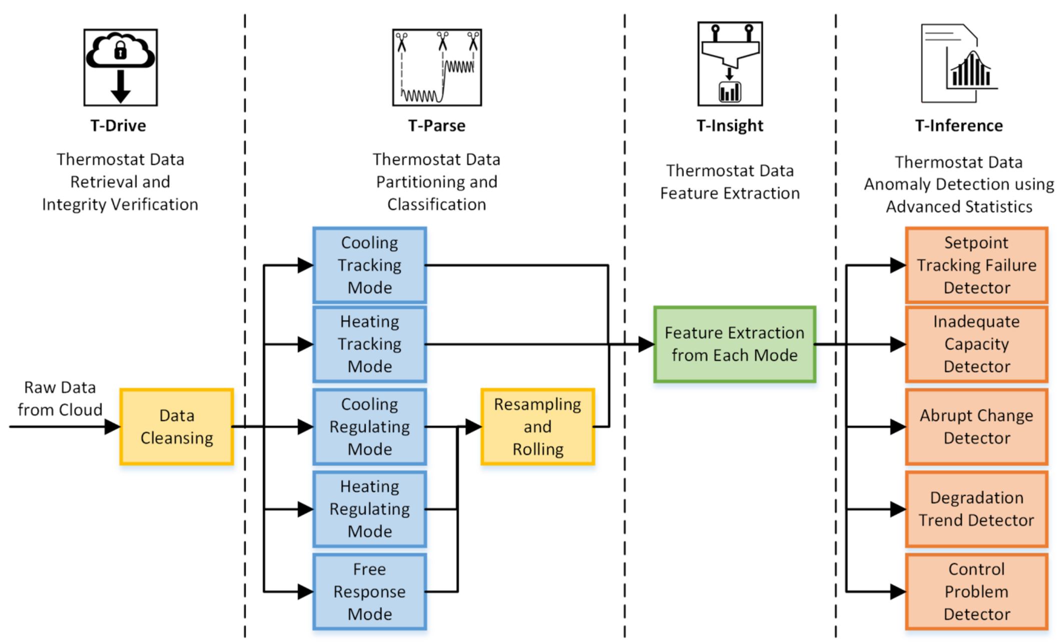


Figure 2.10: Summary of each of the steps involved in the FDD process. Used with permission from [2].

Since a fault detection methodology has now been established the next step in the process is examining what effect the faults have on the occupants. The rest of the thesis explores this question by estimating the impact of faults on energy consumption and thermal comfort of occupants. Finally, a method to combine the impact metrics and determine a severity index for the system will also be presented. In the process the cost and effect of improving the thermal comfort level of the system will also be examined.

3. BACKGROUND & OBJECTIVES

Implementing energy efficiency measures in buildings helps reduce energy costs while maintaining or improving occupant comfort [25]. In addition to identifying the right measures to be taken, an important step in the process is providing feedback to occupants on their usage of the system. This process also referred to as eco-feedback [26] has been expedited considerably by the increased establishment of smart thermostats and home energy management systems.

The process of eco-feedback has typically been employed to inform the occupant of their impact on energy consumption [26]. The process of designing such an eco-feedback systems relies on the appropriate benchmarking of energy consumption of buildings. Building owners as well as policy makers use benchmarking to help identify buildings that need retrofits or buildings that have potential for improvement [27]. However, benchmarking energy more often than not ignores the comfort level of homes which is equally important because of its influence on occupants' health and productivity [27].

The process of benchmarking involves comparing the existing performance of an air-conditioning system, as indicated by both the energy consumption and thermal comfort, with that of a baseline to evaluate its behavior. The baseline that's used for comparison falls into one of two categories:

1. How would the given air-conditioning system perform at baseline conditions?

2. How would a baseline air-conditioning system perform at the given conditions?

The former case could involve comparing the performance of the given system to its own behavior in the past or comparing the performance of the given system at the current set point level to a hypothetical scenario with fixed set points. In case of the latter, a baseline air-conditioning system will be chosen based the observed data and the system performance will be compared to this baseline. Choosing such a baseline can be tricky. One approach is to compare the performance of the air-conditioning system with that of a fixed system that represents the national average [28]. This method while being easier still has to ensure that the baselines are updated especially when considering energy consumption with changing energy demands, fuels prices and other factors [28]. The second method is comparing the performance of two existing air-conditioning systems with each other but this does not help provide an objective worth of the system's performance. A good middle ground between the two is modeling the performance of an archetypal building at the location of the given house and using that as the baseline [28]. This research proposes a baseline based on this concept but goes further to try and attribute the impact to building factors or occupant factors. The following sections in this chapter will discuss how a baseline will be built in each case and how the baseline can used to benchmark the impact of energy consumption and thermal comfort.

3.1 Energy Impact metric

Benchmarking of buildings for energy consumption is typically done by constructing Energy Performance Certificates (EPCs) which give the buyer or the tenant accurate information on the energy performance of the building [28]. Energy Performance Certificates are constructed with the help of indices that quantify relative energy performance called as Energy Performance Indices (EPIs) [29]. One such index that is commonly used in practice is the Energy Usage Intensity (EUI), which is the annual energy consumed by the building per unit gross floor area [27]. EUI attempts to make the comparison of buildings of different sizes fairer by normalizing them by their area. Normalization of buildings is necessary because directly comparing the energy consumption of the two buildings can paint a false picture because non-uniform sizes and usage patterns. In order to avoid any potential pitfalls, benchmarking and comparison are only done for buildings of similar characteristics, of similar size and situated in similar climate regions [29]. In addition to gross floor area, normalization of buildings can also be done by person or person-day [27]. However, irrespective of the parameter(s), normalization is always accompanied by some loss of perspective. For example, by completing the basement, a particular home may have lesser EUI than a home of the same size without a basement, but the comparison in that case is misleading because basements typically consume very less energy in comparison to the rest of the house [30]. Ueno

[30] describes in detail further problems that could arise with using EUI as an energy performance metric and other normalization methods that could be used.

Building energy benchmarking tools that exist today such as the Energy Star Portfolio Manager [31] or the Building Performance Database [32] are typically constructed using regression models on databases comprising of a variety of buildings in the US. The Energy Star Score [33] which is a popularly used benchmakring tool uses a regression model to give each building a score between 1 and 100 based on comparing the given source EUI and the EUI predicted by the model. As stated previously studies have identified the problems with the existing EUI-based metrics [30] and Guillen et. al. [27] go further to demonstrate the variation in benchmarking results due to different metrics. Various statistical methods have been used to improve the regression-based approaches few of which are: using a more diverse set of variables, building complex probabilistic models, clustering the data to find common sets of buildings and then building a regression model in each set [26]. The regression model built for the house is by the means of data collected typically through one of two processes: calculation-based approaches which use the thermophysical characteristics of the buildings, climate and occupant behavior to estimate the energy consumed or measurement-based approaches that are based purely on metered energy data without regards to the building characteristics or occupant behavior [28]. Calculation-based approaches involve a bottom-up approach wherein energy performance of components is modeled individually with the help of a few approximations without regards to the actual energy use which may vary considerably. On the other hand, measurement-based approaches while accurate, do not collect any auxiliary data making the separation of building performance from other factors such as building characteristics and weather very difficult. Since both of the approaches aren't fully accurate and are susceptible to errors the incorporation of smart technology in the form of smart meters or smart thermostats during the recent times has helped bridge the gap as shown by studies done especially in the European Union (EU) [28].

Smart meters measure accurate electricity/gas consumption data with a high temporal resolution. They are step better than using irregular manual meter readings in that no further adjustment or modelling is required before using for analysis [28]. The high temporal resolution nature of the data allows for it be examined against the external weather data or solar irradiance data observed at the location to understand building characteristics. In case of heating systems researchers in the EU have defined two metrics that use smart meters to characterize and compare houses: heating power loss coefficient [34] and heat transfer coefficient [35]. Heating power loss coefficient (HPLC) is the amount of fuel required to maintain a given temperature difference between inside and outside of a building and heat transfer coefficient (HTC) is the amount of heat required to maintain a given temperature difference. Depending on what metric is measured by the smart meter (fuel consumed/heat energy consumed) either of the metrics can be used as a way to characterize the thermal losses due to poor house properties (in case of HTC) or thermal losses due to properties of the house and heating system (in case of HPLC).

Determining either of these metrics empirically can be done by building a mathematical model of the smart meter data. A linear regression model between the daily average fuel power and external temperature of the house represents the simplest model that was built first used in the PRInceton Scorekeeping Method (PRISM) [36] and subsequently improved in the Inverse Modelling Toolkit [37]. However, the linear model is susceptible to inaccuracies due to its exclusion of solar load and internal loads of the house. Accounting for solar load especially, is a challenging task because though solar irradiance value at a location can be estimated, the solar load the house experiences is also a function of the orientation of the house regarding which sufficient data is not available. Under the assumption that heat provided by the heating system is very high in comparison to the heat addition due to solar load and internal load, the linear model can be used to estimate HPLC (as done by [34]) or the HTC (as done by [38]).

Although smart meters are a good source they lack diversity, restricting themselves to only energy consumption data. Estimating building characteristics from just smart meter and weather data so as to compare systems against each other could be inaccurate because they come from different sources and need to be matched. Smart thermostat data on the other hand goes one step further by reporting indoor temperature and relative humidity data as well as the state of the cooling/heating system thereby allowing for a much more precise estimation of energy savings. With the growing installation and employment of smart thermostat in residential buildings across the US, a wealthy trove of data has become available that can be used to characterize and compare buildings.

Existing literature points to various studies that have used smart thermostat data to build an energy impact metric with the most notable and widely used one being the Energy Star metric developed as part of the EPA Energy Star © Connected Thermostat Program [10] [1]. The Energy Star metric estimates the amount of energy savings possible to compare houses. The basis of such a metric as discussed in the introduction to this chapter involves the comparison of the performance of the given system with a baseline. In case of the Energy Star metric the baseline used was put forward by Urban and Roth [11] and it compares the performance of the system with the how the system would have performed at a constant baseline temperature specific to the system. In the case of heating they suggest the baseline temperature as the 90^{th} percentile of indoor temperature and in the case of cooling they suggest the baseline temperature as the 10^{th} percentile of indoor temperature. This was proved to be better than directly assuming the maximum and minimum temperature as the baseline temperatures respectively because it takes care of random spikes in temperature and will likely represent a comfortable temperature with the home being occupied. An improvement to this baseline was suggested by Huchuk et. al. [13] used data collected from people enrolled in EcoBee's Donate Your Data program and segregated users into two groups: users that overrode their setpoints frequently and users that didn't. Since, most users override their setpoints to a more energy intensive value, in the case of the latter group, Huchuk et. al. propose a new baseline performance for each system which is what their setpoint schedules would have been had there been no override. This method will not work for users who frequently override their setpoints because determining what the original setpoint would have been is very difficult. So, for those users they propose using the same baseline as the one used by the EnergyStar metric.

[1] Please note that this Energy Star score is specifically for smart thermostats and is different from the one built for benchmarking in the Energy Star Portfolio Manager [31].

A key issue, however, in using a baseline specific to each system is that it is not objective. In order to compare the energy savings from two systems, a common baseline is desired and assuming a baseline that's specific to that system makes objective comparison difficult. Furthermore, such a baseline does not make use of the wealth of information that the smart thermostat data provides. A dataset of systems can have systems that are all from the same region having similar specifications (single speed, variable speed etc.), installed in houses of similar size; so, there is an incentive to build a baseline that leverages the data from all these similar kinds of system that can be used objectively to compare systems within that dataset.

Analogous to smart meter data, comparing systems on the basis of smart thermostat data requires the construction of a mathematical model as well. The model can be used to estimate system performance at baseline conditions which can then be compared against its performance at the given conditions. In order to calculate the energy consumption of a house under the two scenarios, a widely employed technique is to run building energy simulations of an archetypal model of a house with a similarly sized HVAC system using softwares such as EnergyPlus [39] and BEopt [40]. The results are then extrapolated to simulate the performance of all the houses in the dataset. A good example of such a study was done by Booten et al [41] who calculated the energy impact of thousands of web-connected Trane smart thermostats by taking a typical model of the house in each climate region and conducting simulations in BEopt. A detailed review of all studies that used the above software tools and the baselines used to estimate the energy savings is given in Pang et al [7].

Unfortunately, conducting simulations using the above software tools for all the systems in the dataset has the underlying drawback of being computationally intensive and possibly inaccurate. Smart thermostats often only provide the thermostat usage data and not necessarily any metadata regarding each house. Since, no information regarding the orientation, model and layout of the houses is known, conducting accurate building energy simulation studies is difficult. To simplify the process of constructing a mathematical model the Energy Star metric for connected thermostats builds a linear regression model between daily average heating/cooling runtime (amount of time the system is switched ON) and the daily average temperature difference between indoors and outdoors. This model is similar to those built for smart meter data but have the advantage of being more accurate because of the inclusion of indoor temperature which was previously unavailable. The slope of the linear relationship represents the ability of the house to let heat pass through the walls (i.e., insulation level) and the intercepts refers to all the heat gained in the house due to occupant's body heat, heat from the appliances and the solar load. Since the amount of cooling/heating energy given to a house is directly proportional to the amount of time the system runs, the coefficients of this linear model are found by minimizing the square of the difference between the actual running time and the running time predicted by the model across all the days of interest [10]. Using this relationship and the baseline proposed by Urban and Roth [11] the energy savings metric is calculated as the difference between predicted baseline running time and the actual running time.

This method of simulating the performance of a system however, reduces the variation of temperature and runtime in a day to one number which smooths out a lot of the effects of intermittent changes and the model built is likely to be incomplete. Furthermore, the range of temperature difference values considered will be low because over the course of usage of a cooling system over 3 months in the summer, the variation in temperature difference values on daily basis will not be high. The same variation across the day can however be broader. So, methods that can utilize this variation can be expected to be more accurate.

In order to categorize the severity of faults in a system, both its impact on energy consumption and its impact on thermal comfort need to be examined. The EnergyStar metric represents the accepted state of the art for quantifying the impact of systems on the energy consumption of the house. The next section discusses metrics that have constructed to examine occupants' thermal comfort level.

3.2 Thermal Comfort Impact metric

In comparison to energy benchmarking, comfort benchmarking is not as commonly conducted, though thermal comfort is known to play a crucial role on occupants' health and productivity [42]. Thermal comfort is one of the parameters upon which the satisfaction of an occupant is dependent on which is typically characterized by Indoor Environment Quality (IEQ). IEQ also includes parameters that account for visual comfort, indoor air quality (IAQ), and aural comfort (relating to noise level) [43]. The consequence of IEQ being dependent on so many different parameters makes expressing it in a single metric a difficult task. Typically, an IEQ index is estimated by weighting different individual IEQ factors depending on their influence on occupant's satisfaction [27]. These weights can be estimated by recording occupant responses on how each parameter affects their satisfaction with the environment. Occupant responses give higher weightage to thermal and aural comfort contribute meaning they contribute the most to satisfaction whereas indoor air quality got the least [43]. Although the indoor air quality is the least perceivable among all the IEQ factors, a high concentration of contaminants in air is detrimental to the health of the occupant. ASHRAE recommends guidelines for buildings to follow to maintain acceptable levels of concentration of contaminants [44].

Currently, the most commonly accepted metric to measure thermal comfort level of an environment was developed by OP Fanger and is called the Predicted Mean Vote (PMV) [45]. PMV is defined as "the mean value of the thermal sensation votes of a large group of people on a sensation scale expressed from -3 to +3" [46]. It was developed from a heat balance model built by considering various factors including the metabolic rate of the occupant, the clothing level as well as the environmental conditions that the occupant is in. It is a good representation of the thermal comfort level that an occupant would feel in any environment and is quantified using a 7-point scale from -3 to 3 with -3 being cold, -2 as cool, -1 as slightly cool, 0 as neutral, +1 as slightly warm, +2 as warm and +3 as hot.

Another index that is closely related to the PMV is the Predicted Percentage of Dissatisfied people or PPD. It is formally defined as a "quantitative prediction of the percentage of thermally dissatisfied people determined from PMV". The relationship between PMV and PPD is shown in fig. 3.2. Acceptable levels of thermal environment for general comfort as defined by ASHRAE is when PPD $< 10\%$ or PMV is in the range of (-0.5,0.5) which means to say that an environment is considered to be thermally comfortable if less than 10% of the people are dissatisfied.

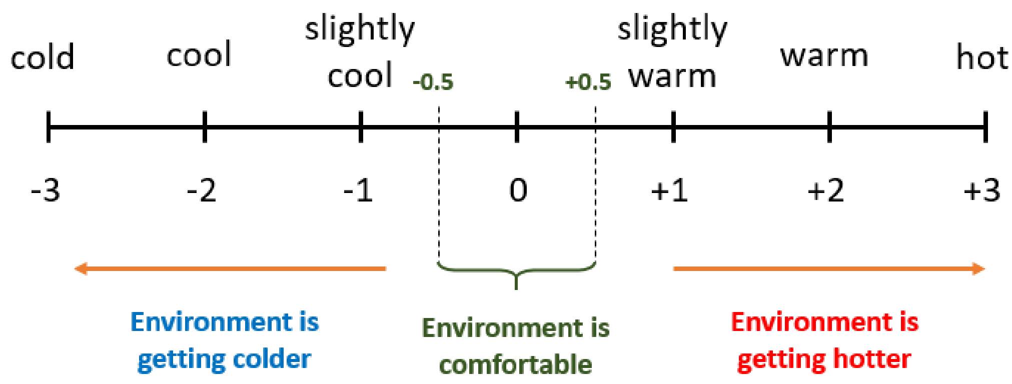


Figure 3.1: Scale showing the possible values of Predicted Mean Vote (PMV).

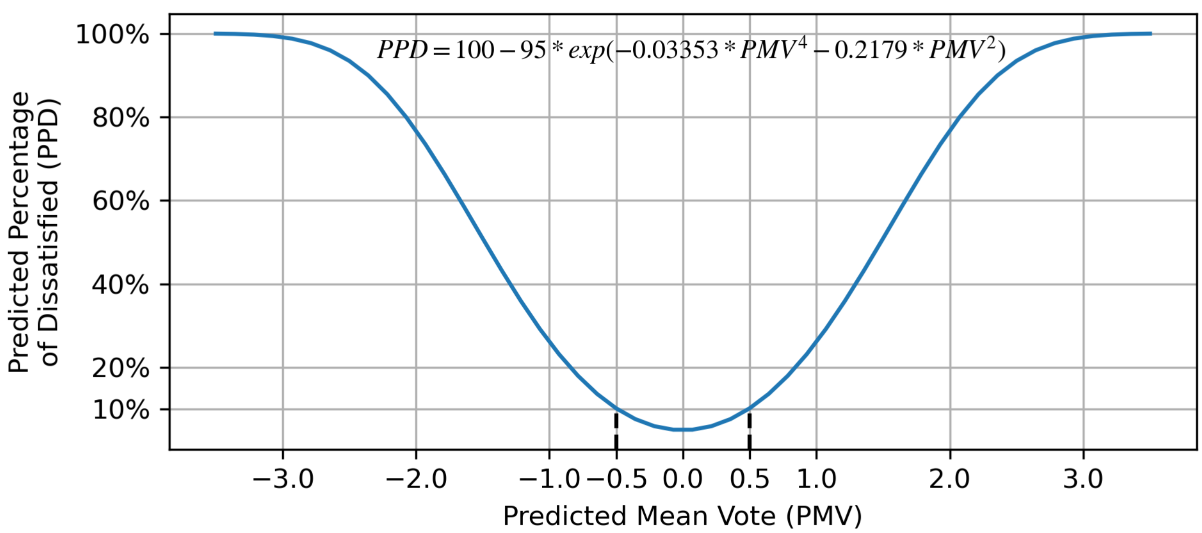


Figure 3.2: Relationship between Predicted Mean Vote (PMV) and Predicted Percentage of Dissatisfied (PPD).

Majority of the existing studies that include thermal comfort as a factor while studying smart thermostat data involve building efficient control algorithms to optimize the thermal comfort of the occupants, characterized as the PMV, by simultaneously increasing energy efficiency of the HVAC system. Only few studies in the literature analyze thermal comfort as a means to measure the severity of the faults in houses. This is because calculation of PMV requires the input of not just the temperature and relative humidity, which is typically the data available from smart thermostats, but also some data regarding occupants such as their level of clothing, level of activity and data regarding the house itself such as the mean temperature of surfaces in the house and the air velocity inside the house. Since, a major chunk of the data required for calculating PMV is not available, researchers have had to either make assumptions about the unavailable data or make comfort predictions using other metrics. An example for the former case is the study conducted by Stopps and Touchie [14] in 59 units of a Multi-unit Residential building in Toronto. The study compared the results obtained from taking surveys of thermal sensation from occupants with PMV values calculated by making some knowledgeable assumptions about the occupants and their units. While their study was more focused on examining the inconsistencies in thermal sensation reported by the occupants and those expected by the building physics, the thermal comfort model they used is a good precedent for how assumptions can be used to calculate the value of PMV.

The severity of thermal discomfort can also be calculated using the graphical method proposed in the ASHRAE standards [46]. The graphical method identifies the states of an environment on the psychrometric chart for which the PMV lies in the comfortable region thereby dividing the chart into regions of comfort and discomfort. The boundaries of the comfort region are when PMV equals 0.5 or -0.5. Since, the PMV is dependent on numerous factors, in order to reduce them to only temperature and relative humidity so that the states can be identified on the chart, the graphical method also uses assumptions regarding clothing level and activity level of occupant as well as general air speed relative to the occupant with the mean radiant temperature equal to the indoor temperature. Since, an environment can be placed in either the comfortable or uncomfortable regions the total level of thermal discomfort of an environment during a time-frame can be measured by the "unmet hours" which is the amount of time or percentage of time the environment is labelled as uncomfortable [47]. This method however does not provide a method to account of severity of discomfort i.e., how close or far is the environment from the boundary of the comfort region [27]. Additionally, total thermal discomfort level can also be measured by total degree hour of the environment away from the comfort region which is difference in temperature at the given state and the "comfortable temperature" multiplied by the duration of time of each uncomfortable period summed across all the uncomfortable periods across the time-frame [47]. However, this does not take into account the effect of other factors that influence the occupant's thermal comfort [48]. In order to account for this severity, studies have tried weighting the total unmet hours and the total degree-hours with the PMV value of the environment [47]. This weighted metric does measure the severity of thermal discomfort, but, it does not account for the amount of time spent in the comfort region. Ideally in order to measure the mean thermal discomfort level of a house, credit should be given to the system that spends more time in the comfortable region, which will be done so in the current thesis.

Another model that's used to judge thermal comfort is the adaptive model of thermal comfort put forth by Dear & Brager [49] that has since been adopted into the ASHRAE-55 [46] standards of thermal comfort. The adaptive model relates the indoor operative temperature as a linear function of the outside temperature and suggests acceptable limits of indoor temperature for optimum comfort. While this model requires less diverse data, it is also known to be more effective in naturally ventilated spaces rather than air-conditioned ones [46]. Huchuk et. al. [50] applied the adaptive model of thermal comfort on smart thermostat data from 10000 EcoBee smart thermostats enrolled in the Donate Your Data program. They found that temperature bounds as described by the adaptive model are consistent from the smart thermostat data in case of cooling seasons but not in so case of heating seasons. Furthermore, they argue that since they applied the model to setpoint temperature, it was under the assumption that setpoint temperature is a good indication of thermal comfort of occupants which may not be true because it does not capture the effects of metabolic rate, air flow and radiant temperatures. Hence, the authors of the current study deem that while the adaptive model might be a good starting place, Fanger's PMV index is a much better criteria to put a number on thermal comfort value of homes as not only does it account the factors aforementioned but also is considered more reliable in air-conditioned homes than the adaptive model.

Another example of a study that does not use the heat balance model with assumptions is the study of manual changes in setpoint done by Kane and Sharma [12] on data obtained from the Ecobee program. They use statistical relationships between the time it to make a setpoint change and the magnitude of change to make inferences about the thermal comfort of the occupants. The methodology they have chosen utilizes occupancy data which is not available in all datasets and also assumes that the setpoint temperature is a good indication of thermal comfort. Furthermore, with the wealth of information provided by the smart thermostat, there is an incentive to diversify the data used for analysis from just the time series of the setpoint temperature and occupancy data.

There exist very few studies in literature that combine the impact of energy consumption and thermal comfort into one single metric that can then be used to judge the performance of systems. Guillen et. al. [27] use the adaptive thermal comfort model to define comfort metric to identify instances of overcooling and overheating. The metric estimates the mean difference between the indoor temperature and the adaptive comfort thresholds thereby identifying systems with potential energy wastage. However, they too conclude that while energy and comfort are coupled factors identifying a clear correlation between them is challenging.

The coupling between energy and thermal comfort has not yet been fully delineated. In order to "rank" or benchmark buildings based on researchers have tried putting the two impact metrics together creating a program that allows the user to choose the weights of a combined index. Energy impact metrics such as EUI and comfort metrics such as PPD and concentration of CO_2 were combined together into a metric and the values of this metric for given system was compared with that of a baseline of the user's choice [51]. In similar vein Kong et. al. [52] established the two metrics individually and instead of combining them into a single metric chose to examine them together in the form of a matrix. This matrix essentially showcases, the rank of thermal comfort impact of an air-conditioning system vs the rank of its energy impact in a given dataset of systems. This helps identifying systems that have low or high impact in both aspects and systems that have low impact in one aspect but high in the other.

There exists a gap in literature regarding how energy and thermal comfort impact metrics are coupled. The latter study is a good example of how both of them can be examined together to identify systems with interesting behavior. An objective metric that combine both of the impacts which can be used to rank and compare systems will be useful to manufacturers and building owners. In the process of creating such an index we also wish to understand gain more insight into system behavior.

3.3 Objectives of the current study

A review of literature provides us with a necessary background to understand how fulfill the objectives of this research. The following section is an overview of the objectives that this research wishes to fulfill. The current thesis firstly proposes to use smart thermostat data to construct an energy impact metric for residential HVAC systems. Through the means of the metric the study proposes to compare the energy performance of systems at the given operating conditions. The creation of the energy impact metric requires the establishment of an appropriate baseline for comparison. Existing baselines built for smart thermostat data are specific to the performance of the given system making them ineffective for objective comparison. Additionally, they do not use the enormous scale of the data available from all the systems contributing to the dataset. Hence, an objective of this research is to build a baseline performance value that leverages the huge amounts of data and leads to a metric that can easily be used to compare to air-conditioning systems objectively. The baseline that the current research has chosen to build uses data from all available air-conditioning system to model an archetypal system of the dataset and then compares each system's performance with how the baseline system would have performed at the given conditions.

The creation of the energy impact metric, therefore, requires a methodology to model the performance of each given air-conditioning system as well as the archetypal system. After considering the limitations of using simulation software tools to build a model for the house, building a linear model was deemed easily scalable and fairly accurate (based on existing literature). However, the linear model methodologies in current literature build a model for the daily average runtime. This leads to smoothing of effects of changes in runtime during the day. In places like Texas the temperature outside can vary from $3°C$ in the morning to nearly $21°C$ in the day and back to sub $5°C$ at night. This broad range of temperatures outside can cause a broad range of runtimes during the day and taking a daily average will reduce this variation to one number. Therefore, another objective of this research is to propose a better methodology to build a model for the performance of the system that can incorporate this broad range.

Next, the creation of a thermal comfort impact metric is taken up. Since, total amount of unmet hours is an inadequate metric to compute the degree of discomfort, the current study provides way to calculate this degree using the Predicted Mean Vote of the indoor environment of the house. The estimation of PMV has to be done based on a few assumptions. An additional objective of this study, therefore, is to provide a way to calculate the impact of thermal comfort in a way that's not affected by the assumptions made. So, if in the future appropriate sensors are installed and metadata about the occupant is more accurately available, no change is needed in the calculation of the impact.

Finally, both of the above impacts need to be put together to gain an idea of their combined affect on the system. This will be done by thermal discomfort of the house to zero and estimating the effect on the value of the energy impact metric. Reduction of the thermal discomfort to zero can be done by making the house comfortable at all points in time, which will give rise to another metric which is the amount of extra cooling hours required to make a house completely comfortable. This along with the change in the value of energy impact metric can then be used to characterize the behavior of various air-conditioning systems and examine the coupled nature of energy and comfort.

3.4 Data used for the current analysis

Based on data selection factors outlined in 2.2 the following study uses data collected between May and September of 2019 from 7352 HVAC systems in the state of Florida (IECC Climate zone 1A) that have a single-speed outdoor cooling unit. In the analysis done during the current study, the event-based data from **regulating periods** of each system was gathered and re-sampled into periods of 2-hours. This means that a 2-hour weighted average of each of the operation parameters was calculated and then the new time-series so formed analyzed. Since, system is in regulating mode or pseudo steady-state mode, the indoor temperature of the house is assumed to remain a constant during the duration of the 2 hours and equal in magnitude to the weighted average of the indoor temperature. Similarly, the rest of the operation parameters are also re-sampled and assumed to be constant during each 2-hour period before beginning the analysis.

4. CONSTRUCTION OF AN ENERGY IMPACT METRIC

As outlined in the objectives section of chapter 3, the creation of a metric to estimate the impact of a fault on the energy consumption of an air-conditioning system requires the construction of a baseline for comparison and methodology to model the energy consumption of any given system. In order to do so the following chapter uses data from 7000 single-speed cooling systems from the same climate region collected during the summer months of the year 2019.

The large-scale nature of the data available provides an opportunity to study several systems at the same time. This results in estimating the energy consumption level of the average system in the dataset. The average air-conditioning system in the dataset represents how much energy on average is consumed by air-conditioning systems in the dataset. Since it will be built using data from all the systems, it is good objective baseline that can be used to compare the energy consumption of any given system. The value of the energy impact metric of an air-conditioning system, therefore, describes how much extra or less energy is consumed by the system to run in the house it's installed in, in comparison to the amount of energy consumed by systems in the dataset on average.

4.1 Rationale behind using the average system as the baseline

Ideally, each AC's energy consumption should be compared to check if it's greater or less than acceptable levels i.e., the energy consumed by a properly sized, fault-free air-conditioning system. However, defining what is acceptable performance level is difficult. As noted in the background section, researchers estimate the acceptable level of consumption by running simulations on tools like Energy Plus or BEopt. However, these simulations are computationally expensive and inapplicable for the systems used in the current study. This is because the smart thermostat data does not contain any metadata regarding the layout, orientation or size of the house. In the absence of such metadata conducting simulation to gauge the acceptable performance level will be inaccurate. Instead, the following study chooses to estimate the performance of the average system to use as the baseline for comparison. Estimating the performance of the average system is data-driven, based on the data from each available system. It is not as expensive to compute as an energy simulation model and essentially represents how much energy the average system in the climate region would require, to create the exact same environment being produced by the given system. Alternatively, it could also be thought of as the amount of energy that the given system would take to produce the same environment in a house with the mean characteristics as those of houses found in the dataset.

The performance of the average system, however, need not be a good representation of acceptable system performance levels. More often than not the average system performs below acceptable levels, so when comparing systems care should be taken. A system that performs better than the average system may not necessarily be good because it may still run at lower than acceptable levels. However, a given air-conditioning system that performs worse than the average system will surely perform at lower than acceptable levels.

When the average air-conditioning system performs lower than acceptable levels i.e., consumes more energy than the what is considered acceptable, it generally implies that a majority of the air-conditioning systems in the dataset perform a lower than acceptable levels. Equal number of systems will consume more and less energy than the average system, but given that the average system itself consumes more energy than a properly sized and perfectly operating system, majority of the systems will also work equally worse. The current study notes that the baseline used maybe inaccurate to judge the relative performance of a given air-conditioning system, however, in the absence of true/acceptable energy consumption levels, it still works as a good proxy in identifying systems that are definitely running at lower than acceptable levels.

In order to design a way to find energy consumed by air-conditioning systems in the dataset on average, first, a method must be constructed to find the model the energy consumption of a single system. In order to build a model for a single system, a typical method used is to regress the average amount of cooling energy consumed by the system in a day against the average value of the difference in indoor and outdoor temperature. However, in an effort to capture the fluctuations in runtime throughout the day, a different design for building an energy consumption model of the system is proposed.

When the air-conditioning system is maintaining a given setpoint then the system is considered to be operating in a pseudo steady state mode. In this mode, the amount of heat added to the house is equal to the amount of heat removed from the house by the cooling system $(\dot{Q}_{cool})$, because the temperature of the house remains a constant. The significant sources of heat addition to the house are: heat conducting through the walls of the house from outside $(\dot{Q}_{cond})$, heat addition due to solar irradiance $(\dot{Q}_{solar})$ and internal load of the house due to occupants' activity $(\dot{Q}_{int})$. Therefore, the equation of the house operating at a steady indoor temperature will look as follows,

$$\dot{Q}_{cool} = \dot{Q}_{cond} + \dot{Q}_{solar} + \dot{Q}_{int} \tag{4.1}$$

Heat is also added to the house because of infiltration of outside air and heat is removed from the house to dehumidify the house but these are much smaller than the sources considered in the above equation and hence are ignored. The amount of cooling provided by the air-conditioning system is dependent on the capacity of the system (C_{sys}), the duty factor of the system, which is the percentage of time the system is switched ON, and the load factor which is percentage of load level at which the system is operating. The product of duty factor and load factor is here on referred to as the cooling effort (E_c) of the air-conditioning system. The load factor of single-speed systems like the ones used in the study is always 1 or 0. Hence, the cooling effort of a single speed air-conditioning system is just the duty factor of system. The conduction load is dependent on the overall heat transfer coefficient of the home (U), the overall area of heat transfer across the walls and the roof (A), and the difference between outdoor and indoor temperature (ΔT_{oi}). The heat balance equation in eqn. 4.1 is now transformed into:

$$C_{sys}E_c = UA\Delta T_{oi} + \dot{Q}_{solar} + \dot{Q}_{int} \tag{4.2}$$

The cooling effort of the air-conditioning system is therefore affected by the capacity of the system installed, the size of the house, the difference in indoor and outdoor temperature, the amount of solar load the house gains and the total amount of activity done by the occupants of the house.

However, a reasonable assumption is that a larger house with large windows that allow for easy penetration of solar load or a house with higher number of occupants will be fitted with an air-conditioning system of higher capacity. Similarly, across the dataset the systems are assumed to have been sized such that the ratio of heat conductivity of wall to system capacity is nearly the same. The increase in internal load due to the residents having a party, or because of opening of a window on a particular day are assumed to not have a large affect because the analysis is done on the scale of months. Therefore, among the variables in the above equation, only the temperature difference affects the cooling effort significantly as they appear to be highly correlated.

The following section presents a methodology to describe this relationship by building a linear regression model between the cooling effort and the values of temperature difference. The energy impact metric of a system can be computed by comparing the cooling effort of the given air-conditioning system with the cooling effort of the average system at the values of temperature difference at which the system operates.

4.2 Computing the relationship between cooling effort (E_c) and temperature difference (ΔT_{oi})

Firstly, the event-based data from the thermostat is resampled into periods of 0.5 hour after which a rolling two-hour average is taken. Only periods when the air-conditioning system maintains a constant temperature i.e., the system operates in a pseudo-steady state mode are considered. In each of the periods the mean cooling effort (E_c) is the percentage of time the AC is switched on in the 2-hour period and the mean difference in indoor and outdoor temperature (ΔT_{oi}) calculated as shown in eqn. 4.3,

$$\Delta T_{oi} = T_o - T_i \tag{4.3}$$

As discussed in eqn. 4.2 the cooling effort of the system is directly proportional to the temperature difference of that 2-hour period.

$$E_c \propto \Delta T_{oi} \tag{4.4}$$

In order to **visualize** the method that will be described, first a probability density function of the joint distribution of cooling effort and temperature difference from each of the 2-hour periods from all the air-conditioning systems belonging to the dataset was constructed. This probability density function was constructed using a non-parametric method (free of assumptions regarding distribution of original data) called as Kernel Density Estimation (KDE). The KDE first chose 20 random data points from each system and built a 2D gaussian kernel of cooling effort and temperature difference at each data point, upon summing of which results in the joint probability density function. This process was repeated multiple times and the densities at each point were averaged across all the samples. KDE used here did not use all the points from all the systems at once because that might lead to overfitting due to excess of data. Instead repeated random samples are taken and densities obtained were averaged. The KDE built is shown in fig. 4.1 where the darker regions of the plot indicate a higher probability of finding datapoint and hence indicate regions of higher concentration of datapoints than the lighter regions. The above distribution can

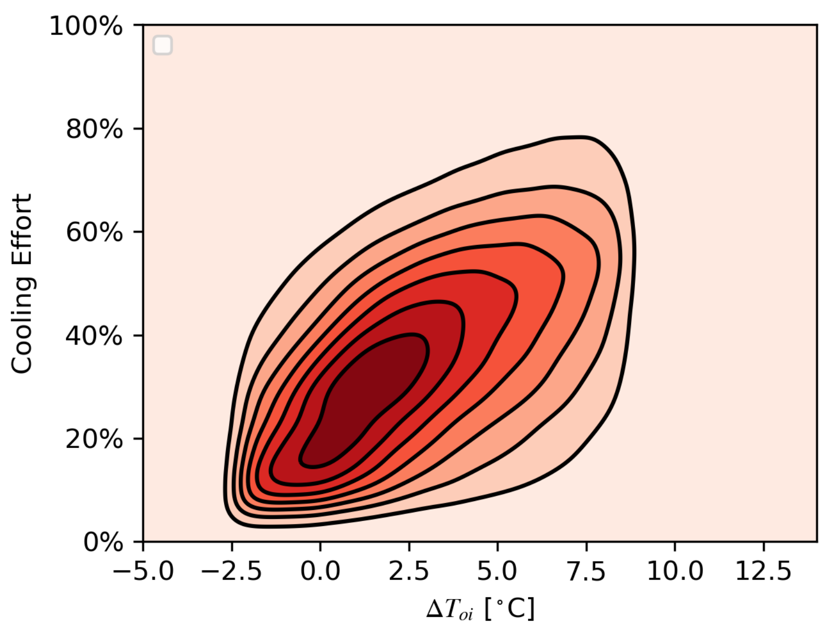


Figure 4.1: Joint probability density function of cooling effort and temperature difference made by using data from all the systems in the dataset containing approximately 7,000 air-conditioning systems in Florida.

be used to visualize the method to estimate the relationship between cooling effort and temperature difference for all systems in the dataset on average which represents the performance of an average system in the dataset. Based on the physics of heat flow into and out of the house, in the equation 4.2 the cooling effort of a house is a linear function of temperature difference between outdoors and indoors of the house. A simple linear regression model can therefore be built to describe this relationship with cooling effort as the dependent variable and temperature difference as the independent variable. Such a model when built using data from all the systems in the dataset will describe how the cooling effort of air-conditioning systems in the dataset is influenced by temperature difference on average. Since the data used to construct such a model will come from 2-hour periods of operation of the system, the model so constructed will describe the cooling effort of the average system in a given 2-hour period against the value of temperature difference in that period. Let the slope of this linear model be denoted by β_1^M and the intercept by β_0^M, the cooling effort of the average system in 2-hour period is given by E_M is,

$$E_M = \beta_0^M + \beta_1^M \Delta T_{oi} \tag{4.5}$$

While constructing such a model from the given data care was taken to avoid over-fitting. This was done by first taking a random sample of 10 periods from each system; using which the coefficients of the model in 4.5 were calculated using the Ordinary Least Squares (OLS) regression method. This process was then repeated multiple times and an average of the coefficients estimated in each case was calculated. The model is shown against the contours of the probability density function in 4.2 Instead of using data from all air-conditioning systems, coefficients of the model for each system can be estimated by using the OLS regression method. The distribution of cooling effort and temperature difference values for a single system was plotted in fig. 4.3. The linear model for the system was built by regressing the cooling effort against the ΔT_{oi} values, like how it was done for data from all the systems in the dataset.

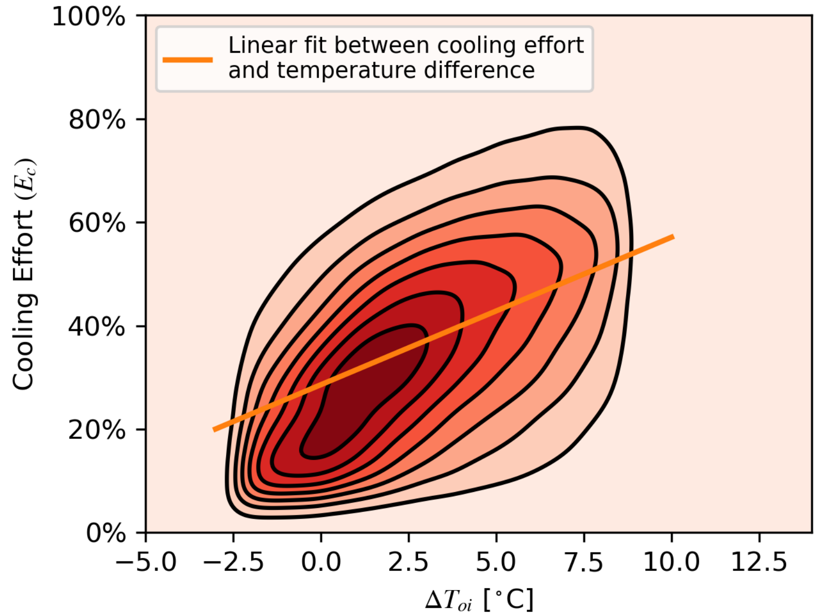


Figure 4.2: Linear model of the energy consumed by the average air-conditioning system in the dataset shown on the contour plot of the joint probability density function built using data from all the systems.

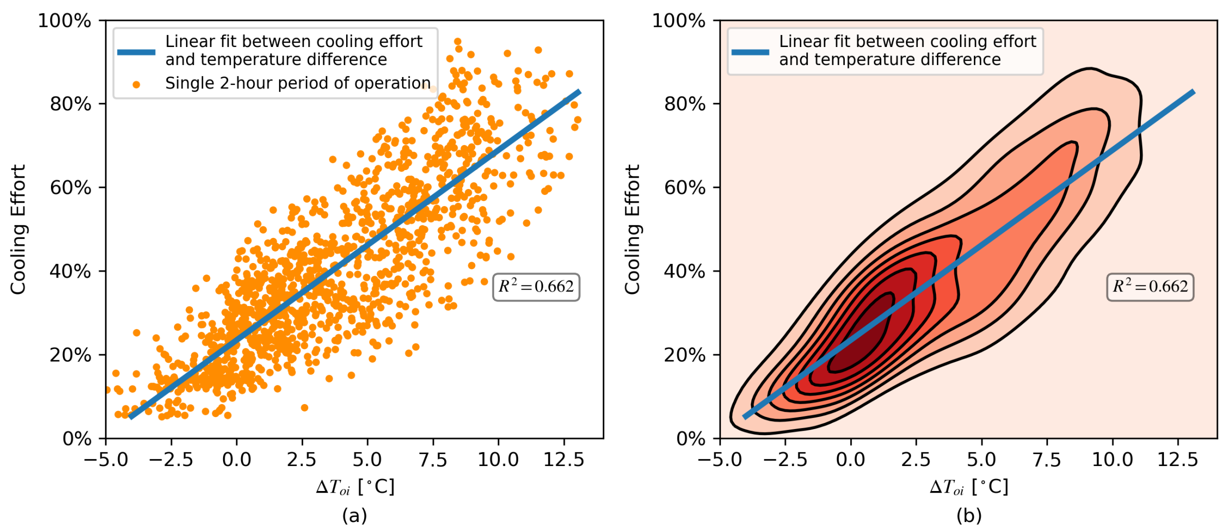


Figure 4.3: Linear model built for the energy consumed by a given air-conditioning system. The subplots are (a) scatter plot of 2-hour data from the system (b) contour plot of the joint probability density function of the same data.

Let β_0^S and β_1^S represent the intercept and slope of the linear model of the system in the house its installed in, then the cooling effort (E_S) of the system is given by,

$$E_S = \beta_0^S + \beta_1^S \Delta T_{oi} \tag{4.6}$$

4.3 The reliability of the above method

In order to test the reliability of the method described above, examining how good of a fit is the model to the data is necessary. In order to examine the goodness-of-fit of the models in equations 4.5 & 4.6, the R^2 value of the model was calculated. The R^2 value of a linear model describes what percentage of the total variability in the data can be explained by the model. The higher the R^2 the better is the fit of the model to the data. The R^2 value can take a maximum value of 1 when the model fit passes through each and every single point in the data i.e., the model fully explains all the variability in the model.

The model of the average system as calculated using the method explained in the preceding section, for the data used in the current study, was, $E_M = 0.2892 + 0.0277\Delta T_{oi}$. Like the coefficients, an R^2 value for each estimation of the model using a random sample of the data was calculated and a mean of the R^2 values was estimated to be equal to 0.262. This implies that the the model of the average system explains 26.2% of the total variability in the data. Similarly, an R^2 value was also calculated for the energy consumption models built for each system in the dataset and a histogram plot of these values is given in fig. 4.4

The "low" R^2 value of the model of the average system and the general variation of R^2 values over a range raises the question of the reliability of the models built. If the models are not able to adequately explain the variability in the data, there is a strong evidence to indicate that perhaps the cooling effort is not fully explained by the temperature difference values itself and there are other variables that need to be included in its model. Based on the heat balance model in equation 4.2, on top of the temperature difference between inside and outside, the cooling effort of the air-conditioning system is also dependent on other loads such as the internal load produced by

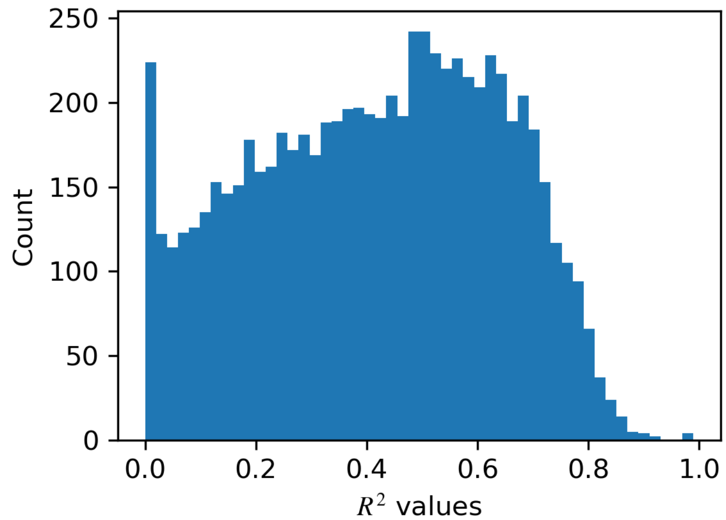


Figure 4.4: Histogram plot of R^2 values obtained for the energy consumption models built for each air-conditioning system of the dataset.

the occupants and the solar load on the house. However, using smart thermostat data, delineating the effect each of these loads have on the cooling effort is not possible because of the lack of requisite data and the uncertainty involved. For example, calculation of solar load will depend on the orientation and layout of the house, the size of the windows and the roof as well as the heat energy received from the sun at that location. While the latter value can be calculated using the zip code of the house, the former require metadata which is unavailable. Furthermore, the amount of internal load too is dependent on the number of occupants in the house and their behavior, regarding which data is not available. Therefore, the model built above was only built with the purpose of using the available data to only capture the effect of temperature difference on the cooling effort of the system. Since cooling effort is also dependent on other factors on top of cooling effort, the constructed model can only explain a part of its total variability. In order to conduct further regression diagnostics, the distribution of residuals of the models were examined. The histogram plot of the distribution of one system is given in fig. 4.5 and shows that the residuals are normally distributed with the Shapiro-Wilk's test for normality having a p-value less than 5% implying that there is significant evidence for the same.

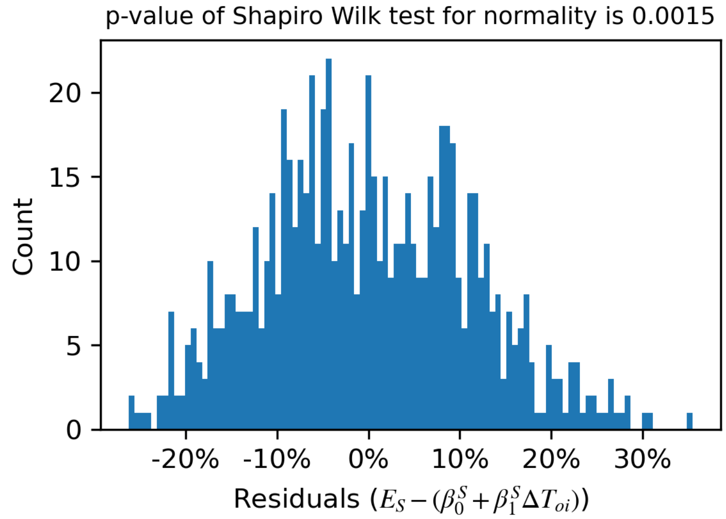


Figure 4.5: Histogram plot of residuals of the model of cooling effort of a system of the dataset.

Predicting occupant behavior in residential buildings is a difficult task. Often high cooling effort in a given 2-hour period could be due to a number of reasons including: occupant left the window open, occupant is having a party, obstruction in the air vents, leakage of refrigerant etc. Since diagnosing the exact cause without further data is not possible, the current study asserts that the model built will be inadequate. However, it can still be used to examine how a unit change in temperature difference affects the cooling effort assuming that the all other parameters are at their "mean" operation level. So care should be taken to note that the R^2 value of the model is misleading, because though the model may be incomplete, for the purposes of this study, it is still valid.

4.4 Interpreting the models

The model for the average system was built by directly regressing the cooling effort values in each 2-hour period against the values of temperature difference. This however, is not the only way to find a model for the average system. Other methods include taking the median/mean or taking the midpoint of the density field at each ΔT_{oi} value. The latter method is explored in greater detail in Guo and Rasmussen [2]. The advantage of the model used here is that, the variables affecting the cooling effort of the air-conditioning system can be split into easily interpretable groups. Since, this is a linear model of the form $E_c = \beta_0 + \beta_1 \Delta T_{oi}$, by comparing it to the heat balance equation in eqn. 4.2, the following analogy can be drawn:

$$\beta_0 \sim \left(\frac{\dot{Q}_{int} + \dot{Q}_{solar}}{C_{sys}} \right) \tag{4.7}$$

$$\beta_1 \sim \left(\frac{UA}{C_{sys}} \right) \tag{4.8}$$

The intercept of energy consumption model of the system could be a representation of the mean value of the total of internal and solar loads on the house per unit capacity of the system. Similarly, the slope could be an indication of the mean insulation level and/or infiltration level of the house normalized by the capacity of the system. Based on this analogy, the following comparisons can be drawn:

1. If the slope of the model of the given air-conditioning system is greater than that of the average system i.e., $\beta_1^S > \beta_1^M$ then there exists a strong likelihood that the insulation level of the house is worse than the mean insulation level found in houses in the dataset, because for the same amount of change in ΔT_{oi}, the given house sees a larger change in cooling effort than the system with mean insulation and vice-versa. The reader must not that with the available data proving the above statement is not possible as it would require metadata regarding the size and layout of the house.

2. Similarly, if the intercept of the model of the given system is greater than that of the average system i.e., $\beta_0^S > \beta_0^M$ then there is a strong likelihood that the (internal+solar) load on the house is more than the (internal+solar) load found on houses of the dataset on average.

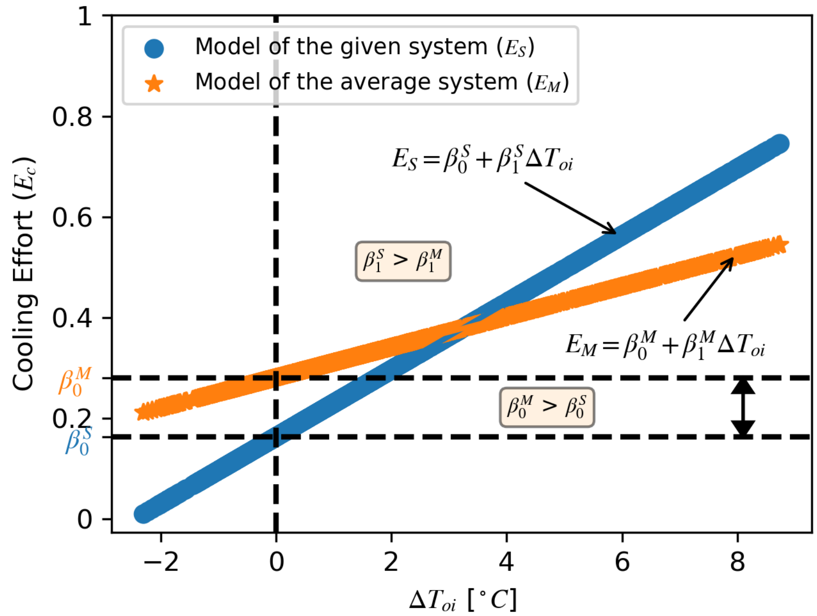


Figure 4.6: Illustration of the energy consumption model of an individual air-conditioning system against that of the average air-conditioning system of the dataset. Based on the analogies established, there is a strong likelihood that the house has poorer insulation but lower internal load than what is found on average for houses in the dataset.

4.5 Calculating an Energy Impact metric

Since a baseline as well as a procedure to model the energy consumption of a given air-conditioning system has been established, the value of the energy impact metric can be estimated. The energy impact metric is defined as the percentage of time, less, the average system would run that the given system in order to produce an environment at the current level of comfort. The energy impact is an estimation of how much impact a fault in the house is causing to change the energy consumed by a given air-conditioning system. This value represented as Γ_E can be calculated as,

$$\Gamma_E = \frac{\sum_{i=1}^{N} ({E^i}_S - {E^i}_M)}{\sum_{i=1}^{N} E_S^i} \tag{4.9}$$

The $'i'$ in the above equation represents each individual 2-hour period for the system and summation is done over the entire operating time frame of the system consisting of N periods. In each of the periods, the temperature difference value (ΔT_{oi}^{i}) is used to estimate the cooling effort of the given air-conditioning system which is then compared with the cooling effort of the average system.

$$\Gamma_E = \frac{\sum_{i=1}^{N}\left(\left(\beta_0^S + \beta_1^S \Delta T_{oi}^{i}\right) - \left(\beta_0^M + \beta_1^M \Delta T_{oi}^{i}\right)\right)}{\sum_{i=1}^{N}\left(\beta_0^S + \beta_1^S \Delta T_{oi}^{i}\right)} \tag{4.10}$$

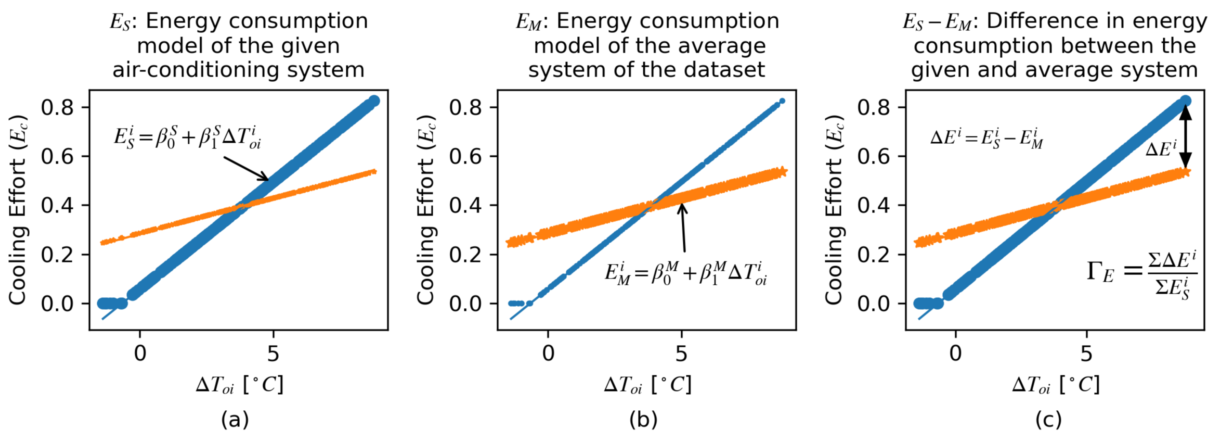


Figure 4.7: Illustration of the method used to calculate the energy impact of a system

Figure 4.7 shows a visual representation of the calculation of energy impact at each value of ΔT_{oi}. The energy consumed by the given and the average system were both calculated from their respective models. Therefore, the value of the energy impact of a system is dependent on the difference between the slope and intercept of the energy consumption model of the given system and the energy consumption model of the average system as well as the range of operation of ΔT_{oi} values.

Calculation procedure for the **Energy Impact Metric** (Γ_E)

1. Modelling the energy consumption of the given system

$$E_S^i = \beta_0^S + \beta_1^S \Delta T_{oi}^i \tag{4.11}$$

2. Modelling the average energy consumption of the systems in the dataset

$$E_M^i = \beta_0^M + \beta_1^M \Delta T_{oi}^i \tag{4.12}$$

3. Γ_E - Estimating the relative difference in total energy consumption of the given system and the average system

$$\Gamma_E = \frac{\sum_{i=1}^{N} ({E^i}_S - {E^i}_M)}{\sum_{i=1}^{N} E_S^i} \tag{4.13}$$

5. CONSTRUCTION OF A THERMAL COMFORT IMPACT METRIC

Along with the energy consumed, faults in air-conditioning system also have a detrimental affect on the thermal comfort of occupants. However, using smart thermostat data to judge the comfort level of homes is a challenging task because of the lack of data. Thermal comfort of occupants is dependent on a myriad of parameters that are often difficult to estimate especially for analyses done on such a large scale. As described in the chapter 3 the thermal comfort level of air-conditioned environments is typically estimated using a metric called as the Predicted Mean Vote (PMV) [46].

PMV of an indoor environment is dependent on six factors: air temperature (T), relative humidity (ϕ), activity level of the occupant (met), the clothing level of the occupant (I_{cl}), average air speed (V_a) around the occupant and finally the mean radiant temperature $(\bar{T_r})$ [46]. Among these temperature and relative humidity are standard parameters, however, the rest are not something that is typically measured in environments. The activity level of the occupant translates to the metabolic rate of the occupant which is the rate at which chemical energy is converted into heat and mechanical work by metabolic activities of an individual per unit area of skin. The amount of clothing that the occupant wears represents the amount of resistance to sensible heat transfer that the occupant carries. The air speed is the relative air speed around the occupant. The mean radiant temperature is the temperature of a uniform, black enclosure that exchanges the same amount of heat by radiation with the occupant as the actual surroundings. This temperature is typically measured using a Globe Thermometer.

$$PMV = f\left(T, \phi, met, I_{cl}, V_a, \bar{T_r}\right) \tag{5.1}$$

5.1 Assumptions required to estimate the PMV

Since obtaining parameters not collected by the smart thermostat data i.e., metabolic rate, activity level, average air speed and mean radiant temperature is not a viable task that could be done here, a few realistic assumptions will be made. The most influential variables during the estimation of PMV are the metabolic rate, air temperature and relative humidity [53]. While air temperature and humidity are available from smart thermostat data, metabolic rate of the occupants has to be assumed. Along with the metabolic rate, the clothing level of the occupants, the air velocity inside the home and the mean radiant temperature of the house have to be assumed as well. As pointed out in chapter 3 there are studies in existing literature showing examples of researchers using reasonable assumptions to estimate the PMV. Lou et al [18] has listed a short review of examples of four studies using assumptions to estimate PMV. Based on existing literature the current study uses the following assumptions to estimate PMV:

1. The metabolic rate of the occupants is measured in units of met wherein $1\ met$ equals $58.2\ W/m^2$. $1 met$ represents the amount of energy generated per unit area of skin by an occupant just sitting at rest [46]. Assuming that the average residential home has 4 occupants and they are performing the following duties each: typing, cooking, sleeping and walking about. The average of the metabolic rate of these activities is $1.325\ met$ which was what was assumed as the metabolic rate of the occupants of the houses.

2. The clothing level is expressed in units of clo which represents insulation provided by the clothing worn by the occupants. $1\ clo$ represents an insulation of $0.155\ m^2C/W$ which typically amounts to wearing a trouser, a t-shirt and a long sleeve sweater [46]. The current study assumes a clothing level of $0.5\ clo$, the mean level for summer months, as suggested by Lou et. al. [18].

3. The relative air velocity inside the residential buildings was assumed to be $0.1\ m/s$ as ASHRAE standards limit the application of this method to indoor air velocity less than $0.2\ m/s$.

4. The mean radiant temperature in the current study was assumed to be equal to air temperature. Lou et. al. argues that this assumption is cause for inaccuracies in estimation of PMV and is also known to not be effective in homes with poor insulation or in homes with significant exposure to the sun [54]. They outline a methodology that could be used to better estimate the MRT but it requires the knowledge of the insulation level of the homes (the resistance value of the walls) and the area of the walls. Since, the current dataset accumulates data from a large number of diverse residential buildings and lacks metadata regarding building type, application of this method is not viable here.

The indoor temperature and relative humidity were available from the smart thermostat data. The assumptions in the study here may not be fully accurate, but the authors of the study assert that the method proposed to calculate an impact metric is still applicable. With more ubiquitous use of smart thermostats and enabling of instant occupant survey and feedback, getting accurate values of PMV is possible but the metric outlined below will still be applicable. Furthermore the above method in essence is similar to the graphical method of assessing thermal comfort proposed by ASHRAE in [46] where in assumptions are chosen for the same four parameters and a comfort region is constructed on the psychrometric chart. The method used in this thesis also does the same by constructing a comfort region of temperature and relative humidity values but instead of time spent outside the comfort region, the predicted mean of thermal sensation votes is used as the metric that is also able to measure the severity of discomfort.

Tartarini and Schiavon [55] have developed a python package called *pythermalcomfort* that was used in the following study to calculate the PMV value at different instants of time. The pmv_ppd function in the package does not take arrays as inputs only accepting scalars instead and so it was modified accordingly in order to reduce computational time. Using the modified version of the package along with the assumptions a metric to gauge the impact faults have on the occupant's thermal comfort was constructed.

5.2 Construction of a Thermal Comfort Impact metric

In order to build a metric first a baseline for comparison must be chosen. In the current study the thermal environment of the house will be compared to the worst possible conditions that the occupant could live in to show the influence of air-conditioning system. In order to construct an easily interpretable metric instead of choosing an extreme case, such as the arctic, as the baseline, the baseline for comparison was chosen to be the outdoor environment that the house is located in. The idea employed is to check, on average how uncomfortable the surroundings in the house are as compared to how uncomfortable the occupant would be if he were living in fully naturally ventilated conditions at the same location as that of the house. Living in fully naturally ventilated conditions would be the same as camping at the same location the house is in. Since the air-conditioning system is installed to mitigate the discomfort caused due to the natural conditions, comparing its impact on the occupants with respect to the outdoor conditions would be a good reference point to compare various systems.

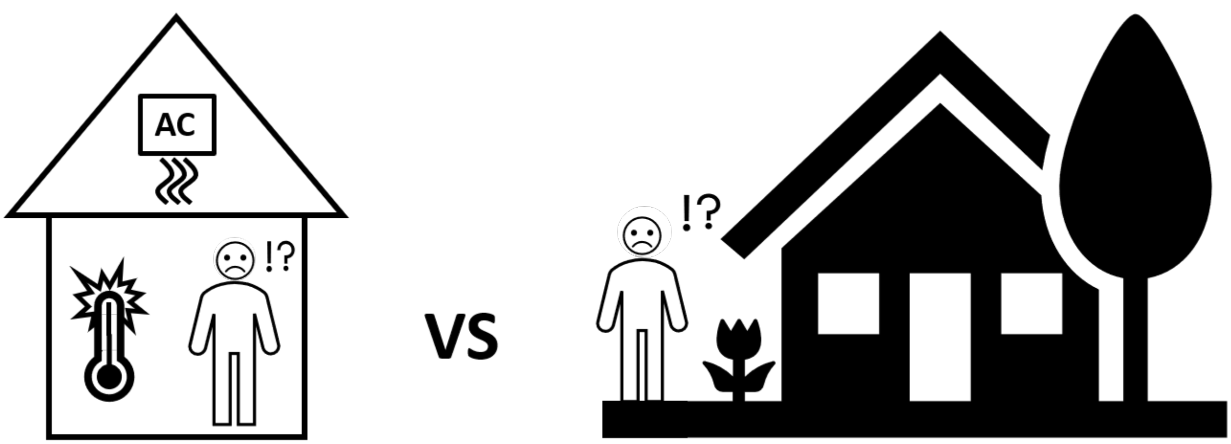


Figure 5.1: Comparing the mean level of discomfort felt by the occupant living in the air-conditioned environment to the mean level of discomfort the occupant would have felt if they were living in the same outdoor environment as the house is located in.

In order to use PMV to estimate discomfort level of an environment, first, the event-based data is resampled into periods of two hours like in the case of calculation of energy impact. The resampling of data ensures that the occupants have enough time to acclimatize themselves to the surroundings of the indoor environment ensuring that it is steady.

An environment is considered to be comfortable when the PMV is between -0.5 and 0.5 [46]. So, for each of the resampled periods of the system after calculating the PMV, the measure of discomfort that the occupants will be experiencing because of the PMV being outside the comfortable range, can be calculated. This measure of discomfort or degree of discomfort is difference between the PMV value of that period to the boundary of the comfort region. That is if the PMV were more than 0.5 than the degree of discomfort can be calculated by subtracting 0.5 from the PMV. If the PMV were less than -0.5 then the degree of discomfort can be calculated by adding 0.5 to the value PMV. However, if the PMV were in between (-0.5,0.5) then the degree of discomfort would be 0.

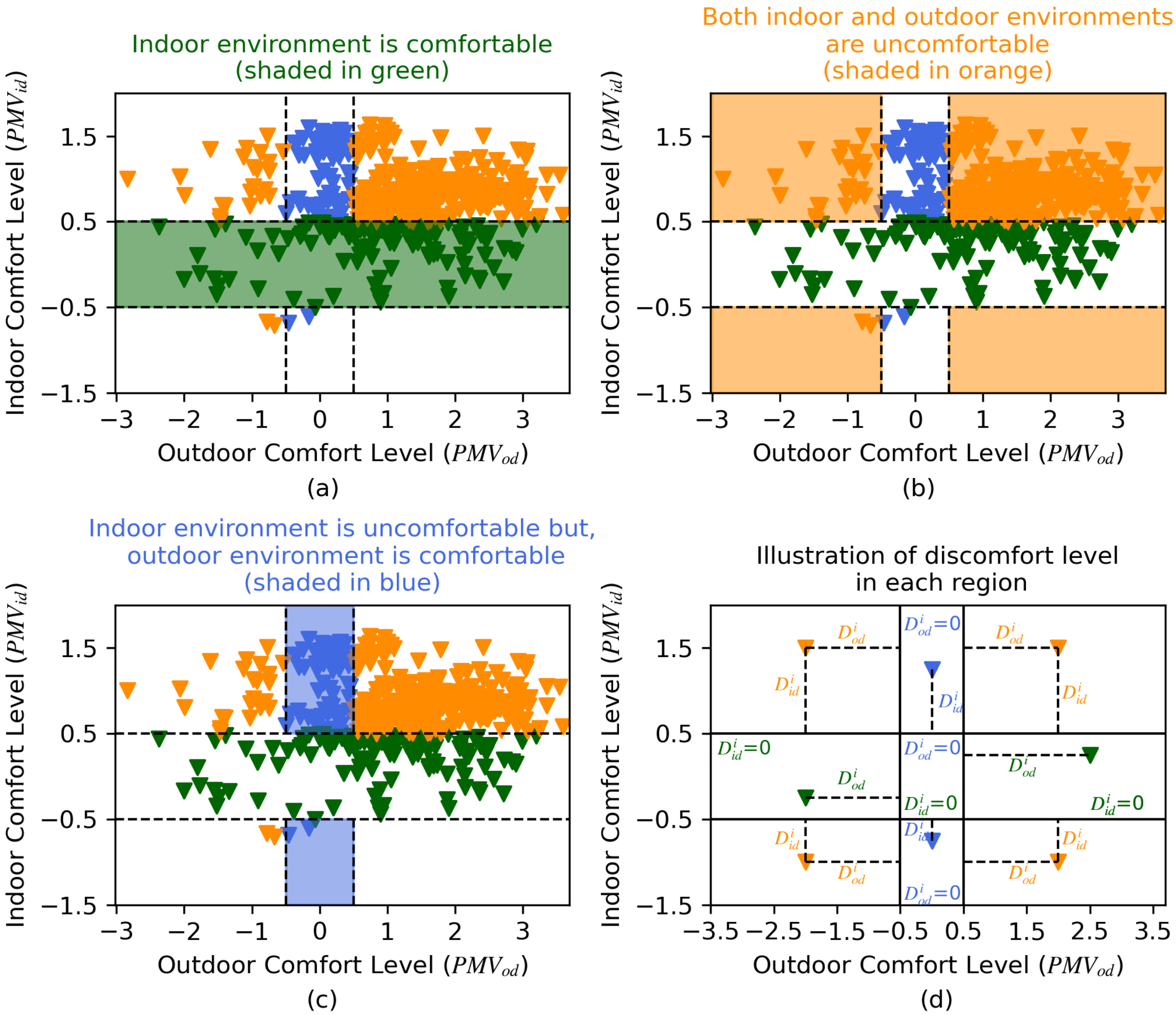


Figure 5.2: Illustration of the procedure used to calculate the degree of discomfort in each of the possible cases of the indoor environment. (a) Indoor is comfortable (b) Both indoor and outdoor are uncomfortable (c) Outdoor is comfortable but indoor is uncomfortable (d) Discomfort is each region.

In fig. 5.2, PMV_{id} represents the predicted comfort level of the home as calculated using the indoor temperature and indoor relative humidity along with the assumption for the rest of the predictors. PMV_{od} represents the predicted comfort level of the home as calculated using the outdoor temperature and outdoor relative humidity along with the same assumptions as those used to calculate PMV_{id}. Indoor discomfort (D^i_{id}) and outdoor discomfort (D^i_{od}) are the values of degree of discomfort as calculated for each 2-hour period of the system, $'i'$, indoors and outdoors respectively.

The degree of discomfort in each region is measured as the distance of the period's PMV to the boundary of the comfort region. So, in case of 5.2(a) the indoor environment is comfortable and so the indoor discomfort (D^i_{id}) for the periods lying in this region is 0. The outdoor discomfort for the region in green is magnitude of the distance between the outdoor comfort level (PMV_{od}) and the boundary of the comfort region given by $PMV_{od} = -0.5$ for all periods where outdoor comfort level is less than -0.5 and $PMV_{od} = 0.5$ for all periods where outdoor comfort level is greater than 0.5.

In the case of points lying in orange region as shown in 5.2(b), both indoor and outdoor environments are uncomfortable. That implies that both of them experience a positive degree of discomfort. Points with indoor and outdoor PMV values greater than 0.5 have $PMV_{id} = 0.5$ & $PMV_{od} = 0.5$ as the boundary of the comfort region. Similarly, with indoor and outdoor PMV values less than -0.5 have $PMV_{id} = -0.5$ & $PMV_{od} = -0.5$ as the boundary of the comfort region. The degrees of discomfort in each case is just the magnitude of the distance to the boundary of the comfort region.

Finally, in case of points lying the region in blue as shown in 5.2(c), the outdoor environment is comfortable but the indoor environment is uncomfortable. Therefore, the outdoor discomfort (D^i_{od}) is equal to 0 whereas the indoor discomfort is the magnitude of difference between the PMV value and boundary of the comfort region. In case of indoor PMV (PMV_{id}) values being greater than 0.5, the comfort boundary is $PMV_{id} = 0.5$ and for indoor PMV values being less than -0.5, the comfort boundary is $PMV_{id} = -0.5$. Thus, based on the region where each the comfort levels of each 2-hour period lies, the degree of discomfort for each 2-hour period can be calculated.

Upon plotting the indoor discomfort of each two hour period against the outdoor discomfort the plot in fig. 5.3 is obtained,

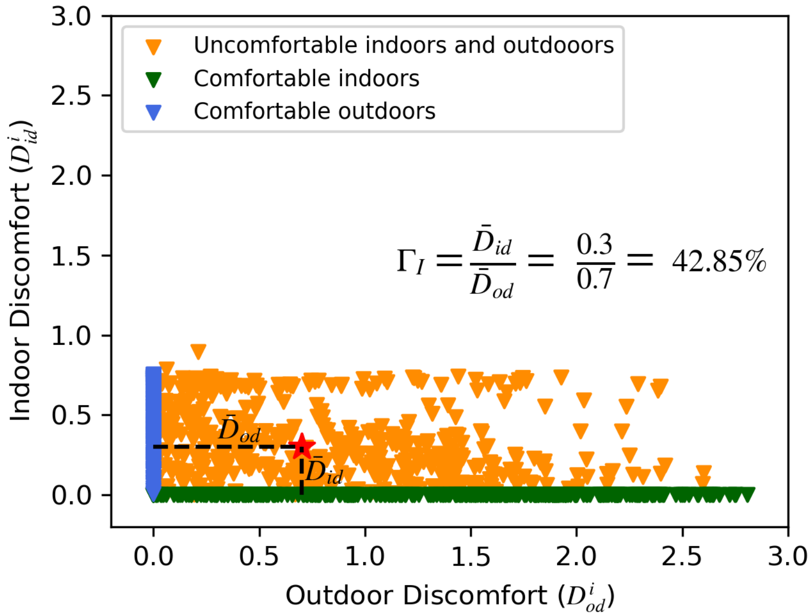


Figure 5.3: Visual representation of the thermal comfort impact. The points are in the same color notation as the one used to define the regions in fig. 5.2

Here the mean indoor discomfort $(\bar{D}_{id})$ and the mean outdoor discomfort $(\bar{D}_{od})$ are calculated as,

$$\bar{D}_{id} = \frac{1}{N}\sum_{i=1}^{N}\left(\lambda^i D^i_{id}\right) \tag{5.2}$$

$$\bar{D}_{od} = \frac{1}{N}\sum_{i=1}^{N}\left(\lambda^i D^i_{od}\right) \tag{5.3}$$

where the summation is done over all the N 2-hour periods of the system. λ^i in the above equation is an indicator variable to signify the occupancy of the house. If the house is occupied then the value of λ^i is equal to 1 and if the house is unoccupied then the value of λ^i is 0. This ensures that the mean discomfort is only calculated for periods wherein the house is occupied, because the thermal comfort only has an impact on the occupant when the house is occupied. The thermal comfort impact metric (Γ_I) is defined as:

$$\Gamma_I = \frac{\bar{D}_{id}}{\bar{D}_{od}} \tag{5.4}$$

While calculating mean indoor discomfort ($\bar{D}_{id}$), and mean outdoor discomfort ($\bar{D}_{od}$), some of the periods may have the indoor PMV (PMV_{id}) or outdoor PMV (PMV_{od}) in between (-0.5, 0.5) in which case indoor discomfort (D^i_{id}) or outdoor discomfort (D^i_{od}) in that period is equal to 0. These periods are not removed from the calculation of mean discomfort because merit should be given to those systems that have a lot of zeros i.e., systems with a lot of comfortable periods. This means that the systems with many comfortable periods will have lower mean indoor discomfort ($\bar{D}_{id}$), implying that on average they are more comfortable systems. Mean outdoor discomfort ($\bar{D}_{od}$) will generally not be equal to 0 because it only means that the outdoor is extremely comfortable and the AC system isn't required. Among the systems analyzed not a single system with mean outdoor discomfort ($\bar{D}_{od}$) equal to 0 was found.

5.3 Solving occupancy issues

One of the main problems with using smart thermostat data to calculate the impact of the HVAC system on the occupant's thermal comfort is that we are unsure whether the occupant is in the house or not during a period of operation. If the occupant is not there in the house then the indoor conditions do not impact the occupant and so those periods should be left out of calculation. Often the indoor conditions might indicate extreme discomfort but the occupant has set them to be so in order to save energy whenever they are not in the house. So, the λ^i in equations 5.2,5.3 has a very important purpose as it filters out all the periods of inoccupancy that can skew the calculation of the impact score. Since smart thermostat data that includes occupancy sensor data is yet to be made available for now a statistical threshold has been developed to classify periods of inoccupancy. Once the occupancy sensor data is added to the smart thermostat data, λ^i value will be more precise, but conceptually the metric will not be affected and will still be valid. The thresholds used to decide if the period is occupied or not are:

1. PMV_{id} is greater than 1 but the cooling effort of the system in that period is 0. This means that the PMV is predicting that the indoor conditions were too hot for occupants to live comfortably but no action was taken for 2 hours to correct such a state. Hence, it is likely that these were periods of inoccupancy.

2. The cooling setpoint cannot be greater than $85°F$. Since, the study analyzes cooling systems, any time the setpoint is more than $85°F$, the period is either unoccupied or too extreme to study.

$$\lambda^i = 0 \quad when \begin{cases} E_c = 0 \; and \; PMV_{id} > 1 \\ CoolingSetpoint > 85°F. \end{cases} \tag{5.5}$$

Based on the regions labelled in fig.5.3 the points in green indicate all the periods of the system wherein the indoor environment is comfortable. A high number of green points indicates a more comfortable system because the value mean indoor discomfort ($\bar{D}_{id}$) will be lower. The points in orange describe all the periods where both the indoor and outdoor are uncomfortable. So, the values of discomfort indoors and outdoors will contribute to the value of the thermal comfort impact metric (Γ_I). Among these points the periods where the indoor discomfort is less than the outdoor discomfort are favorable because they lead to a lower value of the thermal comfort impact metric (Γ_I). These can be easily identified by drawing a line passing through origin inclined at $45°$ to the horizontal in fig. 5.3. All the orange points to the right are points where the indoor discomfort is lower and vice versa. Finally, the points in blue are periods when the indoor is uncomfortable but the outdoor is comfortable. These are worst kind of periods because they indicate that the HVAC system is causing the home to be uncomfortable. They bring down the value of outdoor discomfort but contribute more to the value of indoor discomfort. Hence, they adversely affect value of the thermal comfort impact metric (Γ_I).

Therefore, an air-conditioning system will have a low thermal comfort impact metric if it has most of its periods in the green region, a higher concentration of orange periods to the right of the line passing through the origin inclined at $45°$ to the X-axis and as few blue periods as possible.

The least value that the impact score can take is 0. This is possible when the indoor PMV (PMV_{id}) of the system is between -0.5 and 0.5 for every period of the system, which would make the degree of discomfort in every period equal to 0 and thereby the system would have a thermal comfort impact metric of 0. A thermal comfort impact metric of 100% means that on average the occupant would be as uncomfortable living in the house as they would be if they were camping in the outdoor environment. If the thermal comfort impact metric were more than 100% that would mean that the on average indoor environment is more uncomfortable than the outdoor environment.

Calculation Procedure for the Thermal Comfort Impact Metric (Γ_I)

1. Calculate the Predicted Mean Vote of thermal comfort under indoor (T_{id}, ϕ_{id}) and outdoor environments (T_{od}, ϕ_{od})

$$PMV_{id} = f\left(T_{id}, \phi_{id}, met, I_{cl}, V_a, \bar{T_r^{id}}\right) PMV_{od} = f\left(T_{od}, \phi_{od}, met, I_{cl}, V_a, \bar{T_r^{od}}\right) \tag{5.6}$$

2. Calculate the degree of discomfort under indoor (D_{id}^i) and outdoor environments (D_{od}^i).

$$\bar{D}_{id} = \frac{1}{N}\sum_{i=1}^{N}\left(\lambda^i D_{id}^i\right) \tag{5.7}$$

$$\bar{D}_{od} = \frac{1}{N}\sum_{i=1}^{N}\left(\lambda^i D_{od}^i\right) \tag{5.8}$$

3. Thermal Comfort Impact Metric (Γ_I) - Ratio of mean discomfort indoors ($\bar{D}_{id}$) and mean discomfort outdoors ($\bar{D}_{od}$).

$$\Gamma_I = \frac{\bar{D}_{id}}{\bar{D}_{od}} \tag{5.9}$$

6. COMBINING THE ENERGY AND THERMAL COMFORT IMPACT METRICS TO CONSTRUCT A FAULT SEVERITY INDEX

The previous chapters have outlined a method to calculate the impact a fault has on the energy consumption of an HVAC system and the thermal comfort it provides to the occupant. The more impact a fault in the system has on its occupant, the worse it is and more urgent is the need for repair. However, since the impact in each of the two avenues are equally important, in order to find the more severely affected system, both of them need to be combined into one. This will be done by reducing the thermal discomfort of 2-hour period of the AC to zero, and subsequently examining the change in energy impact. This means that the PMV_{id} for each observed 2-hour period of the system would have to be between -0.5 and 0.5. The time periods that previously had their PMV_{id} between (-0.5, 0.5) will continue to have the same but time periods that were outside will be brought towards to the edges of the comfortable boundary. So for periods with PMV_{id} greater than 0.5, a change is required so the indoor PMV becomes equal 0.5 and similarly for periods with PMV_{id} less than -0.5 a change is required so the indoor PMV becomes equal to -0.5.

Changes in the environment required for a change in PMV_{id} can include: changing layers of clothing, changing the temperature or relative humidity, reducing activity levels, or by reducing penetration of solar heat. However, four of the six inputs to calculate PMV_{id} are based on assumptions; and since sensitivity of the occupants to these assumptions is not calculable using the available data, they are left unchanged. Changes in indoor temperature and indoor relative humidity values, however, can be monitored using the data available from smart thermostats. Since, the linear model developed in equation 4.6 directly models the relationship between the cooling effort of the air-conditioning system and the temperature difference between the outdoors and indoors of the house, changing the indoor temperature of the house to cause a change in PMV can be complemented by using the model to calculate the change in cooling effort required to cause a specific change in indoor temperature (assuming outdoor temperature does not change).

As evidenced by the graphical method of identifying comfort level of a house based on the ASHRAE standards [46], for a constant value of PMV, the indoor temperature will be linearly related to the indoor relative humidity given that assumptions are unchanging. So, to maintain the PMV value of an environment at 0.5 or -0.5, the indoor temperature and relative humidity should lie on a specified line. In order to reduce the PMV of an environment, the current study proposes to change the indoor temperature such that the 2-hour period lies on the edge of the comfort region boundary, keeping the relative humidity a constant. This can be done by building a linear model with PMV as the dependent variable and temperature and relative humidity as the independent variable as shown in equation 6.1 where PMV_{id} represents the indoor PMV, T_{id} the indoor temperature and ϕ_{id} the indoor relative humidity,

$$PMV_{id} = \alpha_0 + \alpha_1 T_{id} + \alpha_2 \phi_{id} \tag{6.1}$$

In equation 6.1, the coefficient of indoor temperature α_1 gives the change in PMV_{id} for $1°C$ change in indoor temperature provided the rest of the assumptions and the indoor relative humidity remain constant. Since the study aims to bring the PMV_{id} of each period down to the comfortable level by changing the indoor temperature (T_{id}), the change in T_{id} required to bring the required change in PMV_{id} can be calculated using the above coefficient α_1. If the new indoor temperature at which the indoor discomfort is 0 is given by $T_{id,c}$ and the change in temperature is given by $\Delta T_{ic} = T_{id} - T_{id,c}$, then,

$$\Delta T_{ic} = T_{id} - T_{id,c} = \begin{cases} PMV_{id} + 0.5\alpha_1, & \text{if } PMV_{id} < -0.5 \\ 0, & \text{if } -0.5 \leq PMV_{id} \leq 0.5 \\ PMV_{id} - 0.5\alpha_1, & \text{if } PMV_{id} > 0.5 \end{cases} \tag{6.2}$$

Using equation 6.15, the new indoor temperature required to make the environment completely comfortable can be calculated. Equation 6.15 indicates that if a 2-hour period of a house is uncomfortable because of it being too hot, then the ΔT_{ic} will be positive, so a reduction in temperature is required to make the house comfortable. However, if the 2-hour period is uncomfortable because of it being too cold then the ΔT_{ic} is negative and so an increase in temperature is required to make the house comfortable.

This change in indoor temperature causes a change in temperature difference between indoor and outdoor environments. A change in temperature difference causes a change in cooling effort. If the new temperature difference were to be represented by $\Delta T_{oi,c}$, then the system is required to run at a new cooling effort say $E_{S,c}$ (%). At this value of $\Delta T_{oi,c}$, the average system of the dataset will have a cooling effort of $E_{M,c}$ (%). The outdoor temperature given by T_{od} remains the same as the original value. The following equations can be used to calculate each of them:

$$\Delta T_{oi,c} = T_{od} - T_{id,c} \tag{6.3}$$

$$E_{S,c} = \beta_0^S + \beta_1^S \Delta T_{oi,c} \tag{6.4}$$

$$E_{M,c} = \beta_0^M + \beta_1^M \Delta T_{oi,c} \tag{6.5}$$

The difference in energy consumption between any given air-conditioning system and the air-conditioning system with average performance level can again be compared and the energy impact metric for the given system when all periods are made comfortable (Γ_E^c) can be calculated:

$$\Gamma_E^c = \frac{\sum (E_{S,c} - E_{M,c})}{\sum E_{S,c}} \tag{6.6}$$

This value of Γ_E^c is the value of energy impact metric that takes into consideration the air-conditioner's effect on occupants' thermal comfort. The value of the thermal comfort impact metric (Γ_I^c) calculated using this new indoor temperature value $(T_{id,c})$ will be close to 0. Therefore, air-conditioning systems can now be objectively compared against each other using a single metric which is the energy impact metric at comfortable operating conditions because at this hypothetical state the fault produces no impact on thermal comfort and only affects the energy consumption.

Hence, this energy impact metric estimated at the new "comfortable" indoor temperature is a direct judge of how severe the affect of the fault is and so it's defined as the fault severity index.

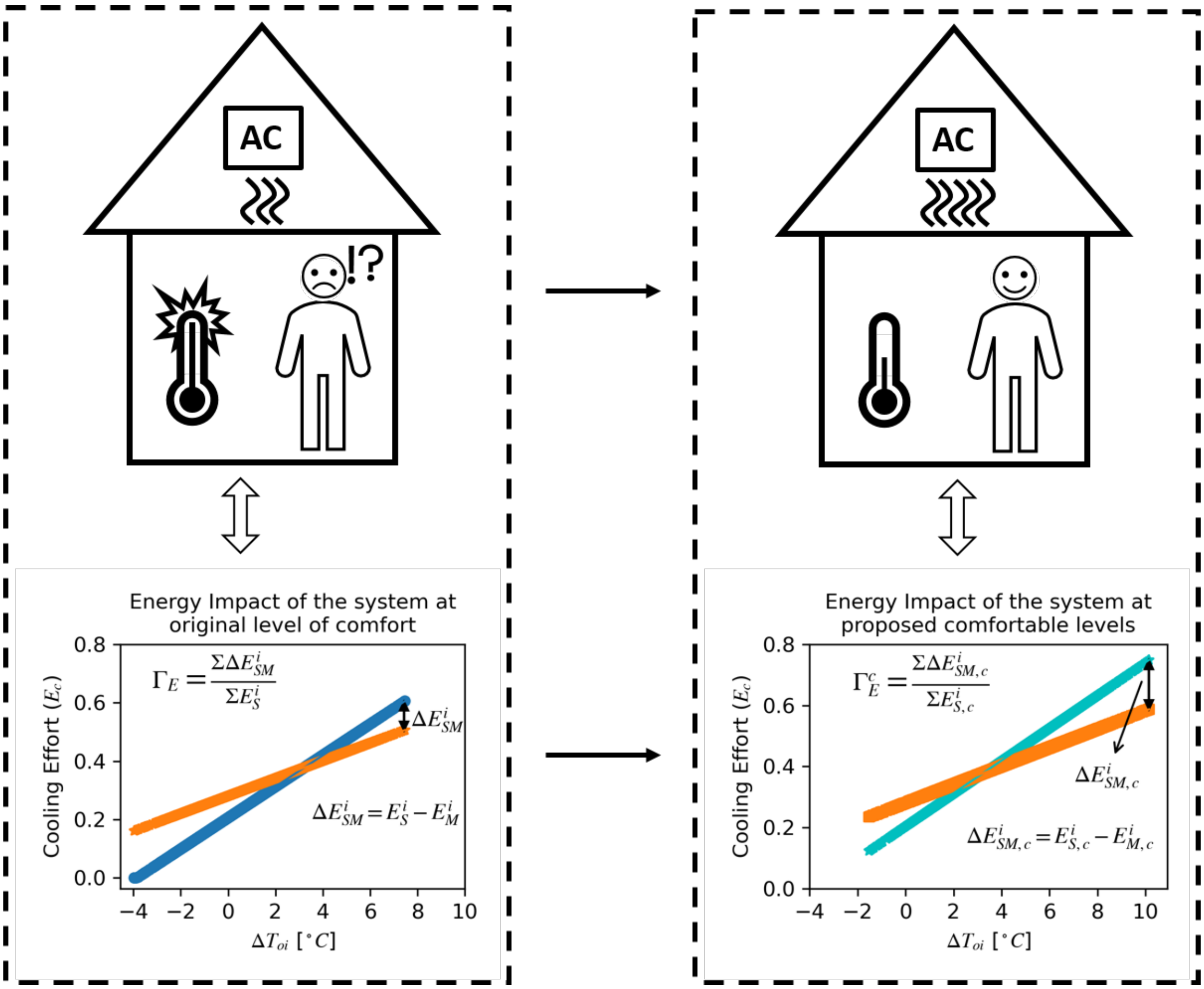


Figure 6.1: Fault severity index of a system is defined as the amount of impact a fault in the system has on its energy consumption given provided it produces no thermal discomfort to the occupants.

6.1 A mathematical exploration of the fault severity index (Γ^c_E)

The impact of a fault in an air-conditioning system can broadly be categorized to fall in one of two areas: the thermal discomfort the occupant feels and the relative extra energy the system consumes. However, these two are not mutually exclusive avenues. For an AC installed in a poorly insulated house its cooling effort is higher than what it would have been in house with the mean insulation level. Simultaneously, this system and its fault (poor insulation) by the virtue of the environment it creates also has an impact on the thermal comfort level of occupant. If the system creates a thermal environment that's on average colder than the comfortable, then the occupant is using extra energy (thereby increasing the value of the energy impact metric) to create an environment that's uncomfortable (thereby increasing the value of the thermal comfort impact metric). Therefore, adjusting the thermal comfort level of the house to the comfortable level will mean that the system will use lesser energy than before which *may* in turn reduce its energy impact. This adjusted energy impact is a good measure of the severity of impact of a fault in the system on its occupant.

This following section will explore the derivation of the fault severity index of an air-conditioning system (Γ_E^c) mathematically. Please note that the summation in the following derivation is done over all 2-hour periods of operation of each AC but for the sake of brevity the representation of each two hour period by the superscript "i" is abandoned. Substituting the models for E_S^c and E_M^c, from equations 6.17, 6.18 into equation 6.19 we get,

$$\Gamma_E^c = \frac{\sum \left(\left(\beta_0^S + \beta_1^S \Delta T_{oi,c} \right) - \left(\beta_0^M + \beta_1^M \Delta T_{oi,c} \right) \right)}{\sum \left(\beta_0^S + \beta_1^S \Delta T_{oi,c} \right)} \tag{6.7}$$

The above equation can further be expanded by using the definition of ΔT_{oi}^c and $T_{id,c}$

$$\Delta T_{oi,c} = T_{od} - T_{id,c} \tag{6.8}$$

$$T_{id,c} = T_{id} - \Delta T_{ic} \tag{6.9}$$

$$\Delta T_{oi,c} = T_{od} - T_{id} + \Delta T_{ic} \tag{6.10}$$

$$\Delta T_{oi,c} = \Delta T_{oi} + \Delta T_{ic} \tag{6.11}$$

The fault severity index (Γ_E^c) therefore, can be written as:

$$\Gamma_E^c = \frac{\sum \left(\left(\beta_0^S - \beta_0^M \right) + \Delta T_{oi} \left(\beta_1^S - \beta_1^M \right) + \Delta T_{ic} \left(\beta_1^S - \beta_1^M \right) \right)}{\sum \left(\beta_0^S + \beta_1^S \Delta T_{oi} + \beta_1^S \Delta T_{ic} \right)} \tag{6.12}$$

$$\Gamma_E^c = \frac{\sum \left((E_S - E_M) + \Delta T_{ic} \left(\beta_1^S - \beta_1^M \right) \right)}{\sum \left(E_S + \beta_1^S \Delta T_{ic} \right)} \tag{6.13}$$

Now that the equation for the fault severity index (Γ_E^c) has been expanded, it can be compared with the equation for a fault's energy impact (Γ_E) from equation 4.9,

$$\Gamma_E = \frac{\sum (E_S - E_M)}{\sum E_S} \tag{6.14}$$

Based on the comparison, the severity index of a fault can be observed to take into account the effects of its energy impact at the original level of comfort and the cost of improvement of comfort. The energy impact metric at the original level of comfort is proportional to the difference between the performance of the given and the average system. The cost of improvement of comfort on the other hand is proportional to the temperature change required which in turn directly affects the cooling effort via. the means of β_1, the coefficient of ΔT_{oi}. Therefore, the fault severity index proves to be a holistic combination of each of the individual impact metrics and a good judge of severity of the fault's cumulative impact on the occupant.

Calculation procedure for the Fault Severity Index (Γ_E^c)

1. Estimate the reduction in indoor temperature required to make each of the 2-hour periods of a system comfortable

$$\Delta T_{ic} = T_{id} - T_{id,c} = \begin{cases} PMV_{id} + 0.5\alpha_1, & \text{if } PMV_{id} < -0.5 \\ 0, & \text{if } -0.5 \leq PMV_{id} \leq 0.5 \\ PMV_{id} - 0.5\alpha_1, & \text{if } PMV_{id} > 0.5 \end{cases} \tag{6.15}$$

2. Estimate the cooling effort of the given air-conditioning system and the cooling effort of the system with average performance level under the new "comfortable" indoor conditions.

$$\Delta T_{oi,c} = T_{od} - T_{id,c} \tag{6.16}$$

$$E_{S,c} = \beta_0^S + \beta_1^S \Delta T_{oi,c} \tag{6.17}$$

$$E_{M,c} = \beta_0^M + \beta_1^M \Delta T_{oi,c} \tag{6.18}$$

3. Fault severity index (Γ_E^c) – Value of the energy impact metric of the given system provided the value of the thermal comfort impact metric is close to 0.

$$\Gamma_E^c = \frac{\sum (E_{S,c} - E_{M,c})}{\sum E_{S,c}} \tag{6.19}$$

6.2 Calculating the percentage change in cooling hours required

Since the calculation of the fault severity index involves estimating the cooling effort of the air-conditioning system when all the observed operating periods of the house are comfortable, another metric that can be estimated along with the severity index is change in cooling hours required to make the house comfortable. Supposing this value is represented by ΔE_c, then it can be calculated using the following equation,

$$\Delta E_c = \frac{\sum E^i_{S,c} - \sum E^i_S}{\sum E^i_S} \tag{6.20}$$

The cooling effort of a given system in a 2-hour period is the proportion of time the system is running during the 2-hour period. So total time the system is running in that 2-hour period is twice the cooling effort in that period. The summation of total cooling hours for the given system is twice the summation of cooling effort of the system in each period. Hence, the total cooling hours of the given system at the original level of comfort is $2\sum(E^i_S)$ and the total cooling hours of the system after making the house comfortable is given by $2\sum(E^i_{S,c})$. The difference between the two is change in cooling hours when the house is made comfortable. Upon expanding equation 6.20 using the definitions of $E^i_{S,c}$ and E^i_S, the percentage of extra cooling hours is transformed into:

$$\Delta E_c = \frac{\sum(\beta_0 + \beta_1 \Delta T_{oi,c}) - \sum(\beta_0 + \beta_1 \Delta T_{oi})}{\sum(\beta_0 + \beta_1 \Delta T_{oi})} \tag{6.21}$$

$$\Delta T_{oi,c} - \Delta T_{oi} = (T_{od} - T_{id,c}) - (T_{od} - T_{id}) \tag{6.22}$$

$$\Delta T_{oi,c} - \Delta T_{oi} = \Delta T_{ic} \tag{6.23}$$

$$\Delta E_c = \frac{\sum \beta_1 \Delta T_{ic}}{\sum(\beta_0 + \beta_1 \Delta T_{oi})} \tag{6.24}$$

Therefore, the percentage change in cooling hours required to make an indoor environment comfortable is dependent on the change in indoor temperature required to make the environment comfortable. So, if the indoor environment is on average "hotter" than comfortable $(PMV_{id} > 0.5)$, then a reduction in temperature is required to make the home comfortable and so the value of $\Delta E_c > 0$ and if the indoor environment is "colder" than comfortable $(PMV_{id} < -0.5)$ then the indoor temperature must be increased to make the home comfortable and so the value of $\Delta E_c < 0$. Conversely, the sign of ΔE_c can be used an indicator to know on average whether the home is hotter than comfortable or colder than comfortable. This metric along with the fault severity index and energy impact metric of the system will now be used to characterize its performance in the following chapters.

7. RELATIONSHIP BETWEEN THE BEHAVIOR OF AN HVAC SYSTEM AND ITS METRICS

Four metrics in total have been defined to represent the performance of an HVAC system and the effect of its faults on the occupants. The following chapter is an explanation of how they could be used individually and together to describe the behavior of the system. Firstly, the four metrics that have been defined are summarized below along with few equations to aid the reader understand this chapter:

1. ***Energy Impact Metric*** (Γ_E): The energy impact metric of an air-conditioning system is the relative difference in total cooling effort of the given system and the system that represents the average performance level of systems in the dataset. The energy models of a single system and the average system for the i^{th} 2-hour period as defined in equations 4.6, 4.5 are given below,

$$E_S^i = \beta_0^S + \beta_1^S \Delta T_{oi}^i \tag{7.1}$$

$$E_M^i = \beta_0^M + \beta_1^M \Delta T_{oi}^i \tag{7.2}$$

   The mathematical definition of the energy impact metric as constructed in equation 4.9 is,

$$\Gamma_E = \frac{\sum_{i=1}^{N}(E_S^i - E_M^i)}{\sum_{i=1}^{N} E_S^i} = \frac{\sum_{i=1}^{N}[(\beta_0^S + \beta_1^S \Delta T_{oi}^i) - (\beta_0^M + \beta_1^M \Delta T_{oi}^i)]}{\sum_{i=1}^{N}(\beta_0^S + \beta_1^S \Delta T_{oi}^i)} \tag{7.3}$$

2. ***Thermal Comfort Impact Metric*** (Γ_I): The thermal comfort impact is the ratio of the average amount of discomfort felt by the occupant living in inside the house to the average amount of discomfort felt by occupant if he were living in the outdoor environment of the house.

3. ***Fault Severity Index*** (Γ_E^c): Fault severity index of an air-conditioning system is defined as the amount of impact a fault in the system has on its energy consumption given provided it produces no thermal discomfort to the occupants. The energy models of a given single

system and the average system for the i^{th} 2-hour period given that that they produce a completely comfortable environment as defined in equations 6.17, 6.18 are given below,

$$E^i_{S,c} = \beta^S_0 + \beta^S_1 \Delta T^i_{oi,c} \tag{7.4}$$

$$E^i_{M,c} = \beta^M_0 + \beta^M_1 \Delta T^i_{oi,c} \tag{7.5}$$

The definition of $\Delta T^i_{oi,c}$ and is, $\Delta T^i_{oi,c} = T^i_{od} - T^i_{id,c}$. This when combined with the temperature difference between current indoor temperature (T^i_{id}) and comfortable indoor temperature ($T^i_{id,c}$), ΔT^i_{ic} leads to equation 6.11 which is $\Delta T^i_{oi,c} = \Delta T^i_{oi} + \Delta T^i_{ic}$. When substituted back into the above equations,

$$E^i_{S,c} = \beta^S_0 + \beta^S_1 \Delta T^i_{oi,c} = (\beta^S_0 + \beta^S_1 \Delta T^i_{oi}) + \beta^S_1 \Delta T^i_{ic} = E^i_S + \beta^S_1 \Delta T^i_{ic} \tag{7.6}$$

$$E^i_{M,c} = \beta^M_0 + \beta^M_1 \Delta T^i_{oi,c} = (\beta^M_0 + \beta^M_1 \Delta T^i_{oi}) + \beta^M_1 \Delta T^i_{ic} = E^i_M + \beta^M_1 \Delta T^i_{ic} \tag{7.7}$$

The definition of the fault severity index of a system given in equation 6.19 can be re-written as,

$$\Gamma^c_E = \frac{\sum(E^i_{S,c} - E^i_{M,c})}{\sum E^i_{S,c}} = \frac{\sum[(E^i_S - E^i_M) + \Delta T^i_{ic}(\beta^S_1 - \beta^M_1)]}{\sum(E^i_S + \beta^S_1 \Delta T^i_{ic})} \tag{7.8}$$

4. ***Percentage change in cooling hours*** (ΔE_c): ΔE_c is the percentage change in cooling hours required to make the system completely comfortable at all points in time. As defined in equation 6.20 the equation for ΔE_c can now be transformed based on the definition of $E^i_{S,c}$ and E^i_S and using 6.11 which is $\Delta T^i_{oi,c} = \Delta T^i_{oi} + \Delta T^i_{ic}$ into,

$$\Delta E_c = \frac{\sum(E^i_{S,c} - E^i_S)}{\sum E^i_S} = \frac{\sum(\beta^S_1 \Delta T^i_{oi,c} - \beta^S_1 \Delta T^i_{oi})}{\sum(\beta^S_0 + \beta^S_1 \Delta T^i_{oi})} \tag{7.9}$$

$$\implies \Delta E_c = \frac{\sum \beta^S_1 \Delta T^i_{ic}}{\sum(\beta^S_0 + \beta^S_1 \Delta T^i_{oi})} \tag{7.10}$$

The energy impact metric of a given air-conditioning system is a measurement of how many hours fewer does a system with the average performance level take than the given system to produce the same environment. Since, this value is calculated using the data obtained directly from the regulating periods of the system it also includes the effect of the thermal discomfort on the energy consumption of the system. The value of the thermal comfort impact metric and the energy impact metric of the system are coupled to each other because they estimate the effect a fault in the AC has on on the occupant in two mutually dependent avenues. This coupled nature of both of the impact metrics makes it harder to compare two systems based on only one metric. Furthermore, the method of coupling is also difficult to decipher. In some cases, the thermal discomfort may cause the system to consume more energy and hence have a higher energy impact. In such cases, adjusting the setpoint to a make the thermal discomfort zero is to the advantage of the occupant because it leads to a decrease in energy impact. However, this can come at the cost of increased energy consumption which may not be acceptable for the occupant; or a decrease in energy consumption in which case there is an extra incentive for the occupant to adjust his setpoint because it leads to better comparative performance (as indicated by a decrease in the value of the energy impact metric) as well as lesser energy consumption (as indicated by decrease in cooling hours). On the other hand, thermal discomfort can cause the energy impact to decrease in which case the true severity of the fault will be hidden behind a low value of the energy impact metric. This also leads to several cases that again need to be considered. Therefore, delineating the relationship between each of the 4 metrics for each system is a complex task but provides immense opportunities to understand system behavior in greater detail.

7.1 Systems with extremely high fault severity index value

The fault severity index of an air-conditioning system (Γ_E^c) is the value of energy impact after the performance of the system has been adjusted to produce no thermal discomfort. Therefore, it can now be used to compare two systems objectively. Comparing the fault severity index implies, comparing the impact of the fault on the energy consumption of the systems provided both systems are producing no discomfort in each of their own individual homes. It is a more objective way to assess how much better or worse one system performs than another, particularly because the discomfort level is a parameter that is local to the system and its effect has thus been removed.

Air-conditioning systems that have a very high fault severity index are therefore most likely to be faulty, irrespective of any other factor. They can be identified by segregating the systems whose severity indices lie within an upper threshold in the distribution of severity indices from all systems. These systems essentially, consume an abnormally high amount of energy than the mean system would have in the same environment to produce a completely comfortable environment. The most probable reason for this is that they might have a mechanical fault causing an inadequate capacity to keep with the heat demands of the house. Mechanical faults can include: less refrigerant flowing in the system, obstruction in air-vents, system is undersized for the house it is installed in, extremely poor insulation, etc. These systems should therefore be screened out and studied separately from the rest of the systems in the dataset. The plot in fig. 7.1 shows the histogram plot of the fault severity index of 7000 systems in Florida (2A climate region) with the systems that lie in the upper 5% threshold highlighted,

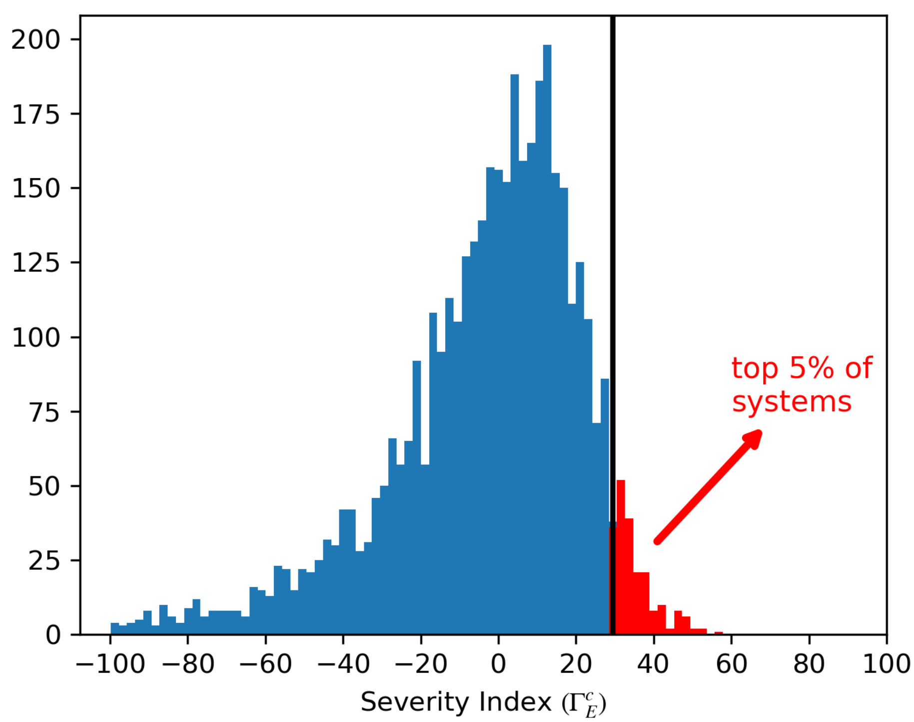


Figure 7.1: Histogram plot of fault severity index of 7000 single-speed cooling systems in Florida with the upper 5% highlighted.

7.2 Air-conditioning systems with a low to moderate fault severity index

In case of systems with a fault severity index below the threshold to be qualified as a system with a mechanical fault, the four metrics can now be further examined to describe system behavior. First, to make the analysis convenient the systems will be divided into two broad groups: Systems with a positive value of fault severity index and systems with a negative value of fault severity index. Air-conditioners with a positive fault severity index ($\Gamma_E^c > 0$) are having to consume more energy than the average system because of the fault and air-conditioners with a negative fault severity index ($\Gamma_E^c < 0$) consume less energy than the average system, provided the thermal discomfort produced by the systems is zero.

7.2.1 Air-conditioning systems with a positive fault severity index ($\Gamma_E^c > 0$) – denoted by (+)

The fault severity index of systems that belong in this category is greater than 0 which means that they consume more energy than the average system to produce the same completely comfortable environment. However, this does not indicate anything about the energy impact metric value of the system at original level of comfort. Correction of thermal discomfort level of the home can lead to an increase in the value of the energy impact metric to higher value or decrease to a lower value. Simultaneously, it can come at the cost of extra energy consumed or at the benefit of savings in energy consumption. Therefore, the category of systems with positive fault severity index (Γ_E^c) can further be divided into sub-categories based on their values of percentage change in cooling hours (ΔE_c) and the energy impact metric at original level of comfort (Γ_E). But before that upon mathematically expanding the terms of the equation representing the fault severity index and considering the fact that cooling effort of the system is always positive ($E_{S,c}^i > 0$) the following condition is obtained,

$$\Gamma_E^c = \frac{\sum \left(E_{S,c}^i - E_{M,c}^i\right)}{\Sigma E_{S,c}^i} > 0 \tag{7.11}$$

$$\implies \sum (E_{S,c}^i - E_{M,c}^i) > 0 \tag{7.12}$$

Upon using the expansions derived in equations 7.6 ($E^i_{S,c} = E^i_S + \beta^S_1 \Delta T^i_{ic}$) and 7.7 ($E^i_{M,c} = E^i_M + \beta^M_1 \Delta T^i_{ic}$) we get,

$$\sum (E^i_{S,c} - E^i_{M,c}) > 0 \tag{7.13}$$

$$\implies \sum [(E^i_S - E^i_M) + \Delta T^i_{ic}(\beta^S_1 - \beta^M_1)] > 0 \tag{7.14}$$

$$\implies \Gamma_E \sum E^i_S + \sum \Delta T^i_{ic}(\beta^S_1 - \beta^M_1)] > 0 \tag{7.15}$$

Therefore this condition, therefore, forms the basis of all the systems that belong to this category. It will be used later on this section.

7.2.1.1 Air-conditioning systems that consume more energy when they are made comfortable ($\Delta E_c > 0$) – denoted by ***(++)***

Air-conditioning systems in this category undergo an increase in total cooling hours upon the reduction of the thermal discomfort. An increase in total number of cooling hours can only happen with an increase in the values of temperature difference, because the coefficients of the energy consumption model of systems are always positive. An increase in temperature difference values happens with a decrease in indoor temperature values, given that the outdoor temperature remains the same. This happens whenever the occupants are keeping their indoor environment at a temperature that is hotter than what would be comfortable and a reduction in temperature is necessary to improve comfort. This is mathematically proved using equation 7.10 and given that $E^i_S > 0$ (cooling effort of a 2-hour period cannot be negative) and $\beta^S_1 > 0$ (coefficient of ΔT_{oi} cannot be negative because a system's cooling effort cannot decrease with increasing temperature difference),

$$\Delta E_c = \frac{\Sigma E^i_{S,c} - \Sigma E^i_S}{\Sigma E^i_S} = \frac{\Sigma \beta_1 \Delta T^i_{ic}}{\Sigma\left(\beta_0 + \beta_1 \Delta T^i_{oi}\right)} > 0 \tag{7.16}$$

$$\implies \beta^S_1 \Sigma \Delta T^i_{ic} > 0 \tag{7.17}$$

$$\implies \Sigma \Delta T^i_{ic} > 0 \tag{7.18}$$

$$\implies \Sigma(T^i_{id} - T^i_{id,c}) > 0 \implies \Sigma T^i_{id,c} < \Sigma T^i_{id} \tag{7.19}$$

Owing to the monetary cost that accompanies the increase in total cooling hours the occupant often has to decide whether or not to reduce the discomfort level of his house and by how much. If the occupant wants to make all the time they in the home thermally comfortable then they will have to increase the cooling hours of their system by a minimum of $\Delta E_c\%$. Subsequently this case now further be divided into two sub cases that each represent the change in the value of the energy impact metric of the system upon adjusting its performance to obtain a negligible value of the thermal comfort impact metric.

1. Correction of discomfort level of the indoor environment is worsening the air-conditioning system's relative performance ($\Gamma_E^c > \Gamma_E$) – denoted by **(+++)**

   This case represents air-conditioning systems where the value of the relative energy consumption of the system increases whenever it is driven to run at a setpoint that produces a comfortable indoor temperature. The energy consumption of a system and the thermal comfort level of the house are coupled with each other. The metrics can help decouple them and aid in improving our understanding as shown in the derivation below. Please note that for the sake of brevity the i's that represent each 2-hour period of operation on top of each of the variables are dropped. Using the definitions of Γ_E^c and Γ_E from equations 7.8 and 7.3 we get,

$$\Gamma_E^c > \Gamma_E \tag{7.20}$$

$$\implies \frac{\sum \left((E_S - E_M) + \Delta T_{ic}\left(\beta_1^S - \beta_1^M\right)\right)}{\sum \left(E_S + \beta_1^S \Delta T_{ic}\right)} > \frac{\sum \left(E_S - E_M\right)}{\sum E_S} \tag{7.21}$$

   Upon cross-multiplying and cancelling the common terms we get,

$$\implies (\beta_1^S - \beta_1^M)(\sum \Delta T_{ic})(\sum E_S) > (\sum(E_S - E_M))(\sum \Delta T_{ic})\beta_1^S \tag{7.22}$$

$$\implies \frac{\sum \left(E_S - E_M\right)}{\sum E_S} < \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.23}$$

$$\implies \Gamma_E < \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.24}$$

This means that systems in this category have an upper bound on the value of the energy impact metric. However, from equations 7.15 and 7.18 and we have the following,

$$\Gamma_E \sum E_S^i + (\beta_1^S - \beta_1^M) \sum \Delta T_{ic}^i > 0 \tag{7.25}$$

where $\sum E_S^i > 0$ and $\sum \Delta T_{ic} > 0$. Since the sign of Γ_E is unknown, both cases of it being positive and negative need to be considered separately. If $\Gamma_E > 0$, according to equation 7.15, $(\beta_1^S - \beta_1^M)$ can be positive or negative, however from equation 7.24,

$$0 < \Gamma_E < \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.26}$$

$$\implies \frac{\beta_1^S - \beta_1^M}{\beta_1^S} > 0 \tag{7.27}$$

$$\implies \beta_1^S > \beta_1^M \tag{7.28}$$

If $\Gamma_E < 0$, from equation 7.24, $(\beta_1^S - \beta_1^M)$ can again be positive or negative, however from equation 7.15,

$$(\sum \Delta T_{ic})(\beta_1^S - \beta_1^M)] > \Gamma_E \sum E_S^i > 0 \tag{7.29}$$

$$\beta_1^S > \beta_1^M \tag{7.30}$$

Therefore, irrespective of the sign of Γ_E, the mathematical condition that holds true in this case is that $\beta_1^S > \beta_1^M$. This signifies that all the air-conditioning systems that belong to (+++) will have their coefficient of temperature difference in their model for energy consumption greater than that of the average system. Since we have previously noted that the coefficient of temperature difference could correspond to the (insulation + infiltration) level per unit capacity, we could infer that for the cases where in the fault severity index is positive, the system requires more energy to become comfortable which causes an increase in the system's relative energy metric the insulation level per unit capacity must be lower (and/or the infiltration level per unit capacity) must be higher than that of the average system. The reader should note that the inverse of the statement is however untrue. The energy consumption models of systems belonging to the case (+++) will have lower slopes than that of the average system but all systems with lower slopes need not belong to this particular case. Further corroboration for this can be derived by plotting the histogram of slopes and intercept of all the models of systems belonging to this case against the slope and intercept of the energy consumption model of the average system as shown in fig. 7.2. Systems lying in this region, therefore, are most likely to posses a mechanical fault, because the thermal comfort and energy impact metrics are getting coupled together in a detrimental way. These faults need to be observed carefully to make sure that the fault severity index value of the system stays under a threshold. The only definitive conclusion that can be made is that in case of systems possessing such kind of high likelihood for the existence of mechanical fault, will have a lower energy model whose slope is greater than slope of the model of the system with the average performance level. This could correspond to a poorer insulation and infiltration level on average.

2. Correction of discomfort level of the indoor environment is improving the air-conditioning system's relative performance ($\Gamma_E^c < \Gamma_E$) – denoted by **(++-)**

   Improving the comfort level of the air-conditioning systems falling in this category is causing an increase in energy consumption but a reduction in the overall value of the energy impact metric. After increasing the number of cooling hours by $\Delta E_c\%$ the resulting performance level of the system is still worse than the performance level of the average system but better than what it used to be before. So systems in this case will benefit from reduction of thermal discomfort of the occupants with regards to relative energy consumption. However, this reduction will come at the cost of more energy being consumed by the system. So, the decision on whether or not to improve their thermal comfort and by what magnitude should be left to the occupant. In order to mathematically explore this case, a procedure similar to the one done for the previous case can be followed keeping in mind that $\sum \Delta T_{ic} > 0$ and

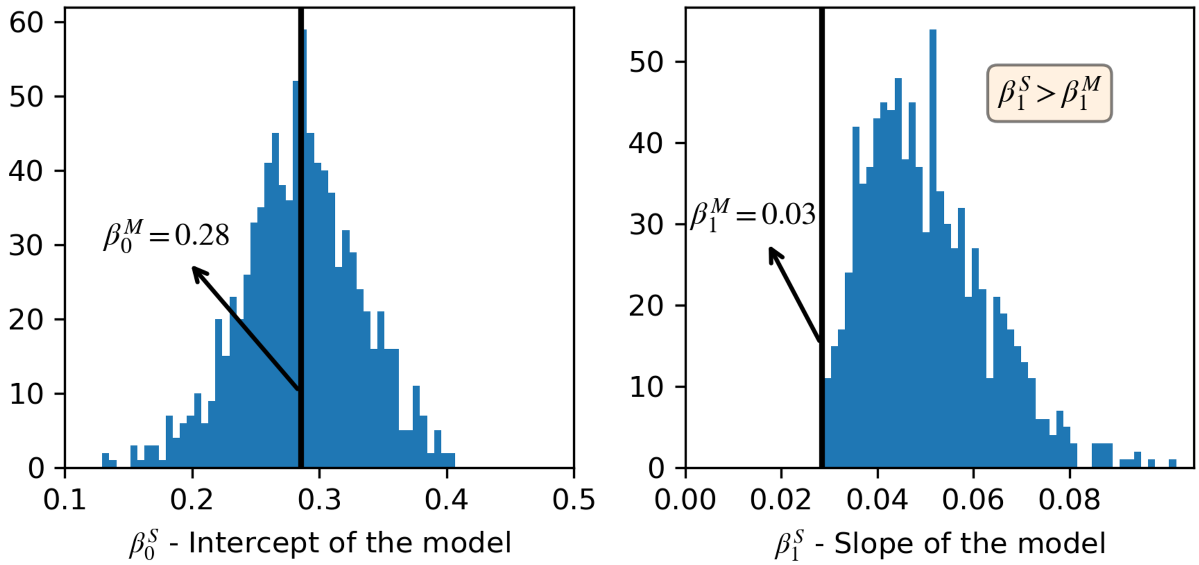


Figure 7.2: Histogram plot of intercept and slopes of models of all air-conditioning systems wherein a fault is causing a severe impact on the thermal comfort of the occupants which when corrected is causing an increase in its energy consumption relative to that of the average system. Systems likely possess mechanical faults like of lack of refrigerant or undersizing issues etc.

using the equations 7.3 and 7.8,

$$\Gamma_E^c < \Gamma_E \tag{7.31}$$

$$\implies \frac{\sum\left[(E_S - E_M) + \Delta T_{ic}\left(\beta_1^S - \beta_1^M\right)\right]}{\sum\left(E_S + \beta_1^S \Delta T_{ic}\right)} < \frac{\sum\left(E_S - E_M\right)}{\sum E_S} \tag{7.32}$$

Upon cross-multiplying and cancelling the common terms we get,

$$\implies (\beta_1^S - \beta_1^M)(\sum \Delta T_{ic})(\sum E_S) < (\sum(E_S - E_M))(\sum \Delta T_{ic})\beta_1^S \tag{7.33}$$

$$\implies \frac{\sum\left(E_S - E_M\right)}{\sum E_S} > \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.34}$$

$$\implies \Gamma_E > \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.35}$$

This implies that air-conditioning systems in this case have a lower bound for the value of the energy impact metric. However upon expanding this further using the expansion of Γ_E

given in equation 7.3 the following is obtained,

$$\Gamma_E > \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.36}$$

$$\frac{\Sigma[(\beta_0^S - \beta_0^M) + \Delta T_{oi}(\beta_1^S - \beta_1^M)]}{\Sigma(\beta_0^S + \beta_1^S \Delta T_{oi})} > \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.37}$$

$$(\beta_0^S - \beta_0^M)\beta_1^S > (\beta_1^S - \beta_1^M)\beta_0^S \tag{7.38}$$

$$\frac{\beta_0^S}{\beta_1^S} > \frac{\beta_0^M}{\beta_1^M} \tag{7.39}$$

Therefore, upon mathematically examining the case we arrive another condition for the systems belonging to this case. In order to understand what it means the histogram plot of the slopes and intercepts is plotted below,

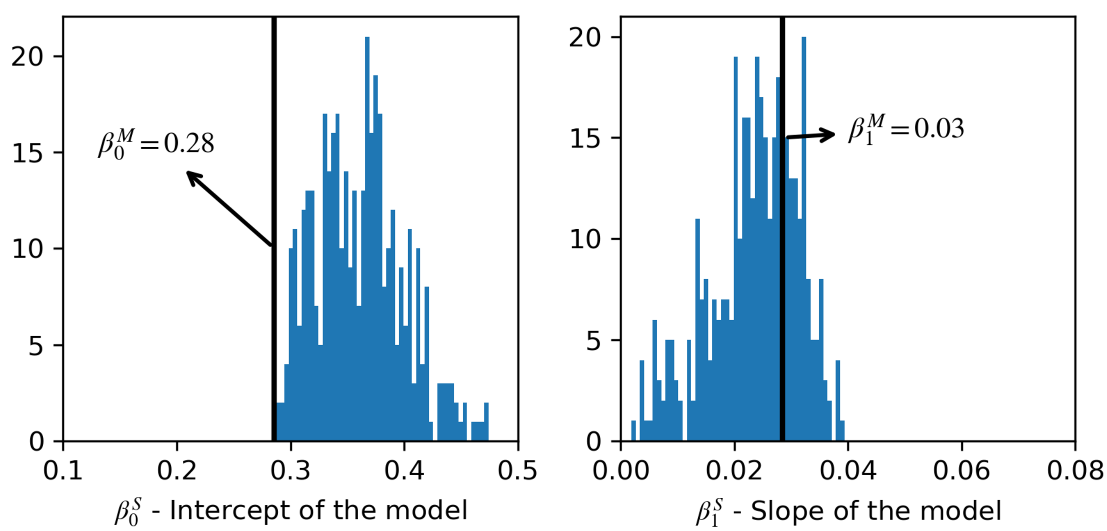


Figure 7.3: Histogram plot of intercept and slopes of models of all air-conditioning systems wherein a fault is causing a severe impact on the thermal comfort of the occupants making them feel hot; which when corrected is causing an decrease in its energy consumption relative to that of the average system.

As a consequence of the slope and intercept of the energy consumption model of the average system being as indicated in fig. 7.3, the condition in equation 7.39, leads to $\beta_0^S > 9.3\beta_1^S$. This condition as expressed in fig. 7.3 forces the intercept of the models of systems falling in this category to be more than that of the average system of the dataset. Going by our analogy, this case in essence captures all houses that on average experience a high internal load and/or a high solar load. Since, the houses already have a high internal load and are being maintained at hotter than comfortable temperatures, reduction in indoor temperature brings the energy consumption level of the system closer to that of the mean system. The higher the reduction in indoor temperature, the lesser will be the influence of the house's high internal load. This implies that the "fault" for systems belonging to this case is that the house they are installed in experience high internal and solar loads on average. By reducing the internal load of the house the difference between the cooling efforts of the given system and the average system will decrease causing a reduction in energy consumption.

*7.2.1.2 Air-conditioning systems that consume less energy upon becoming comfortable ($\Delta E_c < 0$) – denoted by **(+-)***

Air-conditioning systems falling in this category experience a reduction in total cooling hours i.e., total energy consumed, when the thermal discomfort of the house is reduced. These are houses wherein the indoor environments are "colder" than comfortable (according to the PMV index) and hence will benefit from an increase in setpoint. An increase in setpoint implies lesser amount of total cooling and therefore fewer cooling hours in total. Therefore, expressed mathematically using the fact that $E_S > 0$ we get,

$$\Delta E_c = \frac{\Sigma E_{S,c} - \Sigma E_S}{\Sigma E_S} = \frac{\Sigma \beta_1 \Delta T_{ic}}{\Sigma \left(\beta_0 + \beta_1 \Delta T_{oi}\right)} < 0 \tag{7.40}$$

$$\implies \beta_1^S \Sigma \Delta T_{ic} < 0 \tag{7.41}$$

$$\implies \Sigma \Delta T_{ic} < 0 \tag{7.42}$$

$$\implies \Sigma(T_{id} - T_{id,c}) < 0 \implies \Sigma T_{id,c} > \Sigma T_{id} \tag{7.43}$$

Owing to the savings in money that accompanies the improvement of thermal comfort level of the house, occupants of these houses will be typically be keen to adjust their setpoint accordingly. The ΔE_c represents the minimum amount of savings in energy possible when the setpoint of the air-conditioner is altered such that the house is completely comfortable at all points in time. However, as in the case above, the reduction of thermal discomfort can come with an increase in the value of the energy impact metric or a decrease, which leads to the following cases.

1. Correction of discomfort level of the indoor environment is worsening the air-conditioning system's relative performance ($\Gamma_E^c > \Gamma_E$) – denoted by **(+-+)**

   As in the case (+++), air-conditioning systems falling in this category experience a further deviation from the average system whenever the thermal comfort level of the house is improved. Since, the thermal discomfort is in fact causing an increase in the value of the energy impact metric, there is a need to monitor the performance of this system closely because of the possibility of existence of mechanical faults. As previously done, a mathematical exploration of this case is given below keeping in mind that condition in equation 7.42 and using equations 7.3 and 7.8,

$$\Gamma_E^c > \Gamma_E \tag{7.44}$$
$$\implies \frac{\sum \left[(E_S - E_M) + \Delta T_{ic}\left(\beta_1^S - \beta_1^M\right)\right]}{\sum \left(E_S + \beta_1^S \Delta T_{ic}\right)} > \frac{\sum \left(E_S - E_M\right)}{\sum E_S} \tag{7.45}$$

   Upon cross-multiplying and cancelling the common terms we get,

$$\implies (\beta_1^S - \beta_1^M)(\sum \Delta T_{ic})(\sum E_S) > (\sum (E_S - E_M))(\sum \Delta T_{ic})\beta_1^S \tag{7.46}$$
$$\implies \frac{\sum \left(E_S - E_M\right)}{\sum E_S} > \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.47}$$
$$\implies \Gamma_E > \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.48}$$

   As observed in the case (++-) we again have reached a minimum value for the value of the energy impact metric of an air-conditioning system belong to this category. Therefore, upon expanding the above equation as done previously, we reach the same final condition as in equation 7.39

$$\frac{\beta_0^S}{\beta_1^S} > \frac{\beta_0^M}{\beta_1^M} \tag{7.49}$$

Figure 7.4 confirms that since the mathematical condition in 7.49 is identical to the one

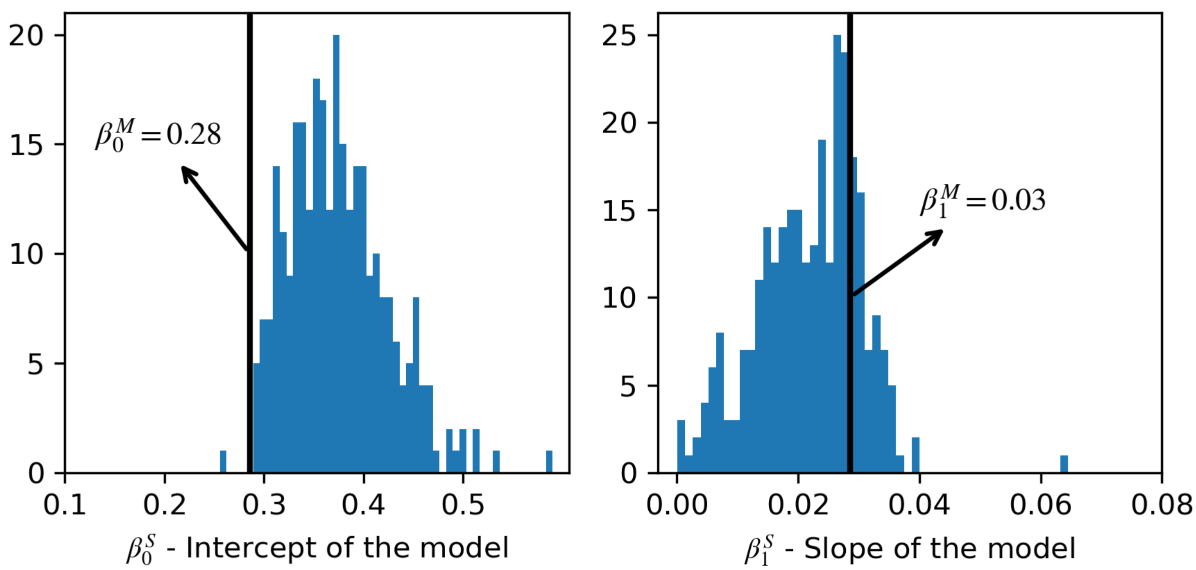


Figure 7.4: Histogram plot of intercept and slopes of models of all the system wherein the house is over-cooled the correction of which causes an increase in relative energy consumption.

observed in case of (++-): the histogram plots also show similar trends. This case too captures air-conditioning systems installed in houses that experience a high internal/solar load on average. Since, an increase in overall temperature is observed, the internal load starts to have a greater effect on the value of the energy impact metric than before causing the difference between the given system and the average system to increase in general. Therefore, the advantage of reducing the thermal discomfort of the house, which is reduction in total cooling hours, should be weighed against the deviation from the average system. As in the

(++-) case, the fault is the high value of the intercept of the energy model which needs to be corrected for better performance.

2. Correction of discomfort level of the indoor environment is improving the air-conditioning system's relative performance ($\Gamma_E^c < \Gamma_E$) – denoted by **(+–)**
A reduction in thermal discomfort of houses for air-conditioning systems belonging to this case is accompanied by a reduction in the value of the energy impact metric as well as a reduction in the total cooling hours of the system. The systems belonging to this case therefore have immense benefit by adjusting their setpoint to produce a more comfortable indoor environment. An adjustment of setpoint is improving the comfort level of the house, reducing the number of cooling hours of the system (and hence cost) and also bringing the energy consumption level of the system closer to that of the average system of the dataset. The "*fault*" for systems in this case is that the systems are being operated at bad operating conditions and the occupants should be notified of the possible improvements. Using the condition derived for $\Delta E_c > 0$ in equation 7.42 and the equations 7.3 and 7.8, the math for this case is examined in the following,

$$\Gamma_E^c < \Gamma_E \tag{7.50}$$

$$\implies \frac{\sum\left[(E_S - E_M) + \Delta T_{ic}\left(\beta_1^S - \beta_1^M\right)\right]}{\sum\left(E_S + \beta_1^S \Delta T_{ic}\right)} < \frac{\sum\left(E_S - E_M\right)}{\sum E_S} \tag{7.51}$$

Upon cross-multiplying and cancelling the common terms we get,

$$\implies (\beta_1^S - \beta_1^M)(\sum \Delta T_{ic})(\sum E_S) > (\sum (E_S - E_M))(\sum \Delta T_{ic})\beta_1^S \tag{7.52}$$

$$\implies \frac{\sum\left(E_S - E_M\right)}{\sum E_S} < \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.53}$$

$$\implies \Gamma_E < \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.54}$$

During the derivation because $\sum \Delta T_{ic} < 0$ (from 7.42) the sign in the above equation switches. The energy impact in this case has an upper bound similar to (+++). Since the severity index is positive ($\Gamma_E^c > 0$) and energy impact is greater than the severity index ($\Gamma_E^c < \Gamma_E$), we also can infer that the energy impact must be strictly positive ($\Gamma_E > 0$). Putting this together with equation 7.54 we get,

$$0 < \Gamma_E < \frac{\beta_1^S - \beta_1^M}{\beta_1^S} \tag{7.55}$$

$$\implies \frac{\beta_1^S - \beta_1^M}{\beta_1^S} > 0 \tag{7.56}$$

$$\implies \beta_1^S > \beta_1^M \tag{7.57}$$

Therefore, in this case too the mathematical condition is the same as in (+++). Equation 7.57 implies that the slope of the energy consumption model of air-conditioning systems lying in this case is greater than the slope of energy consumption model of the average system of the dataset in the climate region. Hence, we also can expect that house the system is installed in this case will have lower insulation levels and/or higher infiltration rates than the average house. The similarity in the properties of system in (+++) and system in (+–) seems counter-intuitive. While (+++) represents systems that very likely possess a mechanical fault, systems in (+–) however, are systems that may not have a mechanical fault but are definitely being *operated at bad operating conditions*. Figure 7.5 provides further evidence for our conclusions as we notice that the slope of models of all the systems falling in this category is greater than that of the average system.

7.2.2 Air-conditioning systems with negative fault severity index ($\Gamma_E^c < 0$) – denoted by (-)

If the value of the fault severity index is negative it implies that the air-conditioning system consumes less energy than the average air-conditioning system to produce the same completely comfortable environment. In this case, there is a lot of incentive to notify the occupant of the possible improvement in comfort which sometimes may come at the cost of extra energy consumption or may not. Since the system will consume less energy than the average system upon correction of thermal discomfort is a given, the decision to adjust the setpoint to a comfortable value is primar-

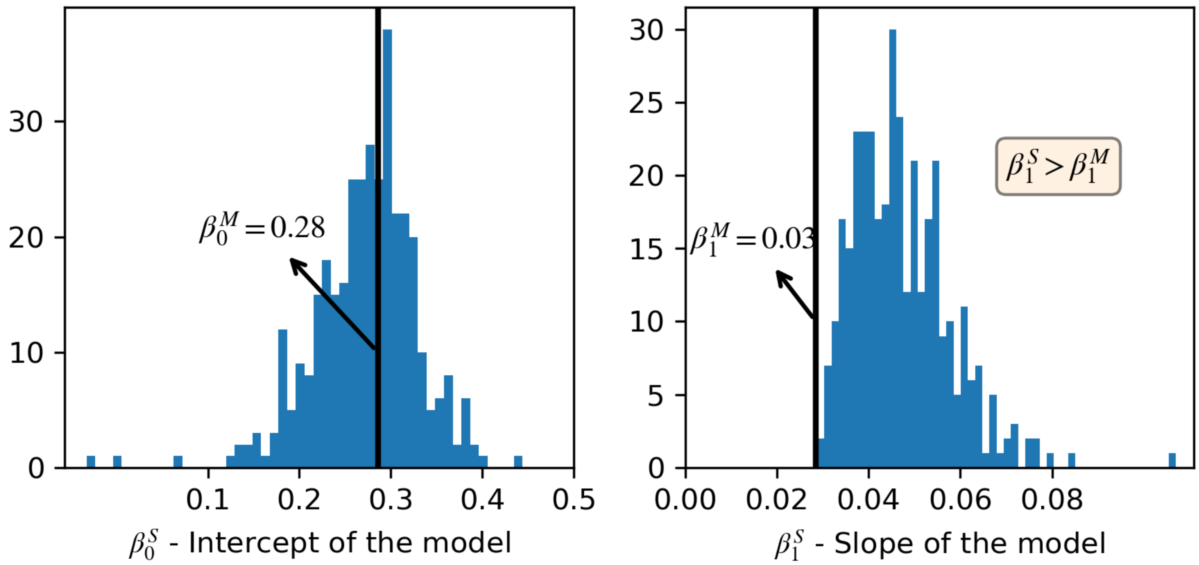


Figure 7.5: Histogram plot of intercept and slopes of models of all air-conditioning systems wherein the house is over-cooled; upon correction of which will reduce the gap between the energy consumption of the system and the average system. Systems are being operated at inefficient operating conditions.

ily dependent on whether or not this adjustment gives rise to an agreeable change in total energy consumption of the system. If the extra cooling hours which will come at the cost of more energy consumed is acceptable to the occupant, then they should be urged to correct their ineffective set-points. The difference in cases that arises because of increase or decrease of the value of the energy impact metric isn't too relevant here because eventually the system will consume less energy than the average system. Hence, there are only two sub-cases for systems in this case. Mathematically, this will be expressed as the inverse of equation 7.15,

$$\Gamma_E \sum E_S^i + \sum \Delta T_{ic}^i (\beta_1^S - \beta_1^M)] < 0 \tag{7.58}$$

7.2.2.1 Air-conditioning systems that consume more energy when they are made comfortable ($\Delta E_c > 0$) – denoted by **(-+)**

Since, making all the periods of operation of the air-conditioning system thermally comfortable, in this case, causes an increase in total cooling hours of the system, though the number of cooling hours is still less than how much the average system would have consumed, different occupants will have different preferences on whether or not the amount of extra energy consumed is acceptable. Additionally, the occupant can choose as to how much of improvement in comfort is acceptable as per the amount of extra energy consumed. As obtained in equation 7.18 systems belonging to this case are at hotter temperatures than comfortable and hence need to be corrected, which can be done by reducing the temperature. A reduction in indoor temperature, however, comes at the cost of extra energy consumption.

Since, systems in this case contain both systems wherein the fault severity index is greater than the energy impact and severity index is less than the energy impact, further mathematical exploration cannot be done without splitting further. Since system falling in this broad category are "better" than the average system, splitting this into further cases will not help in improving our understanding of the system behavior. Figure 7.6 shows the histogram plot of the slope and intercept of the energy consumption models of systems belonging to this case in comparison to the those of the average system of the dataset. 93.5% of the systems belonging to this case have their intercept lower than that of the average system. This implies that "*fault*" that is causing a negative fault severity index is the lesser internal load of the system found on average.

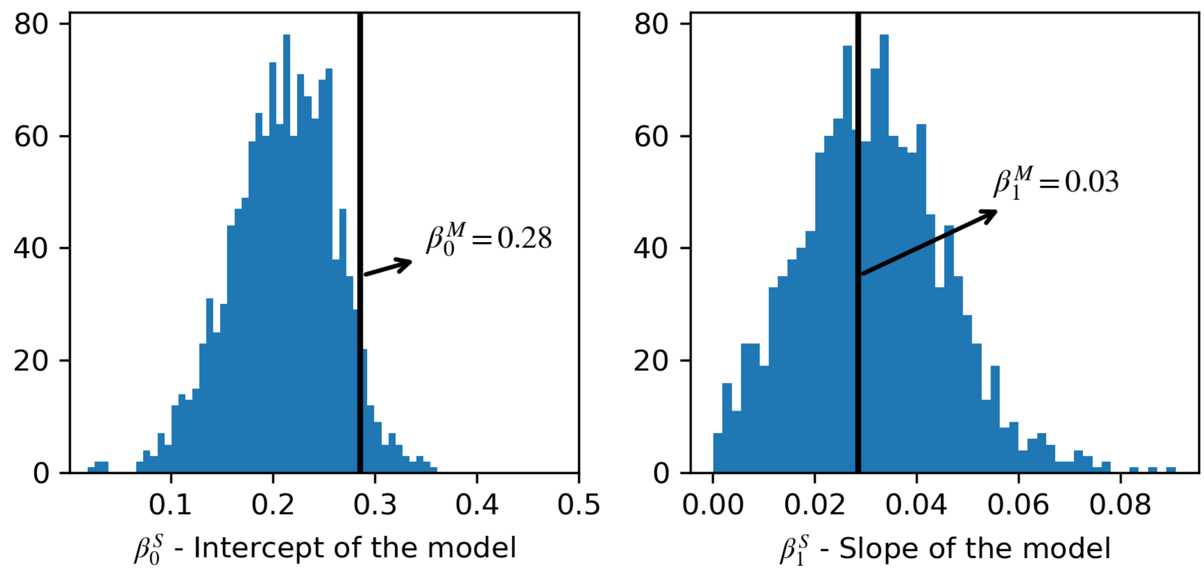


Figure 7.6: Histogram plot of intercept and slopes of models of all air-conditioning systems that will consume lesser energy than the average system upon correction of thermal discomfort but more than what they were consuming at the original level of comfort.

7.2.3 Air-conditioning systems that consume less energy upon becoming comfortable ($\Delta E_c < 0$) – denoted by (–)

This is the ideal case. The value of the energy impact metric of the air-conditioning system is lower than the fault severity index and so is the amount of extra energy required to make the system comfortable. The occupants can experience greater comfort by just adjusting their setpoint to an optimal value. As described in the equation 7.42 the indoor temperatures of the system need to be raised to a higher value so that systems become comfortable. An increase in indoor temperature will come at benefit of lesser energy consumed by the system as lesser cooling will be required.

Figure 7.7 below is the histogram plot of slopes and intercepts of all the systems belonging to this dataset and a common feature with the (-+) case is that 85.5% of the system have a lower value of intercept than the average system of the dataset. The lower internal load value again contributes heavily to the negative value of fault severity index. The reduction in total cooling hours is a consequence of the system being operated at colder than comfortable setpoints as opposed to hotter than comfortable in the (-+) case.

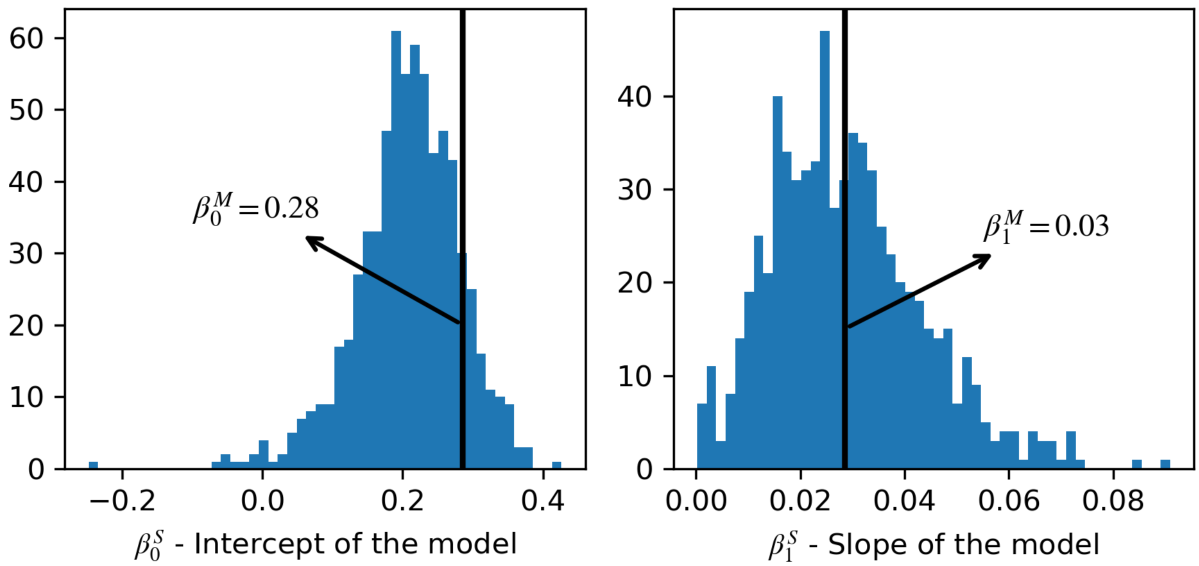


Figure 7.7: Histogram plot of intercept and slopes of models of all air-conditioning systems that are over-cooled but upon correction will consume less energy than the average system would have. Systems are being operated at inefficient operating conditions.

7.3 Summary of the outlined cases

Each fault has an impact on the air-conditioning system and its occupants in complex ways. The metrics developed in this thesis help delineate some of the intertwined relationships between faults and characteristics of the system. In order to aid this process, the air-conditioning systems were divided in to categories based on the values of the metrics. The systems were first divided into two categories based on the sign of the fault severity index. Second, systems in each case were further divided into two categories based on whether they consumed extra energy to become comfortable or if they saved energy by becoming comfortable. Finally, the systems were subsequently divided into two cases that depict how the relative energy consumption value changes when the indoor environments of the houses become comfortable. Therefore, the flowchart shows all the cases into which the air-conditioning systems with a low to moderate value of fault severity index in the dataset are divided, based on the cost & effect of reduction of discomfort,

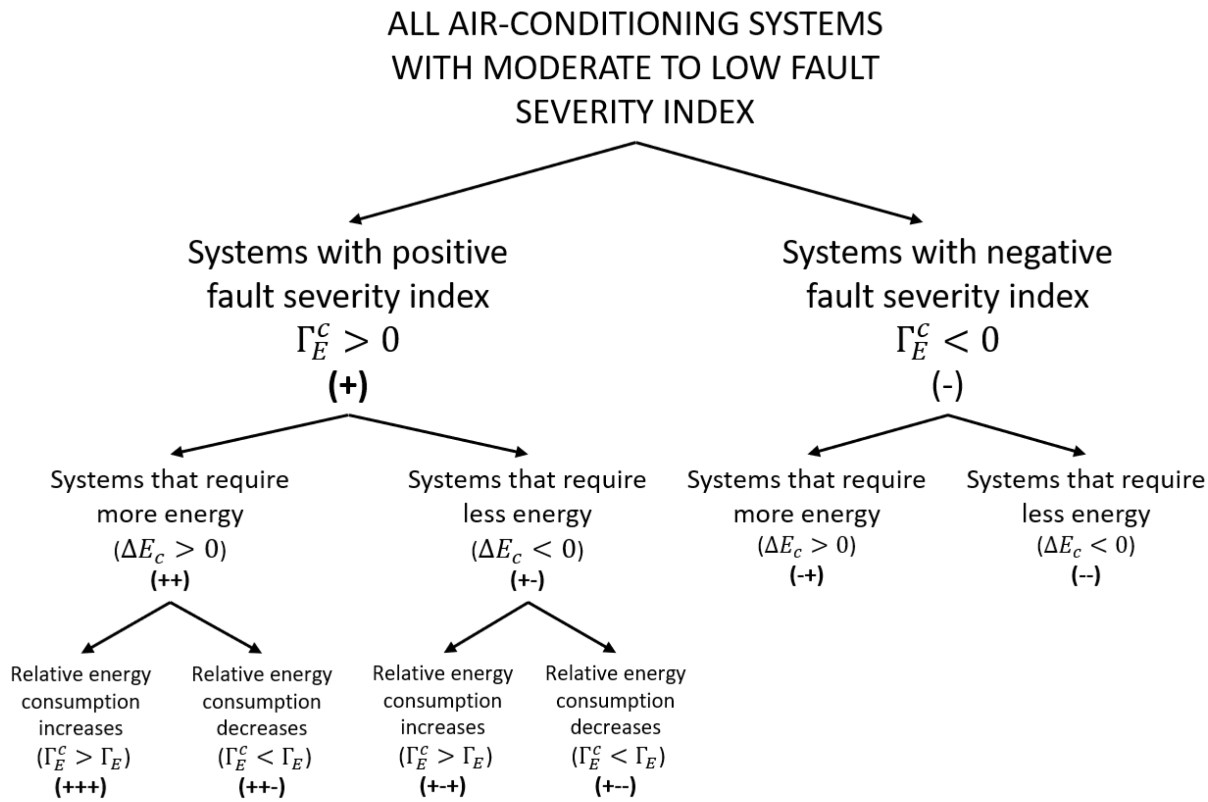


Figure 7.8: Flowchart showing the various cases into which air-conditioning systems with low to moderate fault severity index are sorted into.

In order to make it easier to visualize the systems on a graphical plot, each individual case was allotted a region based on the value of metric they show. Therefore, upon converting the flowchart above into a graphical plot, the following is obtained.

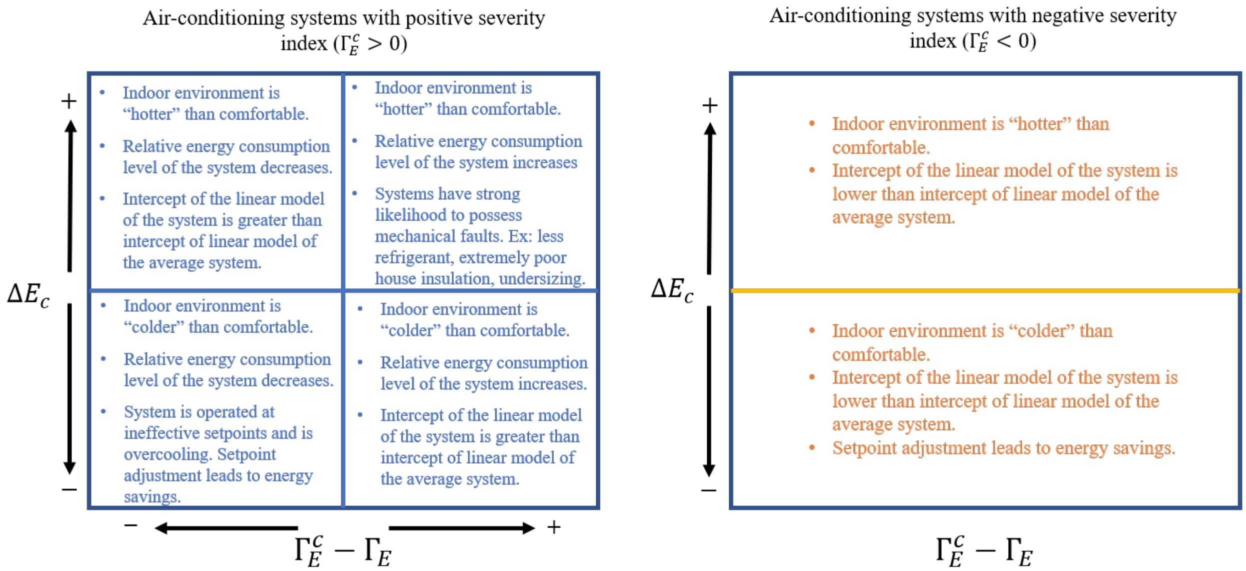


Figure 7.9: Graphical representation of the flowchart showing what happens to the air-conditioning systems upon improvement of comfort level along with few characteristics of systems in each case.

The bubble plot shown in fig. 7.10 plots few of the systems in the format of the graphical plot shown in fig. 7.9. The plot shows the variation of percentage change in total cooling hours (ΔE_c) when all the periods are made completely comfortable against the difference in the value of the energy impact metric ($\Gamma_E^c - \Gamma_E$) after and before making all the time periods comfortable. The size of the bubble indicates the value of fault severity index i.e., the Γ_E^c. The bubbles on the blue plot represent systems with positive fault severity index ($\Gamma_E^c > 0$) and the bubbles in orange represent systems with negative fault severity index ($\Gamma_E^c < 0$). The larger the bubbles in blue & the smaller the bubbles in orange the more severe is the impact of the fault.

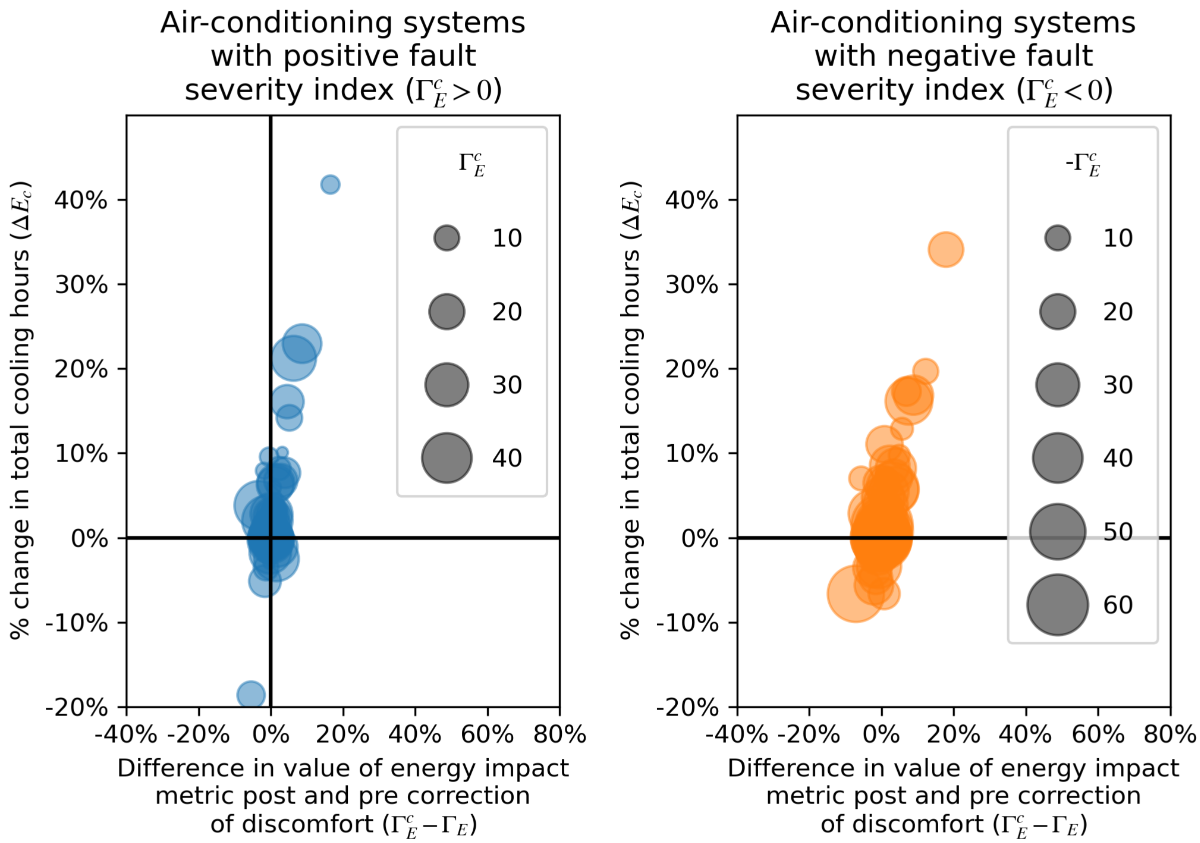


Figure 7.10: Bubble plot showing metrics of 100 systems from the dataset in the same format as the graphical plot in fig. 7.9

Several conclusions have been drawn regarding the behavior of systems in various cases a summary of which is given below:

1. The fault severity index of an air-conditioning system measures the severity of the impact of a fault on the relative energy consumption of the system provided it produces no thermal discomfort to the occupants. It can be used as a metric to rank systems objectively to prioritize them for repair. Systems with with very high value of fault severity index need to be identified and examined separately. Systems with moderate to low value of fault severity index can be examined by dividing them into various cases.

2. The impact metrics of the air-conditioning system measure the cost of reducing the thermal discomfort of the house (ΔE_c) and the effect it has on the relative energy consumption ($\Gamma_E^c - \Gamma_E$). Based on the cost and effect systems are divided into several categories.

3. Air-conditioning systems belonging to the case (+++) show a strong likelihood to possess mechanical faults. These faults could include: reduction in amount of refrigerant, system that was sized incorrectly, houses having obstruction in air-vents, very poor insulation, extremely high solar/internal load etc. All the systems in this case will operate in houses with lower level of insulation and/or higher infiltration rate on average (average here implies normalized by system capacity)than the average house of the dataset. The occupants need to be alerted of the possible faults in their system.

4. Air-conditioning systems belonging to the case (++-) have known to show greater value static load (which in this case is a combination of internal and solar loads i.e., the intercept of the energy consumption model) on average than the average system. This is causing a high value of energy impact metric and correction of thermal comfort level will bring the energy consumption level of the system closer to the average but it comes at the cost of extra energy consumed. The occupant should be notified of the high static load of their system and about the cost of improvement in comfort level.

5. Air-conditioning systems belonging to the case (+-+) have also known to show greater static load than the average system of the dataset where again the static load in this implies a high value of intercept in the energy model. However, they are being operated at colder than comfortable temperatures and so reduction of thermal discomfort will yield savings in energy at the cost of pushing their energy consumption level further away from the average system. The occupant should again be notified of the high value of intercept of their house, the reduction in which will yield significant amount of savings in energy and relative system performance.

6. Air-conditioning systems belonging to the (+–) case are systems that are being operated at ineffective operating conditions. The occupant has a lot of incentive to improve the comfort level of the house as it will benefit in energy savings as well as help bring the energy consumption level of the system closer to that of the average system. Systems belonging to this case also will operate in houses with a lower level of insulation and/or higher infiltration rate on average (essentially the slope of the energy consumption model) than the average house of the dataset, but by operating at cold setpoint temperatures and thereby overcooling their house they end up using a lot of extra energy.

7. Air-condtioning systems with a negative value of fault severity index (cases:(-+) & (–)) generally have a value of the intercept lower than that of the average system. The lower static load helps immensely in keeping the energy consumption level of the system lower than the average system at the proposed level of comfort. The improvement in comfort can come at the cost of extra energy consumption or with savings in energy. The latter represents the cases where the systems are being operated and ineffective setpoints.

8. CASE STUDIES

Chapter 7 outlined the various categories into which air-conditioning systems with a moderate to low fault severity index can belong. The flowchart constructed is given below:

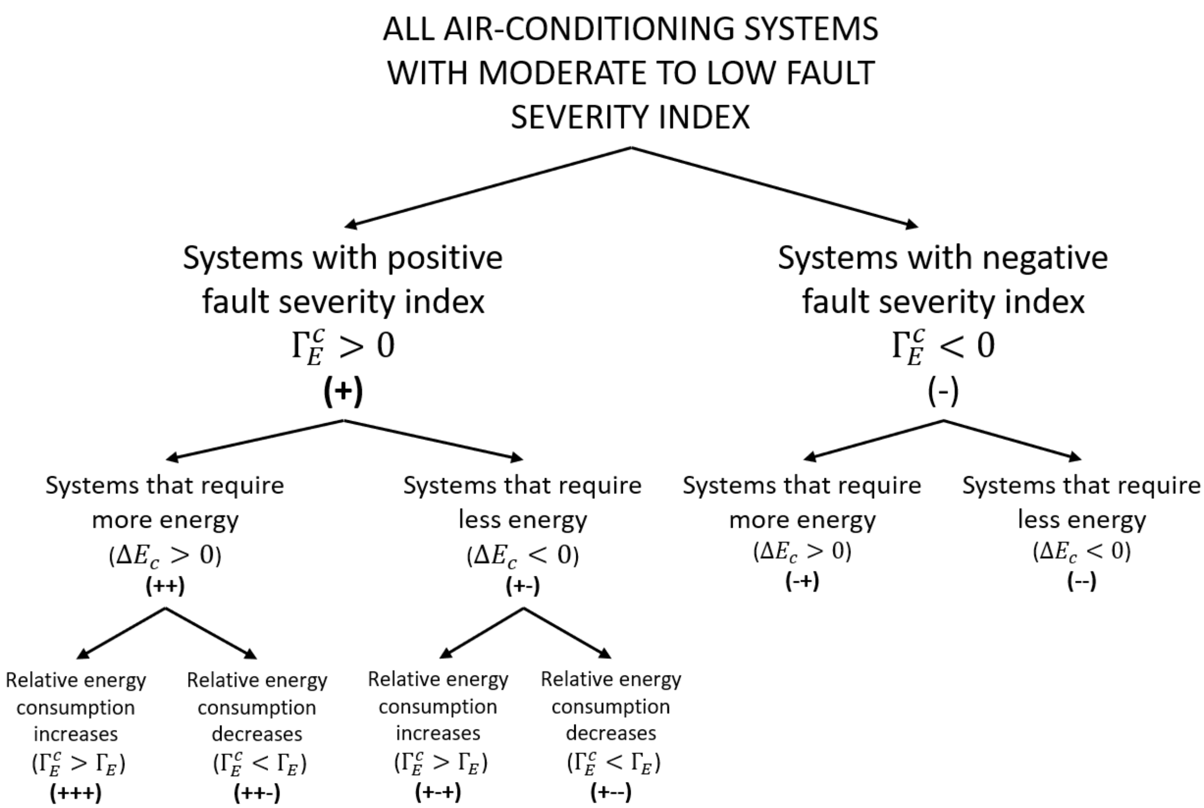


Figure 8.1: Flowchart showing the various cases into which air-conditioning systems with low to moderate fault severity index are sorted into.

Although the previous chapter provides a good description of each case, examining interesting systems in each case can provide further understanding as well as an opportunity to observe the behavior of systems quantitatively rather than qualitatively. Therefore, this chapter chooses to examine an air-conditioning system from each case to better understand its behavior. For each case, the system performance at the original level of comfort is plotted followed by what would happen to its performance at the proposed level of comfort which is when all the operating periods of the system are made comfortable. The reader is urged to pay careful attention to the plot because of the huge amount of information they provide and to understand the myriad of hidden details. Before the case studies are discussed few of the variable that have previously been defined are recapped here:

- E_S^i - Cooling effort of the given air-conditioning system in a 2-hour period as calculated from the energy consumption model of the system
- E_M^i - Cooling effort of the average system operating operating in the climate region in a 2-hour period as calculated from the energy consumption model of the average system
- $E_{S,c}^i$ - Cooling effort of the given air-conditioning system in a 2-hour period when the house is comfortable at all points in time
- $E_{M,c}$ - Cooling effort of the average system operating in the climate region in a 2-hour period when the house is comfortable at all points in time
- $\Delta E_{SM}^i = E_S^i - E_M^i$ - Difference between the cooling effort of the given system and cooling effort of the average system in a 2-hour periods to produce the same temperature difference between outdoor and indoor
- $\Delta E_{SM,c}^i = E_{S,c}^i - E_{M,c}^i$ - Difference between the cooling effort of the given system and cooling effort of the average system in a 2-hour period at the proposed completely comfortable levels
- Γ_I - Thermal Comfort Impact Metric
- Γ_E - Energy Impact Metric (at original level of comfort)
- Γ_E^c - Fault Severity Index (energy impact metric value when house is comfortable at all points in time)

- ΔE_c - Percentage change in cooling hours of the system when all 2-hour periods are made comfortable i.e, at the proposed completely comfortable levels

8.1 (+++): Air-conditioning systems with a positive fault severity index that require more energy to become comfortable in the process of which experience an increase in relative energy consumption ($\Gamma_E^c > 0; \Delta E_c > 0; \Gamma_E^c > \Gamma_E$)

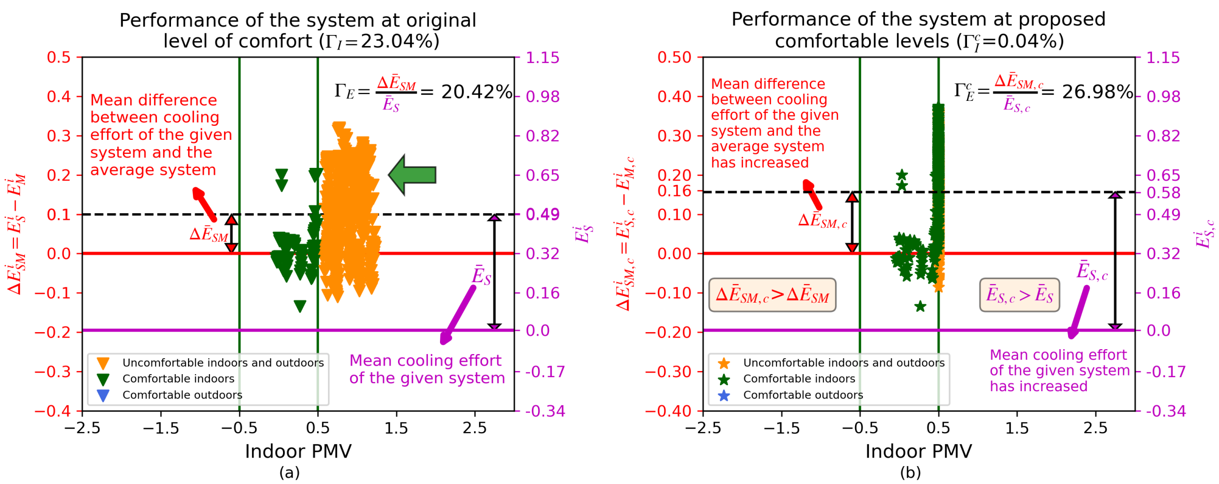


Figure 8.2: Performance of a system belonging to the (+++) case at (a) original level of comfort and (b) when all the operating periods of the system have been comfortable.

Figure 8.2 shows data from an air-conditioning system belonging to the case labelled (+++). First, at the original level of comfort the value of the thermal comfort impact metric (Γ_I) is 23.04%. This implies that on average the occupants are 23% as uncomfortable as they would have been in the absence of the air-conditioner at that geographical location. This can be evidenced by observing the points which each represent a single 2-hour period of operation where the indoors are uncomfortable in the first subplot. Owing to the concentration of 2-hour periods with indoor PMV values greater than 0.5 given by the points in orange the average discomfort indoors is greater than 0. The HVAC system is able to reduce the amount of discomfort the occupant would feel in its absence, but not fully. Hence, there is potential for even further reduction of discomfort, but at a certain cost which is depicted in detail in the second subplot.

In both of the above subplots there are two ordinates one on either side. Each data point on the scatter plot has a corresponding value on each of the axes. The axis in red represents the difference between the amount of cooling effort required for the given system and the amount that would be required by the average system of the dataset to maintain the same indoor temperature in a 2-hour period (ΔE^i_{SM}). The axis in magenta of the subplot is just the cooling effort of the system in a given 2-hour period (E^i_S). Subplot (a) shows the performance of the system at original level of comfort and subplot (b) shows what would happen if the comfort level of the house in each period was improved. According to the definition of energy impact,

$$\Gamma_E = \frac{\Sigma_{i=1}^N (E^i_S - E^i_M)}{\Sigma_{i=1}^N E^i_S} \tag{8.1}$$

Upon dividing the numerator and denominator by the total number of 2-hour periods, N the following is obtained:

$$\Gamma_E = \frac{\Delta \bar{E}_{SM}}{\Delta \bar{E}_S} \tag{8.2}$$

The value of the energy impact metric of the system (Γ_E) as calculated from the original data is 20.42%. That implies that the average system run for 20.42% less amount of time than the given system. So, if the given air-conditioning system were to run 100 hours, then the average system of the dataset in the climate region would only run for $100 - 20.42 = 79.58\ hours$ to provide an indoor environment at the current level of comfort. Upon reducing thermal discomfort the value of the energy impact metric of the system at this new "comfortable" state of operation of the system is termed as the fault severity index of the system. Therefore, upon extending equation 8.2 for the new hypothetical state:

$$\Gamma^c_E = \frac{\Delta \bar{E}_{SM,c}}{\Delta \bar{E}_{S,c}} \tag{8.3}$$

Subplot (a) of fig. 8.2 indicates that the indoor environment of the house is uncomfortable because the periods of the system are on average hotter than the comfortable region. This is indicated by the fact all the orange points that represent each of the 2-hour periods have indoor PMV values beyond 0.5. Improvement of comfort which can be brought about by the reduction indoor temperature, requires the system to increase the amount of cooling it provides. Simultaneously, the amount of cooling required by the average system to produce this comfortable environment would also increase. This can be understood by looking at the plot of the energy consumption models of the given and average system as given in fig. 8.3. Subplot (b) of fig. 8.2 indicates that both of these values increase in such a way that the difference between the two as well as the respective ratios also increase.

$$\Delta\bar{E}_{SM,c} > \Delta\bar{E}_{SM} \tag{8.4}$$

$$\bar{E}_{S,c} > \bar{E}_S \tag{8.5}$$

The fault severity index of the air-conditioning system, Γ_E^c that represents the percentage of time the average system runs lesser than the given system, is now 26.98%. This implies that if the given system were to run for 100 hours to provide a completely comfortable indoor environment, the average system of the dataset in the climate region run for only $100 - 26.98 = 73.02\ hours$. The 6.5% increase in energy impact metric value of the system as the indoor environment goes from current level of comfort to the proposed comfortable level indicates that the thermal discomfort has an adverse effect on the system's energy consumption, which is not being captured by the value of the energy impact metric at the original level of comfort.

Finally, the last metric that gives a complete picture of the system's behavior is the percentage change in cooling hours required to make the system comfortable. Since, the improvement of comfort is accompanied by a reduction in indoor temperature, an increase in temperature difference values can be observed. This increased temperature difference causes an increase in cooling effort of the system, which is computed by the following:

$$\Delta E_c = \frac{\bar{E}_{S,c} - \bar{E}_S}{\bar{E}_S} = 19.54\% \tag{8.6}$$

Further corroboration for the values of the energy impact metric, fault severity index and the percentage change in cooling hours can be found in the following plot which plots the energy consumption model of the given system against that of the average system. The slope of the model of the given system is greater than the slope of the model of the average system is evidence for the system's poorer than average value of (insulation + infiltration) level per unit capacity. Since, the correction of comfort level causes a general increase in temperature difference values, the points will be shifted to the right. This causes an increase in mean difference between the cooling effort of the system and the average system (because of greater slope, $\beta_1^S > \beta_1^M$) as well as an increase mean cooling effort of the given system (because of positive slope, β_1^S).

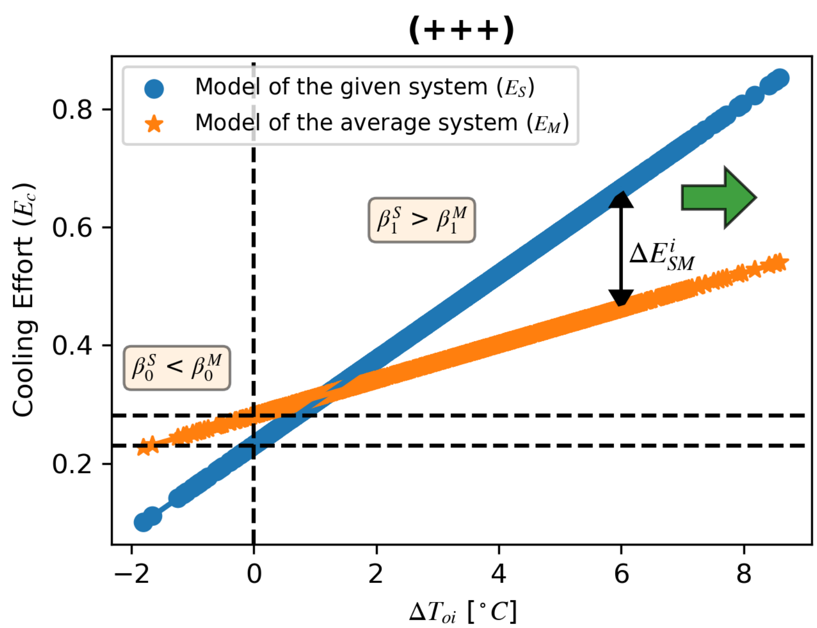


Figure 8.3: Energy consumption model of the given air-conditioning system against that of the average system of the dataset at the original level of comfort. The arrow in green shows the direction of movement of points along their respective lines as the house is made comfortable.

The fourth metric attempts to indicate that, upon improving the comfort level of the indoor environment, the total cooling hours of the system increase by close to 20% whereas the total cooling hours of the average system would increase by a lesser value. Therefore, the thermal comfort impact metric is indicating that the system is being operated at ineffective operating conditions. Upon correction of this leads to the system deviating further away from the average system implying that there perhaps might be a mechanical fault with the system itself.

8.2 (++-): Air-conditioning systems with a positive fault severity index that require more energy to become comfortable in the process of which experience a decrease in relative energy consumption ($\Gamma_E^c > 0; \Delta E_c > 0; \Gamma_E^c < \Gamma_E$)

Air-conditioning systems in this case belong to the second quadrant among the four selected for all systems with a fault severity index greater than 0. Since, by definition the total cooling hours of the system increase upon improvement of comfort in the house in this case, it indicates that indoor temperature of the house has to be reduced. The 2-hour periods of the system must be hotter than comfortable causing a discomfort which can be corrected by increase the total cooling hours of the system. Evidence for this can be seen in the plot of an example system plotted in fig. 8.4 by noticing the values of indoor PMV on subplot (a).

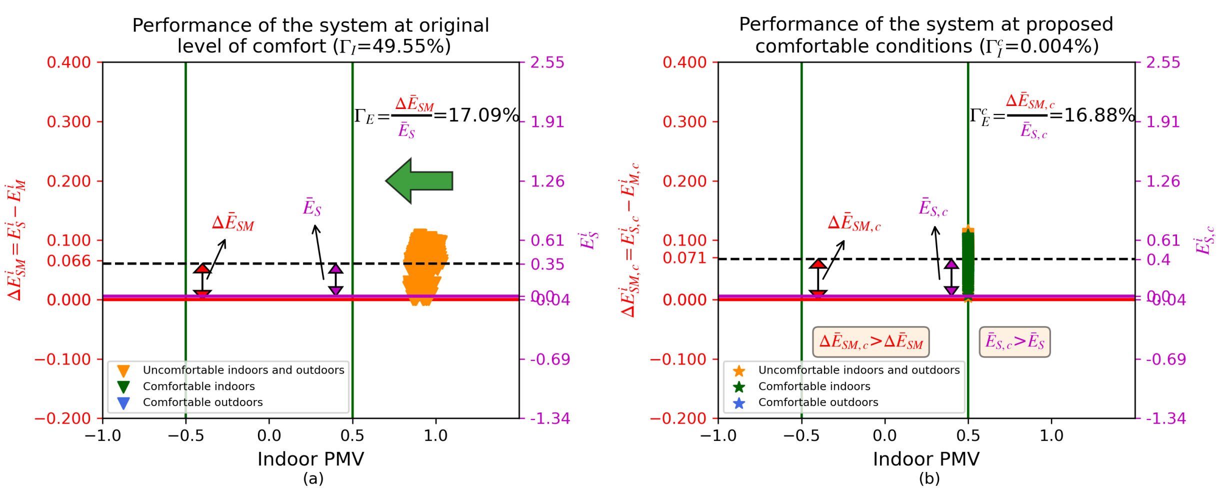


Figure 8.4: Performance of a system belonging to the (++-) case at (a) original level of comfort and (b) when all the periods of the system have been comfortable.

Although the change is very marginal there is an increase in the value of mean cooling effort of the system.

$$\bar{E}_{S,c} > \bar{E}_S \tag{8.7}$$

The value of the thermal comfort impact metric of the system indicates that the indoor environment of the house is about half as uncomfortable as the environment outside. To reduce thermal discomfort the periods of the system must be pulled back to the comfortable region and this can be done by increasing the total cooling hours of the system by $\Delta E_c = 14.36\%$. The readers must be cautious and notice that the scale of the two ordinates in 8.4 are wildly different. This is because, the difference in slopes of the model of the given system and the model of the average system is very low as shown in the fig. 8.5.

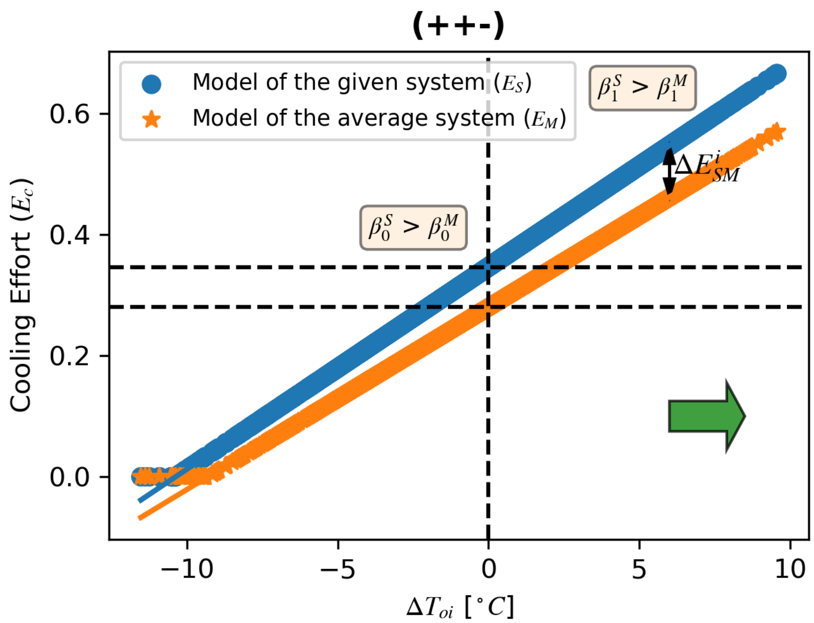


Figure 8.5: Energy consumption model of the given air-conditioning system against the model of the average system of the dataset at the original level of comfort. The arrow in green shows the direction of movement of points along their respective lines as the system is made comfortable.

Since the slopes are very close to each other, a change in the value of ΔT_{oi} will cause a much lesser change in the difference of cooling efforts of the given and average system in comparison to the change caused in the cooling effort of the given system. The plot in figure 8.5 indicate two important details: The difference between the slope of the given system and the average system is very low and the range of operation of ΔT_{oi} values are more negative than positive (fig. 8.6). The correction of level of comfort in the house as discussed above requires the increase in temperature difference values. As observed in the plot in fig. 8.5 an increase in temperature difference values will cause a marginal increase in difference between the cooling effort of the given system and the mean system. This is further corroborated in the plot in fig. 8.4

$$\Delta \bar{E}_{SM,c} > \Delta \bar{E}_{SM} \tag{8.8}$$

Although both the numerator and denominator increase, the energy impact metric of the air-conditioning system has decreased implying that the system's energy consumption level gets closer to that of the average system. Although the slope of the model of the given system is greater than that of the average system even in case of (+++) and (+–), the reason why this system behaves differently can be attributed to the following reasons: reducing thermal discomfort requires reduction of indoor temperature, the difference between the two slopes is very low, the intercept of the model of the system is greater than that of the average system and finally the range of operation of ΔT_{oi} values.

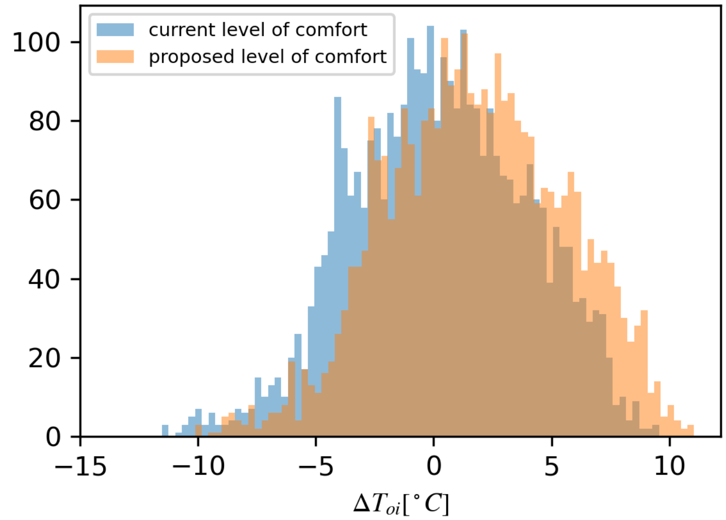


Figure 8.6: Histogram of temperature difference values shows that the range of operation of temperature difference value at original level of comfort is lower than the range observed at the comfortable level. Also notice the high concentration of values close to 0 for the original level of comfort (81$\%$ of values between -5$^\circ C$ and 5$^\circ C$). This is peculiar system behavior because it means that for many periods of operation the mean indoor and outdoor temperatures are approximately equal.

8.3 (+-+): Air-conditioning systems with a positive fault severity index that require less energy to become comfortable in the process of which experience a increase in relative energy consumption ($\Gamma_E^c > 0; \Delta E_c < 0; \Gamma_E^c > \Gamma_E$)

Air-conditioning systems in this case experience an increase in the value of the energy impact metric when the comfort level is improved but benefit from savings in energy. Since a reduction in total number of cooling hours was observed upon improvement of thermal comfort, the temperature difference values in general must have decreased. Such a decrease is only possible with the increase in indoor temperature. Since, the improvement of thermal comfort is accompanied by an increase in indoor temperature, one can conclude that the 2-hour periods of the system are cooler than comfortable. As in all the cases above consider the performance of the system given in the figure below:

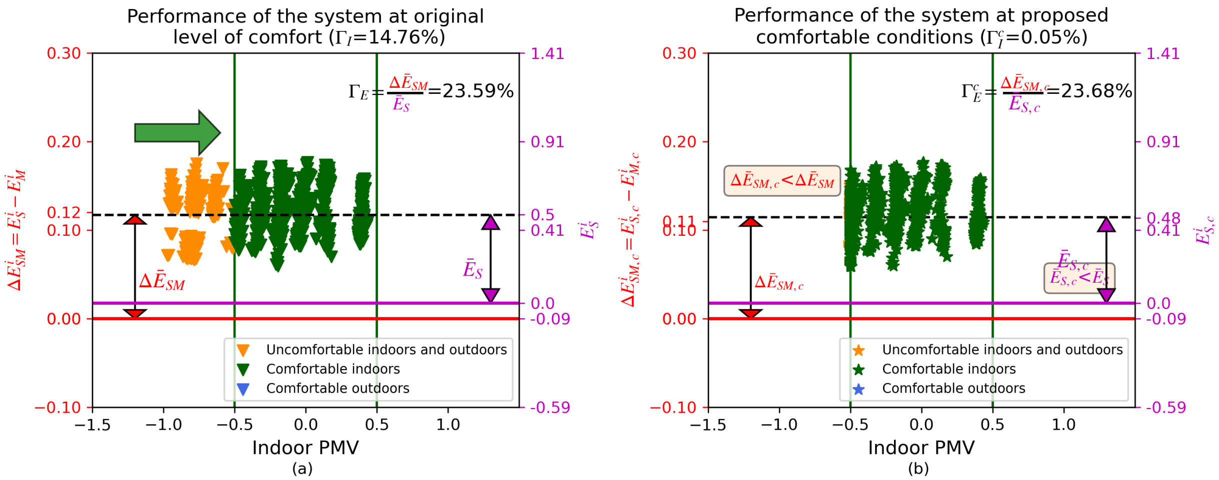


Figure 8.7: Performance of a system belonging to the (+-+) case at (a) original level of comfort and (b) when all the periods of the system have been comfortable.

As expected, a concentration of orange data points is observed with their indoor PMV value less than -0.5 implying that all these periods are colder than comfortable. The mean cooling effort of the system has decreased upon bring all the orange points to the edge of the comfortable region implying a reduction in total cooling hours as shown in 8.7.

$$\bar{E}_{S,c} < \bar{E}_S \tag{8.9}$$

The total number of cooling hours of the air-conditioning system reduce by $\Delta E_c = 2.49\%$. In order to examine the behavior of the difference between the cooling effort of the system and that of the average system, the models of both systems are plotted against each other in fig 8.8. Similar to the system in the (++-) case, the slope of the model of the given system is marginally greater than that of the average system of the dataset. The difference between the cooling effort of the given system and the cooling effort of the average system is estimated by the distance between each corresponding point on the two models. As the values of temperature difference decrease, the points will get shifted to the left causing a decrease in the mean difference between the cooling effort of the given and the average system. This is corroborated by the plot in fig. 8.7,

$$\Delta \bar{E}_{SM,c} < \Delta \bar{E}_{SM} \tag{8.10}$$

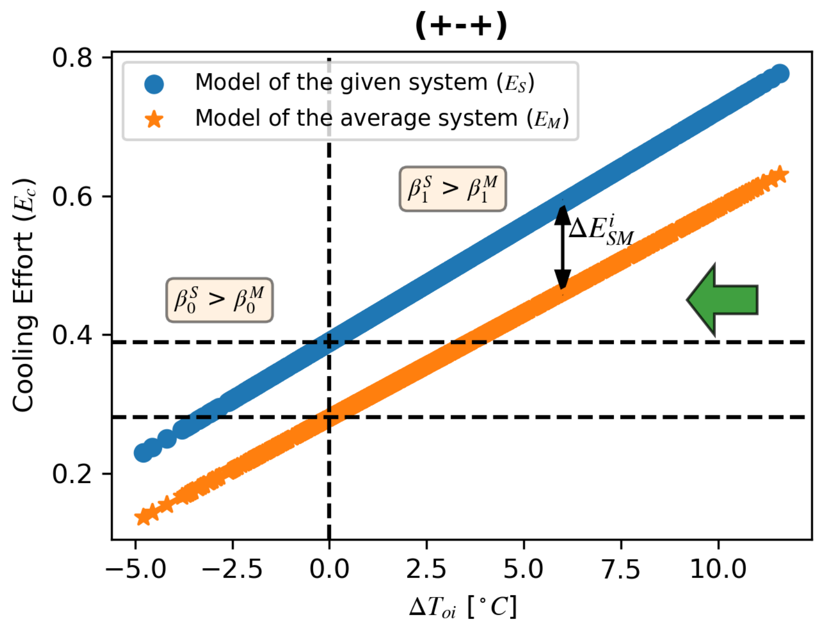


Figure 8.8: Energy consumption model of the given air-conditioning system against the model of the average system of the dataset at the original level of comfort. The arrow in green shows the direction of movement of points along their respective lines as the system is made comfortable.

Although both the numerator and the denominator decrease (contrary to the case (++-)) the energy impact value has however increased from 23.59% to 23.68%. Therefore, the correction of thermal discomfort is causing the system's energy consumption level to move farther away from the average system, which can be attributed to marginal difference in slopes, the positive difference in intercepts and the current level of comfort which requires a temperature increase for correction. Air-conditioning systems falling in this category do not necessarily have a mechanical fault because they see further deviation from the average system but it is likely. While occupants will find it beneficial to improve their comfort level because of the gains in comfort and energy, manufacturers are urged to keep a watchful eye on the systems' behavior to understand the reasons for its deterioration.

8.4 (+–): Air-conditioning systems with a positive fault severity index that require less energy to become comfortable in the process of which experience a decrease in relative energy consumption ($\Gamma_E^c > 0; \Delta E_c < 0; \Gamma_E^c < \Gamma_E$)

This case is the exact opposite of the (+++) case as indicated in fig. 8.9 below:

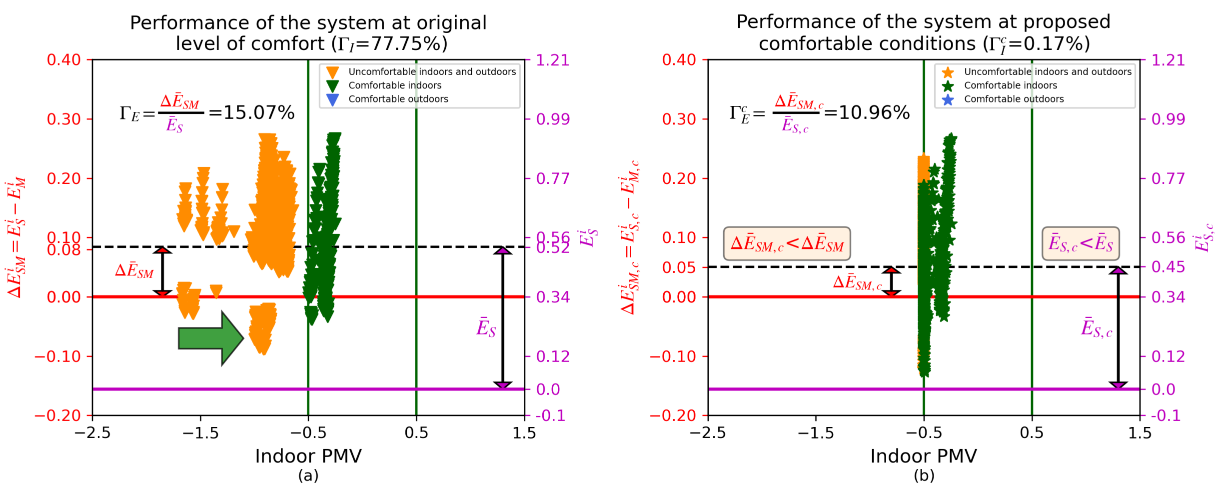


Figure 8.9: Performance of a system belonging to the (+–) case at (a) original level of comfort and (b) when all the periods of the system have been comfortable.

The occupants of the house in which the above air-conditioning system is installed experience an average of 77.75% the amount of discomfort as compared to the amount of discomfort they would have experienced if there were no air-conditioning system ($\Gamma_I = 77.75\%$). The data points in orange indicate that the discomfort caused is because a good concentration of 2-hour periods of the system indicate an indoor PMV value of less than -0.5, i.e., the house on average is colder than normal.

In order to reduce the discomfort felt by the occupants the value of indoor temperature must be raised which can be done by reducing the amount of time the cooling system runs. A reduction in total cooling time is tantamount to a reduction in mean cooling effort of the system, which is indicated in fig. 8.9. Notice that the value of mean cooling effort decreases at the new level of comfort:

$$\bar{E}_{S,c} < \bar{E}_S \tag{8.11}$$

Since an increase in indoor temperature causes a reduction in the values of temperature difference on average, the mean cooling effort of the given system as well as the average system decreases. In order to examine the change in their difference, a plot of the models represented by both systems is given below:

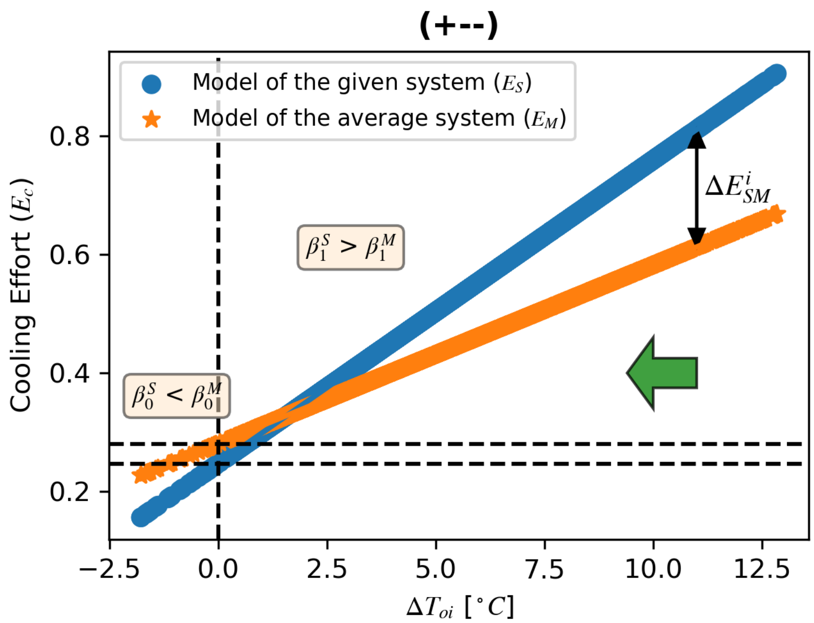


Figure 8.10: Energy consumption model of the given air-conditioning system against the model of the average system of the dataset at the original level of comfort. The arrow in green shows the direction of movement of points along their respective lines as the system is made comfortable.

Similar to the system in (+++), the slope of the energy consumption model of the system in this case too is greater than the that of the average system. Here a reduction in temperature difference values causes the points to be shifted to the left which implies that the difference between the two cooling efforts decreases. This is further corroborated in the plot in fig. 8.9:

$$\Delta\bar{E}_{SM,c} < \Delta\bar{E}_{SM} \tag{8.12}$$

Combining both of these effects the energy impact metric of the system reduces from 15.07% at the original level of comfort to 10.96% at the proposed level of comfort. This indicates that energy consumption level of the system is closer to the average system than what was initially indicated at the original level of comfort. This must because the occupant is operating their system at an ineffective operating condition. The occupant has set his setpoint at a temperature that is lower than what would be comfortable and that's causing extra consumption of energy in comparison to the average system and in general too. The % change in total cooling hours of the system is indicated by the value of ΔE_c which is,

$$\Delta E_c = \frac{\bar{E}_{S,c} - \bar{E}_S}{\bar{E}_S} = -14.02\% \tag{8.13}$$

Therefore, adjustment of setpoint temperature is very beneficial to the occupant because it will reduce the energy consumed by the system by 14.02% and will also cause the performance to shift closer to the average system of the dataset.

8.5 (-+): Air-conditioning systems with a negative fault severity index that require more energy to become comfortable ($\Gamma_E^c < 0; \Delta E_c > 0$)

All the air-conditioning systems that have a negative value of fault severity index, essentially consume lesser energy than the average system all the while providing a completely comfortable indoor environment. However, systems falling in (-+) case require more energy to make their indoor environment comfortable because systems in this case will require a reduction in indoor temperature. The plot depicting the system performance below shows why,

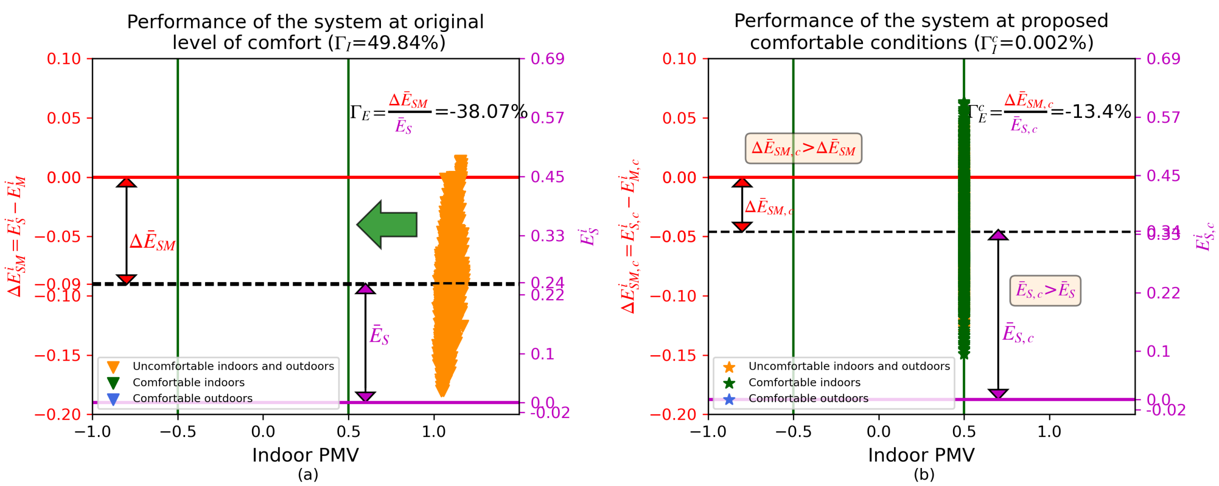


Figure 8.11: Performance of an air-conditioning system belonging to the (-+) case at (a) original level of comfort and (b) when all the periods of the system have been comfortable.

A huge concentration of orange points is observed have an indoor PMV value more than 0.5. In order to make these periods comfortable the PMV value must be pulled back to at least the boundary of the comfortable region. This can be done by reducing the indoor temperature. As concluded in the previous cases the reduction in indoor temperature on average is only possible with the increase in energy consumption of the AC. The increase in energy consumption can be observed in the plot by comparing the mean value of cooling effort of the system in the both cases.

$$\bar{E}_{S,c} > \bar{E}_S \tag{8.14}$$

In order to examine how the difference in cooling effort of the given system and the mean system changes, the energy consumption model of the given system is plotted against the model of the average system of the dataset in 8.12

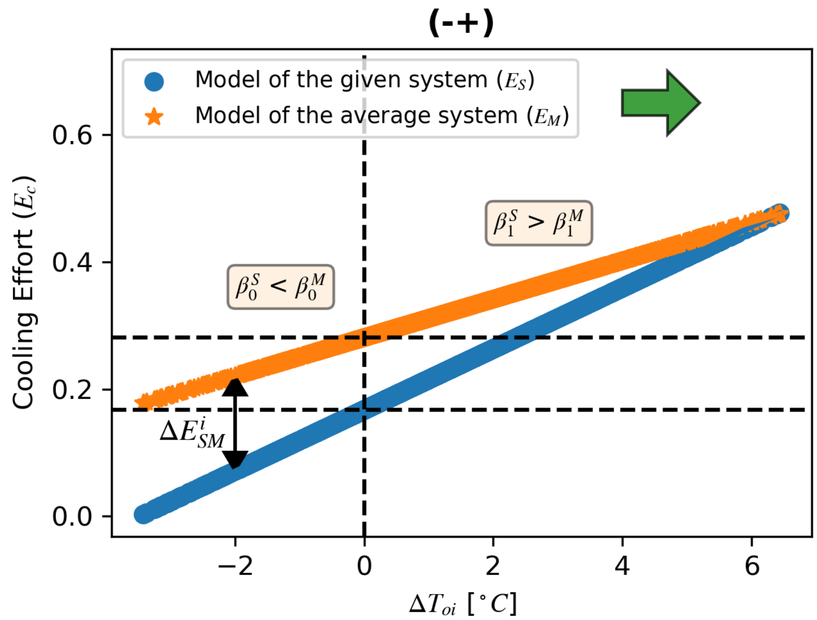


Figure 8.12: Energy consumption model of the given air-conditioning system against the model of the average system of the dataset at the original level of comfort. The arrow in green shows the direction of movement of points along their respective lines as the system is made comfortable.

As noted in chapter 7 majority of the air-conditioning systems belonging to this case have lower intercept than the average system. Therefore, as observed in 8.12 when the values of temperature difference increase, the points will shift to the right, thereby bringing the points on the two models closer to each other. That means that the magnitude of difference in energy consumption levels between two systems gets smaller. However, since the mean difference is negative originally a decrease in its magnitude actually implies an increase in mean difference. This is corroborated by the fig. 8.11 wherein,

$$|\Delta \bar{E}_{SM,c}| < |\Delta \bar{E}_{SM}| \tag{8.15}$$

Therefore, an increase in the value of mean cooling effort and decrease in magnitude of mean difference causes the value of the energy impact metric to decrease in magnitude and the system gets closer in performance to the average system. However, since the given air-conditioning system performs better than the average system at original levels of comfort itself, the reduction in magnitude of the energy impact metric actually means the performance of the system is getting worse. By operating the system at ineffective setpoints, the occupants were able to conserve energy and have their system consume less energy in comparison to the average system of the dataset. However, the system when operated at comfortable setpoints still consumes less energy that what the average system would have but not as less as what it was before correction. The number of cooling hours of the system are $\Delta E_c = 44.04\%$ more. Occupants based on their budget can choose by how much to reduce their thermal discomfort using this value as the maximum required to achieve fully comfortable conditions. The reader is urged to note that this particular example had an increase in energy impact metric value upon improvement of thermal comfort, however, that is not always the case. Systems in this case show the reverse trend too, wherein an improvement of thermal comfort decreases the energy impact metric further, which in most cases is more desirable. Regardless of how the energy impact metric changes, the systems in this case will always experience an increase in total cooling hours of the system and therefore, will always have a "cost" associated with comfort improvement.

8.6 (–): Air-conditioning systems with a negative fault severity index that require less energy to become comfortable ($\Gamma_E < 0; \Delta E_c < 0$)

Finally, this is the last case defined in chapter 7. When the indoor periods of the air-conditioning system are initially cooler than necessary but now have become comfortable by increase of temperature, a reduction in total cooling hours of the system in possible. This can be observed in the plot in fig. 8.13 An increase in temperature of the indoor 2-hour periods decreases the temperature difference values and thereby the total cooling hours of the system. Figure 8.13 points out that the mean cooling effort of the system decreases,

$$\bar{E}_{S,c} < \bar{E}_S \tag{8.16}$$

Figure 8.14 shows that the intercept of the energy consumption model of the system shown in 8.13 way lesser than the intercept of the model of the average system. In fact the intercept is shown to

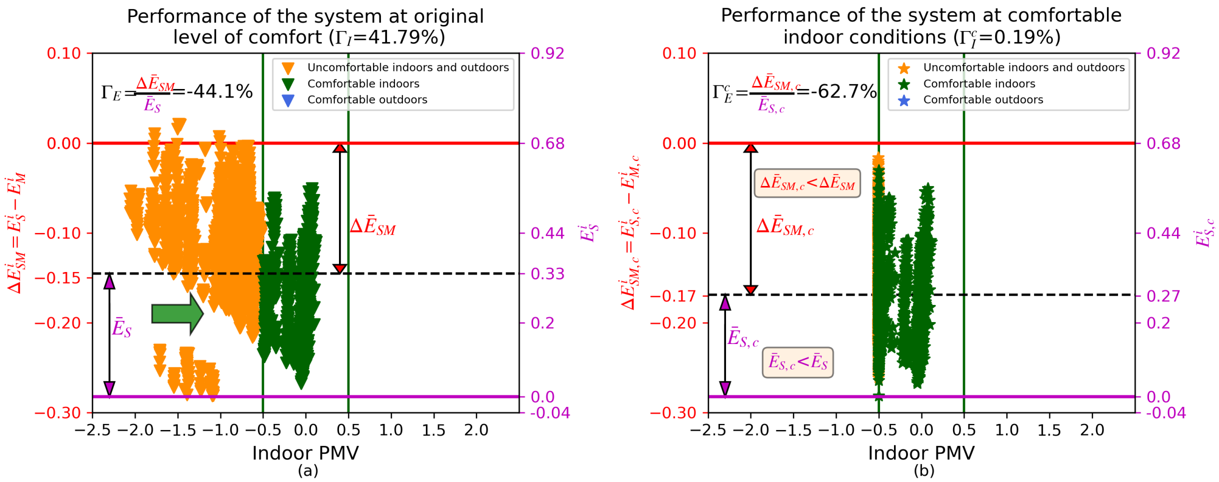


Figure 8.13: Performance of a system belonging to the (–) case at (a) original level of comfort and (b) when all the periods of the system have been comfortable.

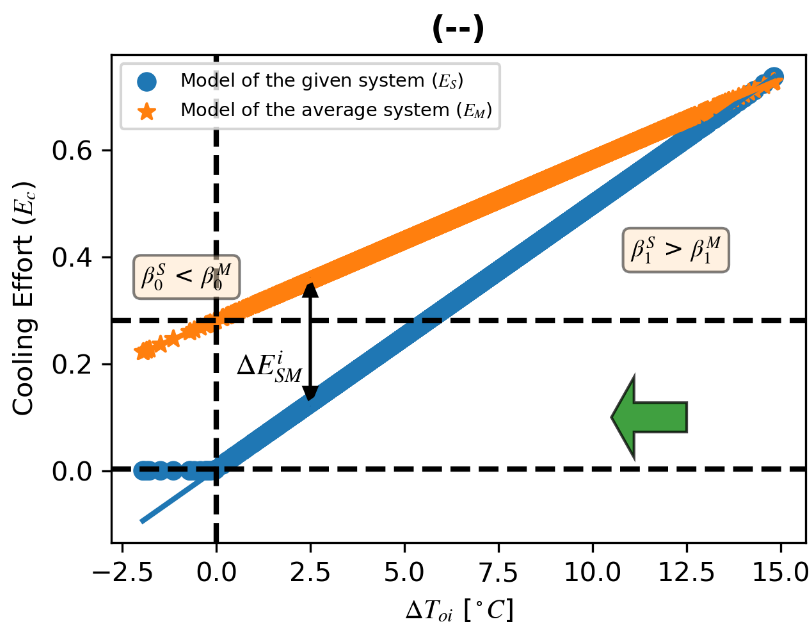


Figure 8.14: Model of the given system against the model of the average system of the dataset at the original level of comfort. The arrow in green shows the direction of movement of points along their respective lines as the system is made comfortable.

be very close to 0. And so, as the temperature difference values reach 0 and go negative, the system will have 0 cooling effort. That means that the system will stop cooling completely in the 2-hour periods when the mean indoor temperature is more than the outdoor temperature. The low internal load of the system removes the need for cooling at negative temperature difference values. The difference of cooling effort between the system and the average system will decrease (increase in magnitude but sign is negative) which can be observed in fig. 8.13

$$|\Delta\bar{E}_{SM,c}| > |\Delta\bar{E}_{SM}| \tag{8.17}$$

The reduction in both the numerator and denominator happens to be such that the value of energy impact metric also decreases as the system is made comfortable. So, the air-conditioning system's energy consumption level decreases than what it was in comparison to that of the average system. This is a very desirable scenario and is a clear case of occupant operating their system at bad operating setpoints. An adjustment will benefit the occupant in terms of actual energy saved, relative energy saved and increased comfort level. The total cooling hours of the air-conditioining system reduce by 18.27% as the average discomfort level of the indoor environment will drop from 41% to 0.19% in comparison to the average discomfort the occupant would have experienced if they were outside in the absence of the HVAC system. As in the (-+) case the systems can also undergo an increase in energy impact metric and not just a decrease, but the final value of the energy impact metric will still be negative indicating that the system consumes lesser energy than the average system of the dataset to provide the same comfortable environment. Additionally, the improvement in comfort will come at the benefit of savings in energy.

In summary, this chapter studied an air-conditioning system belonging in each case to examine its behavior with respect to its metrics. Corroboration for the values of four metrics as well as an improved understanding of the behavior of systems in each case was found. Breaking down the performance of the air-conditioning system in terms of values of its metrics and examining its behavior through the lens of these metrics provides a new insight into understanding characteristics of systems.

9. SUMMARY AND CONCLUSIONS

The adoption of smart thermostats in residential buildings and the increased aggregation of data from homes has provided an opportunity to conduct large scale data analytics for identifying air-conditioning systems with interesting behavior. The lack of huge diversity in data acquired from smart thermostats coupled with the large number of homes available for analysis has forced researchers to construct innovative ways to understand and characterize the behavior of various systems. In particular manufacturers and occupants are interested in understanding the severity of impact of a fault in the air-conditioning system. Based on the severity the manufacturer can then rank systems to prioritize them for repair and the occupant can be notified the costs associated with improvement of the system's performance. The impact of a fault in the air-conditioning system can affect its energy consumption and the thermal comfort of the occupants.

Studies in literature have attempted to build impact metrics for air-conditioning systems using smart thermostat data but a common theme among the methods that have previously been used to study the data is to compare the performance of a given air-conditioning system with a fixed baseline subjective to that system. The huge amount of data available presents an opportunity to compare the performance of a given air-conditioning system with respect to other systems operating in similarly sized houses in the same climate region. To this end the current thesis attempts to provide metrics for each air-conditioning system in a dataset of systems in the same climate region that are a description of its performance with respect to other systems as well as a description of the system's inertia to change. The thesis first explores the construction of an energy impact metric and a thermal comfort impact metric. The two impact metrics are then combined to form an aggregate fault severity index which can be used to characterize the behavior of various systems. However, before describing each of the methods, first, chapter 2 provides an overview of the research conducted by two former graduate students to give the reader a context into which this research will fit. Chapter 3 then discusses existing studies in literature and the gaps therein and proposes ideas to fill them which will be done so in the following chapters.

Chapter 4 explores the construction of an energy impact metric. An objective way to compare two air-conditioning systems relative energy consumption is by comparing the extra energy consumed by the system in comparison to what would have been consumed by systems in the dataset on average to produce the same indoor environment. This entails the construction of a baseline performance for reference which is the performance of the average air-conditioning system of the dataset. Energy consumed by an air-conditioning system can be estimated using simulation tools such as Energy Plus, but owing to their high computational expense and their requirement of unavailable metadata regarding a house they aren't viable for the current study. Therefore, this chapter proposes linear regression based method to model the energy consumption of a given system as well as the average system of the dataset. The models developed were then used to quantify the relative energy consumption level of each air-conditioning system.

Similarly, chapter 5 explores the construction of a thermal comfort impact metric by examining the average discomfort level of the house. In order to do so, first a method to calculate the discomfort level at any instant in time using the Predicted Mean Vote (PMV) value of the environment is given. A baseline environment once again was chosen for reference which was the thermal discomfort the occupant would have experienced if they were living outside without the presence of the HVAC system. By comparing the mean discomfort level in the house to the mean discomfort level at the baseline environment a thermal comfort impact metric of the house was estimated.

Subsequently, chapter 6 attempts to put both of the metrics together. This is done by firstly considering a hypothetical state of the air-conditioning system when the house is completely comfortable at all points in time. The change in indoor temperature required to reach this hypothetical state is first estimated. The value of the energy impact metric at this operating state of the system can be used to compare systems against each other objectively as all systems that have reached this state produce no thermal discomfort. Hence, the value of the energy impact metric when the house is completely comfortable is defined as the fault severity index of the air-conditioning system. Reducing the thermal discomfort to zero also provides an opportunity to calculate the change in cooling hours required to improve the comfort level of the house and hence forms the fourth and last metric that can describe the system completely.

In order to understand the relationships between the 4 metrics and what they indicate about the characteristics and behavior of system chapter 7 builds 6 categories into which all the air-conditioning systems of the dataset are divided into. The definition of each case was explored mathematically as well as empirically and interesting conclusions about the systems were drawn. Subsequently, chapter 8 is provided to quantitatively examine the air-conditioning systems in each case. An example from each case is chosen and its performance and characteristics with respect to its metrics was observed and discussed. This chapter provides the reader with a direction on how the metrics built in this study can be used to understand air-conditioning systems in residential buildings. Notably, the chapters attempt to provide a method to use the metrics to segregate systems that operate under ineffective operating conditions with systems that possess mechanical faults.

While this study presents an approach to reduce the behavior of air-conditioning systems to metrics that can be easily identified and understood, the author recognizes that several updates can be made to make this calculation more accurate. The above analysis only considers those periods where the system is trying to maintain a given setpoint and hence is in pseudo steady state. Data from transient periods of system operation should also be added to the calculation of the energy impact. Comparing the performance of the system when the cooling is turned off to the performance of the system when the system is actively cooling the house to a new setpoint will present interesting conclusions regarding the air-conditioning system and the house. When cooling is switched off the variation in indoor temperature is governed only by the characteristics of the house which can observed based on its rate of decay. Upon comparing the model so developed to the model developed when the system is actively cooling where the variation in indoor temperature depends both on the characteristics of the house and the system, faulty systems can be segregated from systems operating in poor houses. This when put together with the observations from steady-state data will help improve the diagnosis of faults in air-conditioning system.

Furthermore, the construction of the energy impact metric requires the construction of a model for the average system of the dataset as well as each given system. An approach on how to build the model was presented here but several avenues can still be explored that can lead to better results. Especially, the effectiveness of black box models should be examined and a trade-off study between using grey-box models as done here versus using black box models will also be useful. In order to improve the estimation of thermal comfort impact, future work should focus on integrating data from occupancy sensors as well as better estimations of clothing, activity level and radiant temperature that could lead to the calculation of a more accurate value the thermal comfort impact metric. Data acquisition will be challenging but with the cooperation of the occupant it can still be done a good example for which is through a mobile phone application. Furthermore, this study utilizes the heat-based thermal comfort model to predict comfort levels in indoor environments. However, with the advancement and growing popularity of personal comfort models, the estimation of comfort level can become more accurate and more occupant-specific. Although several methods can be chosen to improve the accuracy of calculation of both metrics, the basic concept and construction of metric will not change with refinement of data used to calculate it. The author believes that the analysis approach used in the study will still be relevant and can still be used an approach to understand the behavior of residential HVAC systems.

Large scale analytic methods for HVAC systems therefore can be used in innovative ways a few of which have been presented in this work. Since, the behavior of HVAC systems in residential buildings especially is very dynamic in nature and often dependent on a myriad of interdependent factors, smart thermostat data offers endless opportunities to build new perspectives to observe the relationship between the occupant and his system.

REFERENCES

[1] A. P. Rogers, "Fault detection and diagnosis methods for residential air conditioning systems using cloud-based data.," *Doctoral dissertation, Texas AM University. Available electronically from https : / /hdl .handle .net /1969 .1 /186408.*, 2019.

[2] F. Guo, "Large scale data analytics for fault detection and diagnosis of residential hvac systems," *Doctoral dissertation, Texas AM University.*, 2021.

[3] EIA, "Residential energy consumption survey (recs), energy information administration . washington, dc.," 2015.

[4] T. Peffer, M. Pritoni, A. Meier, C. Aragon, and D. Perry, "How people use thermostats in homes: A review," *Building and Environment*, vol. 46, no. 12, pp. 2529–2541, 2011.

[5] A. Meier, C. Aragon, T. Peffer, D. Perry, and M. Pritoni, "Usability of residential thermostats: Preliminary investigations," *Building and Environment*, vol. 46, no. 10, pp. 1891–1898, 2011.

[6] M. Pritoni, A. K. Meier, C. Aragon, D. Perry, and T. Peffer, "Energy efficiency and the misuse of programmable thermostats: The effectiveness of crowdsourcing for understanding household behavior," *Energy Research & Social Science*, vol. 8, pp. 190–197, 2015.

[7] Z. Pang, Y. Chen, J. Zhang, Z. O'Neill, H. Cheng, and B. Dong, "How much hvac energy could be saved from the occupant-centric smart home thermostat: A nationwide simulation study," *Applied Energy*, vol. 283, p. 116251, 2021.

[8] H. Stopps and M. F. Touchie, "Residential smart thermostat use: An exploration of thermostat programming, environmental attitudes, and the influence of smart controls on energy savings," *Energy and Buildings*, vol. 238, p. 110834, 2021.

[9] A. A. Daken, A. K. Meier, and D. G. Frazee, *Do Connected Thermostats Save Energy?* eScholarship, University of California, 2019.

[10] "Epa, us energy star ® certified smart thermostats : Fact sheet for manufacturers and service providers," . *Tech. rep. January*, 2018.

[11] B. Urban and K. Roth, "A data-driven framework for comparing residential thermostat energy performance," *Fraunhofer Center for Sustainable Energy Systems, Final Report| Co-Developed with NEST*, 2014.

[12] M. B. Kane and K. Sharma, "Data-driven identification of occupant thermostat-behavior dynamics," *arXiv preprint arXiv:1912.06705*, 2019.

[13] B. Huchuk, W. O'brien, and S. Sanner, "Exploring smart thermostat users' schedule override behaviors and the energy consequences," *Science and Technology for the Built Environment*, vol. 27, no. 2, pp. 195–210, 2020.

[14] H. Stopps and M. F. Touchie, "Managing thermal comfort in contemporary high-rise residential buildings: Using smart thermostats and surveys to identify energy efficiency and comfort opportunities," *Building and Environment*, vol. 173, p. 106748, 2020.

[15] G. Baasch, A. Wicikowski, G. Faure, and R. Evins, "Comparing gray box methods to derive building properties from smart thermostat data," in *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation*, pp. 223–232, 2019.

[16] S. A. Tabatabaei, W. Van der Ham, M. CA Klein, and J. Treur, "A data analysis technique to estimate the thermal characteristics of a house," *Energies*, vol. 10, no. 9, p. 1358, 2017.

[17] G. H. Merabet, M. Essaaidi, M. Ben-Haddou, B. Qolomany, J. Qadir, M. Anan, A. Al-Fuqaha, R. M. Abid, and D. Benhaddou, "Artificial intelligence-assisted energy and thermal comfort control for sustainable buildings: An extended representation of the systematic review," 2020.

[18] R. Lou, K. P. Hallinan, K. Huang, and T. Reissman, "Smart wifi thermostat-enabled thermal comfort control in residences," *Sustainability*, vol. 12, no. 5, p. 1919, 2020.

[19] L. Sarran, H. B. Gunay, W. O'Brien, C. A. Hviid, and C. Rode, "A data-driven study of thermostat overrides during demand response events," *Energy Policy*, vol. 153, p. 112290, 2021.

[20] A. Rogers, J. Martinez, F. Guo, and B. Rasmussen, "Labeling modes of operation and extracting features for fault detection with cloud-based thermostat data.," *ASHRAE Transactions*, vol. 126, no. 2, 2020.

[21] F. Guo and B. P. Rasmussen, "Fault detection and diagnosis for residential hvac systems using transient cloud-based thermostat data," 2021.

[22] A. P. Rogers, F. Guo, and B. P. Rasmussen, "A change point detection algorithm with application to smart thermostat data.," *ASHRAE Transactions*, vol. 126, no. 1, 2020.

[23] F. Guo and B. Rasmussen, "Detection of gradual capacity degradation for residential hvac systems using smart thermostat data," *Applied Thermal Engineering*, In Review.

[24] F. Guo, A. P. Rogers, and B. P. Rasmussen, "Multivariate fault detection for residential hvac systems using cloud-based thermostat data, part ii: Case studies," *Science and Technology for the Built Environment*, pp. 1–16, 2021.

[25] D. Popescu, S. Bienert, C. Schützenhofer, and R. Boazu, "Impact of energy efficiency measures on the economic value of buildings," *Applied Energy*, vol. 89, no. 1, pp. 454–463, 2012.

[26] S. w. Ham, P. Karava, I. Bilionis, and J. Braun, "A data-driven model for building energy normalization to enable eco-feedback in multi-family residential buildings with smart and connected technology," *Journal of Building Performance Simulation*, vol. 14, no. 4, pp. 343–365, 2021.

[27] E. E. Guillen, H. W. Samuelson, and J. G. C. Laurent, "Comparing energy and comfort metrics for building benchmarking," *Energy and Buildings*, vol. 205, p. 109539, 2019.

[28] J. Crawley, E. McKenna, V. Gori, and T. Oreszczyn, "Creating domestic building thermal performance ratings using smart meter data," *Buildings & Cities*, vol. 1, no. 1, pp. 1–13, 2020.

[29] L. Pérez-Lombard, J. Ortiz, R. González, and I. R. Maestre, "A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes," *Energy and Buildings*, vol. 41, no. 3, pp. 272–278, 2009.

[30] K. Ueno, "https://buildingscience.com/documents/digests/bsd152-building-energy-performance-metrics," *Build. Sci. Dig., 152 (2010)*.

[31] "U.s. enivronmental protection agency (epa)," *Energy Star Portfolio Manager*, 2018.

[32] P. Mathew, N. Addy, T. Walter, and USDOE, "Building performance database api (bpd api) v2.1," 2 2021.

[33] "Technical reference: Energy star score," *Energy Star*, 2018.

[34] J. D. Chambers and T. Oreszczyn, "Deconstruct: A scalable method of as-built heat power loss coefficient inference for uk dwellings using smart meter data," *Energy and Buildings*, vol. 183, pp. 443–453, 2019.

[35] E. ISO, "52016-1. energy performance of buildings-energy needs for heating and cooling, internal temperatures and sensible and latent heat loads-part 1: Calculation procedures," 2017.

[36] M. F. Fels *et al.*, "Prism: An introduction," *Energy and Buildings*, vol. 9, no. 1-2, pp. 5–18, 1986.

[37] J. K. Kissock, J. S. Haberl, and D. E. Claridge, "Inverse modeling toolkit: Numerical algorithms," *ASHRAE transactions*, vol. 109, p. 425, 2003.

[38] P. Gianniou, C. Reinhart, D. Hsu, A. Heller, and C. Rode, "Estimation of temperature setpoints and heat transfer coefficients among residential buildings in denmark based on smart meter data," *Building and Environment*, vol. 139, pp. 125–133, 2018.

[39] D. B. Crawley, L. K. Lawrie, C. O. Pedersen, and F. C. Winkelmann, "Energy plus: energy simulation program," *ASHRAE journal*, vol. 42, no. 4, pp. 49–56, 2000.

[40] C. Christensen, R. Anderson, S. Horowitz, A. Courtney, and J. Spencer, "Beopt(tm) software for building energy optimization: Features and capabilities," 8 2006.

[41] C. Booten, J. Robertson, D. Christensen, M. Heaney, D. Brown, P. Norton, and C. Smith, "Residential indoor temperature study," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2017.

[42] A. M. Bueno, A. A. de Paula Xavier, and E. E. Broday, "Evaluating the connection between thermal comfort and productivity in buildings: a systematic literature review," *Buildings*, vol. 11, no. 6, p. 244, 2021.

[43] A. Lai, K. W. Mui, L. T. Wong, and L. Law, "An evaluation model for indoor environmental quality (ieq) acceptance in residential buildings," *Energy and buildings*, vol. 41, no. 9, pp. 930–936, 2009.

[44] "American society of heating, refrigerating and air-conditioning engineers. ventilation for acceptable indoor air quality : Ashrae standard. atlanta, ga: American society of heating, refrigerating and air-conditioning engineers.,," 2007.

[45] P. O. Fanger *et al.*, "Thermal comfort. analysis and applications in environmental engineering.," *Thermal comfort. Analysis and applications in environmental engineering.*, 1970.

[46] "American society of heating, refrigerating and air-conditioning engineers. ashrae standard thermal environmental conditions for human occupancy. atlanta, ga. :the society,," 1992.

[47] A. 55, "Thermal environmental conditions for human occupancy," *Atlanta, GA:American Society of Heating, Refrigerating and Air-conditioning Engineers Inc; 2017:155.*

[48] S. Salimi, E. Estrella Guillén, and H. Samuelson, "Exceedance degree-hours: A new method for assessing long-term thermal conditions," *Indoor air*, vol. 31, no. 6, pp. 2296–2311, 2021.

[49] R. J. De Dear and G. S. Brager, "Thermal comfort in naturally ventilated buildings: revisions to ashrae standard 55," *Energy and buildings*, vol. 34, no. 6, pp. 549–561, 2002.

[50] B. Huchuk, W. O'Brien, and S. Sanner, "A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data," *Building and Environment*, vol. 139, pp. 199–210, 2018.

[51] C.-A. Roulet, F. Flourentzou, H. Labben, M. Santamouris, I. Koronaki, E. Dascalaki, and V. Richalet, "Orme: A multicriteria rating methodology for buildings," *Building and Environment*, vol. 37, no. 6, pp. 579–586, 2002.

[52] X.-f. Kong, S.-l. Lü, Y.-j. Xin, and W. Wu, "Energy consumption, indoor environmental quality, and benchmark for office buildings in hainan province of china," *Journal of central south university*, vol. 19, no. 3, pp. 783–790, 2012.

[53] L. Ciabattoni, G. Cimini, F. Ferracuti, M. Grisostomi, G. Ippoliti, and M. Pirro, "Indoor thermal comfort control through fuzzy logic pmv optimization," in *2015 International joint conference on neural networks (IJCNN)*, pp. 1–6, IEEE, 2015.

[54] T. Chaudhuri, Y. C. Soh, S. Bose, L. Xie, and H. Li, "On assuming mean radiant temperature equal to air temperature during pmv-based thermal comfort study in air-conditioned buildings," in *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, pp. 7065–7070, 2016.

[55] F. Tartarini and S. Schiavon, "pythermalcomfort: A python package for thermal comfort research," *SoftwareX*, vol. 12, p. 100578, 2020.