A Study of Building Thermal Dynamics from Large Data Sets:

An Application for Residential Smart Thermostats

Camille John

A Thesis
in
The Department
of
Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science (Building Engineering) at
Concordia University
Montreal, Quebec, Canada

August 2021

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Camille John
Entitled: A Study of Building Thermal Dynamics from Large Data Sets: An Application for Residential Smart Thermostats

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Building Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

Chair

Dr. Bruno Lee

Examiner

Dr. Bruno Lee

Examiner

Dr. Mohamed Ouf

Co-supervisor

Dr. Andreas Athienitis

Co-supervisor

Dr. José Candanedo

Approved by:

Dr. Michelle Nokken, Graduate Program Director
Department of Building, Civil and Environmental

August 30, 2021

Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

ABSTRACT

A Study of Building Thermal Dynamics from Large Data Sets:
An Application for Residential Smart Thermostats

Camille John

This thesis focuses on identifying the Thermal Time Constant (TTC), a thermal performance indicator related the building's effective thermal insulation, airtightness, and thermal storage capacity. Using data from over 15,000 smart thermostats, data mining is applied to identify patterns in the short-term transient thermal response of Canadian and American dwellings. The data used consist of contextual information (i.e. metadata) and one year of measurements recorded at 5-minute intervals (i.e. indoor air temperature, outdoor air temperature, and Heating, Ventilation and Air-Conditioning (HVAC) equipment run times). The TTC is captured from the data by tracking the indoor temperature response of the free-running dwelling over a specific time period, and by also assuming this response can be accurately described by the characteristic exponential decay (or growth) of a first-order resistance-capacitance thermal model. Consequently, the results show significant differences between estimated TTC values for the summer and winter months across ASHRAE climate zones 1 through 7. In winter, the mean TTC related to these climate zones ranges from 7 to 47 hours. In contrast, the summer mean values vary between a lower and narrower range of 6 to 19 hours which can presumably be attributed to occupants opening the windows, and thus effectively reducing their dwelling's overall thermal resistance. Towards the larger objectives of thermal resilience, energy savings and grid reliability, the estimated TTC values can be used in the residential sector to quickly identify buildings eligible for building enclosure retrofits or to rapidly generate a simple model to inform thermal load estimation and management.

Dedication

To my beloved parents,
Clover and Othello,
without whom none of my success
would be possible.

To my sister Candice,
for being my inspiration.

To the rest of my family,
for all your love and support.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors, Dr. A. K. Athienitis and Dr. J. A. Candanedo, whose expertise and encouragement added significantly to my graduate studies. In addition to the privilege of having worked with both of them, they have granted me a number of enriching experiences: the 2017 Big Data in Building Operation Workshop in Ottawa, the 2018 eSim Conference in Montreal, the 7th International Building Physics Conference in Syracuse, the 4th International Symposium on Occupant Behaviour in Ottawa, and numerous workshops under both the CZEBS Network and the NSERC/Hydro-Québec Industrial Research Chair. This thesis would not appear in its present form without their kind assistance and support.

I am also grateful to Michaël Fournier from Laboratoire des Technologies de l'Énergie (LTE) for sharing his insights and experience from working with data collected from residential thermostats.

Special thanks to my colleagues and friends in both the Concordia Solar Lab and the CanmetENERGY-Varennes Research Centre for their friendship and assistance. This experience would not have been same without the memories I shared with all of them. In particular, I wish to thank Charalampos Vallianos and Vasken Dermardiros, for the enlightening discussions we shared, the guidance they bestowed, and the helping hand they always offered.

My wholehearted thanks to my family and friends for their unwavering support and love. Particular thanks to my partner Costa, my aunt Barbara, my dearest Daphne, Erika, Audrey, and Jessica. I am forever grateful.

Financial support of this work was provided by both the NSERC/Hydro-Québec Industrial Research Chair held by Dr. Athienitis and the NSERC Discovery Grant held by Dr. Candanedo. I would also like to gratefully acknowledge ecobee, a Canadian home automation company, for providing the data used in this study.

Finally, I would like to thank my defence committee for their time and their constructive feedback.

Contents

List of Figures

List of Tables

List of Abbreviations

ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers.
BAS	Building Automation System.
BEM	Building Energy Modeling.
CRISP-DM	Cross-Industry Standard Process for Data Mining.
DM	Data Mining.
DR	Demand Response.
DSM	Demand Side Management.
DYD	Donate Your Data.
HVAC	Heating, Ventilation and Air-Conditioning.
IoT	Internet of Things.
IQR	Interquartile Range.
MPC	Model-based Predictive Control.
MT	Manual Thermostat.
PDF	Probability Density Function.
PT	Programmable Thermostat.
RC	Resistance-Capacitance.
RMSE	Root-Mean-Square Error.
ROM	Reduced-Order Model.

ST Smart Thermostat.

TTC Thermal Time Constant.

List of Symbols

A	total heat transfer surface area of dwelling	m^2
a	shape parameter for Johnson SB PDF	
ACH	number of air changes per hour	
b	shape parameter for Johnson SB PDF	
C	thermal capacitance	$J \cdot K^{-1}$
C_{eff}	effective thermal capacitance	$J \cdot kg^{-1} \cdot K^{-1}$
C_{in}	overall effective thermal capacitance of dwelling	$J \cdot kg^{-1} \cdot K^{-1}$
$c_{\text{p,air}}$	specific heat capacity of air	$J \cdot kg^{-1} \cdot K^{-1}$
E_{i}	total internal energy of dwelling	J
IQR	Interquartile Range	
loc	parameter used to shift the Johnson SB PDF	
m	variable of Johnson SB PDF used to incorporate parameters loc and $scale$	
$\dot{Q}_{\text{hvac}}$	heat transfer rate due to HVAC system	W
$\dot{Q}_{\text{ig}}$	heat transfer rate due to internal heat gains	W
$\dot{Q}_{\text{net}}$	net heat transfer rate into the dwelling	W
$\dot{Q}_{\text{sol}}$	heat transfer rate due to solar heat gains	W
R	thermal resistance	$K \cdot W^{-1}$
R_{eff}	effective thermal resistance	$K \cdot W^{-1}$
$R_{\text{in,ext}}$	overall effective thermal resistance of dwelling	$K \cdot W^{-1}$

R'	R-value	$K \cdot m^2 \cdot W^{-1}$
$RMSE_j$	Root-Mean-Square Error associated to the estimation of $\tau_{j,k}$	°C
$scale$	parameter used to scale the Johnson SB PDF	
S_h	cumulative sum of predicted frequencies across bin range 1 to h	
$\hat{S}_h$	cumulative sum of observed frequencies across bin range 1 to h	
t	time	s
t_{d}	time	s
T_{in}	indoor temperature of dwelling	K
$T_{\mathrm{in,o}}$	initial indoor temperature of dwelling	K
T_{ext}	outdoor temperature of dwelling	K
$(T_{\mathrm{ext}})_{j,k}$	estimated outdoor temperature of dwelling based on analysis period j and dwelling k	K
U_{s}	thermal conductance of the building envelope's physical components (i.e. the walls, windows, doors, floors and ceilings)	$J \cdot K^{-1}$
U_{inf}	thermal conductance linked to the building envelope's openings (i.e. doors, windows, cracks and gaps)	$J \cdot K^{-1}$
V_{air}	total air volume of dwelling	m^3
w_j	weighting factor for $\tau_{j,k}$ in calculation of $\tau_{mth,k}$	
x	independent variable of Johnson SB PDF	
Y_i	predicted value	
$\hat{Y}_i$	actual value	

Greek Symbols

ΔT_{in}	indoor temperature difference	K
ΔT_{o}	initial indoor-outdoor temperature difference	K
$(\Delta T_{\mathrm{o}})_{j,k}$	initial indoor-outdoor temperature difference based on analysis period j and dwelling k	K

ρ_{air}	density of air	$\text{kg} \cdot \text{m}^{-3}$
τ	thermal time constant (TTC)	s
$\tau_{j,k}$	estimated Thermal Time Constant based on analysis period *j* of dwelling *k*	s
$\tau_{mth,k}$	weighted average of all $\tau_{j,k}$ values estimated during month *mth* for dwelling *k*	s
$\tau_{ssn,cl}$	mean value of $\tau_{mth,k}$ based on fitted distribution for season *ssn* and the ASHRAE climate zone cl	s
ϕ	normal PDF	
χ^2	chi-squared value	

Chapter 1

Introduction

Canadians and Americans spend approximately 90% of their time indoors [1]; they find themselves aware of this fact now more than ever as the present COVID-19 pandemic has accelerated the trend toward telecommuting. The elevated amount of time that North Americans are spending in their residential buildings may extend past the current pandemic into their regular routines and permanently change the way many individuals interact with their homes. Using residential smart thermostat data and Data Mining (DM), this thesis seeks to provide an innovative, scalable, and timely methodology to estimating a dwelling's Thermal Time Constant (TTC); a thermal property that could help characterize a dwelling's passive transient response. The estimated TTC could be used to rate the passive thermal performance of dwellings and provide valuable insights related to energy conservation in the residential sector.

1.1 Background

1.1.1 Energy Use in the Canadian and US Residential Sectors

The operation of buildings, homes in particular, is a major contributor to the national energy consumption of both Canada and the United States (US). The Canadian and US building sectors are responsible for approximately 28% and 40% of their national secondary energy use, and 22% and 21% of their greenhouse gas (GHG) emissions, respectively [2]–[4]. Over 15 million Canadian and approximately 118 million US households [5] use energy everyday – to heat or cool spaces, heat water, operate lighting, power appliances, electronics and more. Consequently, the Canadian and US residential sectors respectively make up 59% and 53% of the secondary energy use in their corresponding building sectors. Typically, more than half of a household's annual energy consumption is just used to heat and cool occupied spaces. In Canada, heating

and cooling the residential market's nearly 2 billion square meters of floor space accounts for 62% of annual residential energy use. Meanwhile, the US residential market represents over 22 billion square meters of floor space and space conditioning consumes 51% of annual residential energy use. Moreover, more than 40% of both Canadian and US households use natural gas as their main source of heating [6]. As the building industry strives to conserve energy for environmental and economic reasons, energy-efficient building assemblies and space heating equipment, as well as optimized building operation are increasingly recognized as priorities.

Over the next several decades, the majority of opportunities to improve energy efficiency in the residential sector will be in the existing building stock. Although new buildings are being constructed to higher performance standards than ever before, existing buildings — homes in particular — are lagging behind on energy efficiency. Despite the efficiency efforts made in new constructions, the total end-use energy demand and costs are expected to continue trending upwards as a consequence of (i) extreme weather conditions due to climate change, (ii) population growth, (iii) increases in the residential floor space per occupant and (iv) increases in household plug loads [6]–[8]. In both Canada and the US, less than 1% of the residential floor space is newly constructed each year and more than 70% belongs to buildings over 20 years old [2], [5]. The energy reduction potential of the existing housing stock is severely restricted by old equipment, aging infrastructure, improper maintenance and inefficient operation practices. A widespread improvement of thermal performance in existing residential buildings — through building retrofitting and other energy conservation measures- represents a high-volume approach with significant reductions in operation costs and greenhouse gas emissions for both the Canadian and US building sectors. Moreover, protecting and upgrading existing building infrastructure not only improves energy efficiency but also tackles issues of thermal resilience and indoor environmental quality. Considering that more than half of the 2050 building stock has already been built, these buildings should not be allowed to spend the next 30 years: wasting resources, generating pollution that destabilizes our climate, and jeopardizing occupant health and comfort. The challenge of future-proofing residential buildings over the coming decades is a formidable one, but it also represents an amazing opportunity to revitalize the economy post-pandemic as well as the spaces where we both live and work.

Two clear incentives for homeowners to adopt energy conservation measures are rising energy costs and health concerns. In both Canada and the US, a household spends an average of about 3% of their total income on utilities. Energy plays a large role in the lives of the average Canadian and American for whom current annual energy expenses per dwelling are approximately $2,460 CAD and $1,836 USD, respectively [5], [9]. However, there are many households that are limited by their ability to afford the rising costs of energy services. A household may be described as experiencing energy poverty when it allocates more than 10% of its total income to energy expenditures [10], [11]. By this measure, the burden of energy poverty,

also known as household energy insecurity, is carried by an estimated 8% (1.2 million) of Canadian and 31% (37 million) of US households. Unable to afford basic household energy needs, these individuals often turn to coping methods (e.g. extreme reductions in space conditioning) that can result in negative health consequences [12]. In addition to financial incentives from the government, the uptake of energy conservation strategies by existing dwellings would be facilitated by providing homeowners and other building stakeholders with simple and accessible indicators for eligibility.

1.1.2 The Built Environment and the Internet of Things

Driven by economic and environmental concerns, both Canada and the US are searching for ways to facilitate a major evolution of their energy infrastructures — from outdated, centralized and carbon-intensive to resilient, distributed and renewable [13]. The rising energy needs of a growing digital economy and the ever-changing effects of climate change have pushed the North American energy infrastructure to perform beyond its original design capabilities [14]. The next-generation energy system, commonly referred to as the smart grid, is part of the Internet of Things (IoT) framework; IoT refers to a system of physical objects, connected via the internet and communication protocols (e.g. bluetooth, BACnet), that can collect, exchange and act on data acquired from their environment. Geared toward providing the reliable, resilient and sustainable energy services needed for future generations, the smart grid is being developed to:

- support bi-directional flow which allows both electricity and information to be exchanged between a utility and its consumers [15];
- manage energy demand and supply, in real time, using a network of controls, computers, meters, sensors, and other IoT applications;
- accommodate the electrification of the energy system and the integration of renewable energy, thermal storage and energy-efficient technologies.

Low-energy buildings play a foundational role in the effective implementation of the smart grid. To address energy insecurity and improve grid stability, provincial and federal governments have introduced rebates and financial incentives to encourage the adoption of energy conservation measures. Over the years, a number of innovative Demand Side Management (DSM) strategies have been deployed in an effort to minimize the energy use, operating costs and environmental impact of buildings without compromising functionality and occupant comfort. For example, the lower cost and improved performance of sensors and controllers have lead to the development of smart building features such as continuous performance monitoring, automated diagnostics, and optimized controls. Just over a decade has passed since the

introduction of the Smart Thermostat (ST); one of many IoT applications that have inspired visions of a very different future in the built environment - particularly homes. An ST allows homeowners to monitor and automatically regulate the temperature of their dwellings while reducing both their energy consumption and operating costs [16]. As a component of a Building Automation System (BAS), an ST relies on user-defined setpoint schedules, remote sensors and web-enabled data to minimize a building's energy demand; it operates the space conditioning system only as needed to maintain the building's desired indoor temperature. In addition, an ST can also coordinate with utilities to avoid operation during peak demand times and the data it collects could provide insights into a building's thermal performance. Developing a more automated framework for building controls and diagnostics in residential buildings could have a significant impact not only on their operation but on their design and the associated costs as well.

As IoT capabilities become increasingly more affordable and available in the built environment, home automation applications, such as STs, are becoming a popular energy management option among homeowners [17], [18]. A recent study predicted that, by 2025, IoT would offer a potential economic impact of $200-350 billion a year in relation to the residential sector alone [8]. With this increased adoption of IoT applications, sources of high-volume data are becoming more prominent - like in Ecobee's "Donate Your Data" program. Launched in 2016, Ecobee's smart thermostat users started donating anonymized operational data from their buildings for research purposes. Data collection and (shared) access could progressively become the norm. As a result, the building industry will face both opportunities and challenges, as it determines how data from potentially thousands — or even millions — of buildings will be reasonably integrated into the design and energy management of the built environment. Using data science, actionable insights derived from this unprecedented level of data access could guide the building industry on how to improve a number of building aspects related to the development of the smart grid, including a building's passive thermal performance [19].

1.1.3 The Need for Dynamic Thermal Characterization

A building's passive thermal response is a key consideration for reducing the energy required for space conditioning. The energy demand related to heating or cooling a building is not only linked to the operation of its HVAC system; energy demand is also strongly linked to a building's geographic location and the thermal performance of its construction [20]. The use of high-performing HVAC equipment or a well-tuned BAS is far from optimal if used in a building that is poorly insulated and drafty; the inefficiency of such a building can obscure the untapped potential of these technologies for energy and GHG reductions. The North American housing stock is very large and represents a wide range of dwelling types and construction methods

spread across a number of different climate zones. How does one quickly identify the existing buildings which are wasting energy due to their aging or damaged infrastructure?

At present, one of the most common and accessible thermal performance indicators used in the building industry is the apparent (or nominal) R-value but it does not provide the full picture in regards to a building's thermal behaviour [21]. The R-value represents the thermal insulation provided by a building enclosure component (such as a wall, roof or floor) and therefore its ability to resist one-dimensional conductive heat flow, under steady heat flow conditions. As a metric that is simple to find and widely accepted, the R-values of a building's materials are often the only information provided to reflect the thermal behavior of a building. However, relying on the apparent R-value as the only means of characterizing thermal behavior fails to encompass the effects of thermal bridging, building enclosure defects, thermal mass, and air leakage [22].

Using real building performance data, the estimated TTC provides an indication of the time required for the building (or building suite) — without the influence of a space conditioning system — to cool down or heat up in response to a temperature change in its surroundings. The estimated TTC is an indicator of a thermal zone's passive thermal response as a result of the building's effective thermal insulation, airtightness, and thermal mass (i.e. the interior layer of building envelope, internal partitions, furniture and zone air). The TTC is a thermal property whose value is less commonly known and more difficult to determine when time, budget and information are limited; however, knowing the TTC can provide a better understanding of a building's thermal behavior under real conditions.

A simple and scalable TTC estimation method would provide building stakeholders with: (i) a convenient and reliable method of comparing the natural thermal response between buildings, and (ii) the foundations for quantifying poor passive thermal performance if standards are set for TTC values. Buildings with poor passive thermal control could be quickly targeted for energy conservation measures which could make them easier to heat and cool as well as less costly to operate. In addition, understanding the natural thermal response of a dwelling can also help to: (i) determine how long a dwelling can maintain comfortable thermal conditions after a power cut, and (ii) determine which dwellings can act as short-term thermal storage units for load shifting purposes. Moreover, identifying the distribution of TTC values for a particular climate zone (i.e. buildings built for similar thermal conditions) provides energy modellers and policy makers with a reference for setting standards related to the TTC and facilitating the comparison of thermal performance on a larger scale.

1.2 Motivation

A ST generates a large dataset which creates new opportunities towards facilitating the characterization and evaluation of a building's natural thermal response. With little effort and building information required, the collected ST data can be computationally analyzed to reveal patterns, trends, and associations related to a building's transient thermal performance. Normally, a detailed investigation of the thermal behavior of just one building would require extensive on-site monitoring and analysis which, in practice, would only be applicable to a small fraction of the total existing building stock. However, the prospective increase in available ST data from the residential sector could provide a less expensive, less time-consuming, and more scalable manner of characterizing the thermal performance of homes. Due to practical constraints, previous studies related to the TTC have been restricted by a lack of available field data. Nowadays, STs can now be used as a non-intrusive and convenient solution to provide: (i) larger sample sizes, (ii) longer observation periods, and (iii) easier data collection at an urban scale.

Using smart thermostat data from thousands of real homes across North America, this study presents a simple data-driven method for the characterization of a dwelling's natural thermal response; the proposed method uses a simplified grey-box model in order to the estimate of the TTC of a dwelling. To the best of the author's knowledge, no study has yet attempted to use a source of field data this large to estimate the TTC, or to investigate its dependency on local construction practices and time of year. When time and available information are limited, the methodology provided can use large sources of data to:

- provide useful insights into the thermal performance and thermal resilience of dwellings;
- quantify the effective passive response of a dwelling under transient heat flow conditions;
- assist in the development or adoption of energy conservation strategies for dwellings.

1.3 Thesis Objectives

Using smart thermostat data from over 15,000 real homes, recorded over the period of one year, the objectives of this study are as follows:

- To develop a data-driven statistical methodology to estimate the TTC of a dwelling based on low-order Resistance-Capacitance (RC) thermal networks;

- To investigate the dependence of the TTC on other variables such as geographical location and time of year for Canadian and American dwellings;
- To identify the distribution of TTC values for Canadian and American dwellings in order to provide building stakeholders with a reference for decision-making at an urban scale;
- To start a discussion about the TTC and the possible implications surrounding its use as a thermal performance and thermal resilience indicator in DSM strategies (e.g. building retrofits, Model-based Predictive Control (MPC), and Demand Response (DR)).

The results of this thesis are limited to dwellings in various climate zones across Canada and US. However, the proposed methodology is applicable to dwellings in climate zones that fall outside the scope of this thesis under the assumption that similar datasets are available to researchers or building engineers through STs or BASs, in general. In addition, The exploration of the TTC as a thermal resilience indicator is limited to cold-weather habitability. Finally, the impact assessment of the model developed in this methodology on DSM strategies is beyond the scope of this study.

1.4 Thesis Overview

The thesis is structured as follows; Chapter 2 provides a review of the relevant literature on the dynamic thermal response of a building, building modelling approaches, how time constants have been estimated previously, the evolution of thermostat functionality, and how to approach data mining. Chapter 3 provides a description of the large residential ST dataset used in the study and the data preparation process. This chapter also describes the data mining techniques implemented to estimate the TTC value for each Canadian and US dwelling in the study sample. Chapter 4 presents the data mining results and illustrates the dependence of estimated TTC values for single-family detached homes on local construction practices and time of year. This chapter also explores the possible uses of the building thermal time constant as an indicator of both thermal performance and thermal resilience for existing buildings. Finally, Chapter 6 summarizes the main conclusions from the study.

Chapter 2

Literature Review

This thesis investigates how time-series data collected from thousands of residential smart thermostats can be used to develop a simplified and scalable method for estimating the Thermal Time Constant (TTC) of a dwelling. Valuable information can be extracted from this large data source in order to characterize the passive thermal response of a North American dwelling and inform decision-making regarding its Demand Side Management (DSM) strategies (e.g. building retrofits, Model-based Predictive Control (MPC), and Demand Response (DR)). In order to develop the methodology, the following specific research questions (RQ) must be considered:

- **RQ1** What are the factors affecting the flow of sensible heat into and out of a dwelling?
- **RQ2** What is the appropriate modelling approach to estimate a dwelling's TTC?
- **RQ3** What steps must be followed in order to uncover the patterns, trends, and correlations within a large dataset?

The investigation into the first question (RQ1), regarding the factors influencing the thermal behaviour of a dwelling, is explored in 2.1. Meanwhile, the investigation into the second question (RQ2), found in sections 2.2, focuses on the different approaches used for modelling building thermal zones, and previous studies related to TTC estimation. Finally, in response to the third question (RQ3), 2.4 reviews the steps that a data mining process will typically consist of.

2.1 Thermal Dynamics: Understanding Heat Loss And Gain

This section offers a basic overview of heat transfer and the thermal control systems used to manage the heat gains and losses in a building.

2.1.1 Heat transfer in Buildings

Heat transfer is defined as the exchange of thermal energy between two physical systems (e.g. a building/ thermal zone and its surroundings) as a result of a temperature difference [23]. Heat naturally flows in the direction of decreasing temperature — from an area of higher temperature to one with a lower temperature. For example, on cold winter's day, a heated building will naturally lose heat to the outdoors. Temperature difference is the driving force of heat transfer and as such a larger temperature difference corresponds to a higher heat transfer rate. The temperature difference across a building enclosure — in the case of both heat gain and loss — can manifest itself in response to external weather and climate conditions (e.g. indoor-outdoor temperature difference, solar radiation, and wind speed). Interior heat generation can also affect temperature difference and therefore the flow of heat across the building envelope; these internal heat gains are generated inside a building space by the occupants and their activities (e.g. exercising and cooking) as well as by the operation of lighting and equipment (e.g. computers and appliances). Over time, the temperature difference between two systems will naturally decay; as the two systems approach the same temperature, the rate of heat transfer decreases. Once a state of equilibrium has been reached (i.e when the temperature difference equals zero), the net heat transfer between the two systems will cease to exist.

In regards to time, heat flow can occur under either transient or steady state conditions. Under steady state conditions, both the temperature at any particular point in the physical system and the heat flow remain constant with respect to time [23]. An example of steady state heat flow conditions is a building on a cold winter's day whose rate of heat loss is equal to the rate at which it is heated, therefore its indoor temperature is maintained. In contrast, heat flow is described as transient when its magnitude or direction changes with respect to time. Under transient conditions, heat enters and exits the physical system at different rates, and the spatial distribution of temperature within the physical system also keeps changing with time. An example of transient heat flow conditions is the thermal response of the previously mentioned building in free-running mode (i.e. with its heating system now switched off); the temperature continuously decreasing before reaching the ambient temperature at equilibrium.

The thermal behaviour of a building or one of its thermal zones is a function of the dynamic relationship between the different heat transfer mechanisms [23]. Sensible heat is the thermal

energy — independent of phase changes — needed to increase or decrease the temperature of some substance; it is transferred to, from, and within the system by three different mechanisms, acting separately or in combination: conduction, convection and radiation.

Heat enters, leaves, and moves within a building in many different ways; however, as it travels from one place to another, heat will always flow along the easiest path available and use the most efficient transfer method [23]. During the process of heat flow through and within building systems, the mode of heat transfer can change frequently [24]. For example, the sun transmits heat by radiation to the earth, where it can be absorbed by the building's opaque exterior walls or transmitted through a window to interior surfaces. From the exterior surface of the opaque wall, the heat then travels by conduction through the wall assembly where it is transferred to the indoor air by convection or to the indoor surfaces by radiation. Moreover, heat can also be directly transferred into the building through the infiltration of outdoor air or out of the building through the exfiltration of indoor air; air leakage can occur through openings in a building enclosure (e.g. gaps, cracks and holes). Understanding the fundamentals of how heat is gained and lost is critical when assessing the efficacy of a building's thermal control systems; effective thermal control is achieved by matching the appropriate measures to the heat flow observed. Arguably, the primary concerns of building owners, architects, engineers and utilities is providing thermal comfort to building occupants without excessive energy use and space conditioning costs; thermal control systems are therefore an integral part of virtually all buildings.

2.1.2 Thermal Control: Passive Versus Active Systems

Over the useful life of the building, thermal control is used to maintain the indoor temperature within a range required for the personal comfort, productivity and health of the occupants[25]. There are two common approaches to managing heat flows in buildings: passive and active systems.

The passive systems use the external environment and the inherent properties of the building to moderate the indoor temperature [25]. A building's passive thermal performance is almost entirely determined by the degree of thermal control provided by the building enclosure; it is responsible for passively managing the exchanges of heat, air, moisture and solar radiation that occur between the indoor and outdoor environment. Examples of passive thermal controls include the building's orientation, geometry, effective thermal insulation, thermal capacitance, airtightness, fenestration and shading devices.

The active systems are energy-consuming technologies (e.g. HVAC systems using fans and pumps) that provide auxiliary heating or cooling [25]. Examples of active systems include,

air-conditioning, heat pumps, radiant heating, heat recovery ventilators, solar electric and solar thermal panels, and geothermal energy exchangers. Normally found within boundaries of the building enclosure, they are used to extend or augment the thermal performance of the building's passive systems. Most modern buildings employ a complex combination of both passive and active thermal control systems in order to efficiently deliver the desired level of control and functionality to the indoor environment. The proper coordination of a building's passive and active systems is crucial not only to thermal comfort but energy efficiency, thermal resilience and durability as well.

Higher levels of energy efficiency can be achieved when using passive systems to the greatest practical extent and only employing active systems to supplement them [25]. Before the Industrial Revolution, the thermal control of buildings was rooted in passive systems; however, the abundance of affordable fossil fuels and the invention of powerful space conditioning systems, in the 1900s, led to the heavy reliance of buildings on active thermal control systems. In recent years, the scarcity of fossil fuels, volatile energy costs and climate change have made this approach obsolete. A more powerful HVAC system is not a sustainable substitute when the building fabric provides a poor passive thermal control; active systems should only be practically used to close the energy deficit gaps that cannot be managed through passive measures. Passive measures — when properly incorporated — can conserve the energy sources used to power active systems.

Today's buildings are being designed with energy conservation and efficiency in mind in order to: (a) provide better level of energy security if supply issues were to occur, (b) extend the useful capacity of the grid and (c) positively impact climate change mitigation. As a result, building designers are reevaluating how passive and active measures are to be best applied in new buildings and retrofits.

The thermal resilience and durability of a building are also best served when passive measures are the foundation of the building thermal design [25]. In the event that active systems for space conditioning have been disabled (e.g. due to an extended power outage), the building thermal resilience is the measure of how long its occupants can safely inhabit it before the indoor temperature becomes too cold or too hot. In both Canada and the US, extended power outages are often the result of a strained energy infrastructure, extreme weather events or natural disasters [12], [14], [26]. Taking into account that extreme weather events are expected to become more frequent and more intense, thermal resilience is becoming an increasingly important design consideration — particularly for residential buildings. Moreover, a poorly insulated and drafty building would not only be uncomfortable to its occupants, but be vulnerable to a number of durability issues, due to large fluctuations in temperature and the uncontrolled entry of humid air (e.g. mold growth, decay, corrosion, cracks, frost damage and freezing water pipes). Furthermore, the proper operation of the HVAC system can also be affected by the poor

performance of passive measures such as an inferior enclosure, therefore requiring increased equipment run times which can translate to a shorter service life and higher maintenance costs.

2.2 Modelling Heat Transfer in Buildings

Whether in the name of the occupant thermal comfort or the performance of building systems, Building Energy Modeling (BEM) is often used to get a better grasp on the heat transfer between a thermal system and its surroundings; models help assess the impact of the complex synergistic relationship between a building's passive and active systems.

At all stages of a building's life-cycle, BEM is used to facilitate the design, analysis and optimization of modern buildings systems. In this section, BEM is reviewed, with the estimation of the TTC and the other study objectives in mind. The thermal model required for the TTC estimation needs to be simple enough so that it can be found with the limited building information from smart thermostats yet accurate enough to inform decision-making in the early design stages of a retrofit, or during the building's operation.. The following three categories of building energy models are defined and their suitability is discussed: white-box models, black-box models and grey-box models.

2.2.1 White-box Model

A white-box model is a very detailed physics-based description of a thermal system and can provide an accurate prediction of a building energy system by explicitly modelling all physical processes that play a significant role in its energy balance [17]. By incorporating a comprehensive set of parameters, white-box models consider a building's physical properties, its systems and its environmental conditions. Due to their accuracy, there are numerous examples of building simulation software tools that rely on a white-box modeling approach including TRNSYS [27], EnergyPlus [28], and DOE-2 [29].

White-box models can achieve a high level of detail in their predictions; these models require numerous inputs, extensive computational power, and high development and implementation costs [30]. The model inputs required can normally be found using design plans, manufacturing catalogues or on-site inspections; however, some of these inputs can often be difficult to obtain or even unavailable at times. White-box models are favored during the design of a new building - when the required building specifications are more accessible and performance data is not yet available [17]. With the model inputs in hand, a significant amount of time is still required for building professionals and energy modellers to develop the white-box model

for a particular building. Due to their exhaustive examination of a building, white-box models are impractical for a large-scale application on the existing housing stock.

2.2.2 Black Box Model

Black-box models use a data-driven (or inverse) approach to model development, relying solely on data science techniques and performance data from the building considered [30]. With a very short development time, black-box models can develop a mathematical representation that captures the relationship between known input and output variables (i.e. important influencing variables and the observed building behavior, respectively); this mathematical model and its estimated coefficients can be used to reproduce the building system behaviour for future inputs.

Although they have proven to produce accurate results, black-box models cannot offer any insight into the physical processes that drives those results, and are also limited by the quantity and quality of the data provided [17]. The lack of any definitive link to the physical environment means that black-box models are building-specific and cannot be used determine the effect of design or operational changes [30]. Lacking generality, black-box models would therefore not be suitable for comparing thermal behaviour across the wide variety of buildings found in the existing housing stock. There are several examples reported in literature of data-driven modelling techniques used for BEM including multiple linear regression [31], neural networks, and support vector machines [32].

2.2.3 Grey Box Model

Grey-box models represent a compromise between white-box and black-box models. They are simplified physics-based models that use an inverse modelling approach in order to estimate key system parameters based on measured building data [17]. Once their parameters have been identified, the system can be solved numerically. Grey-box models exploit the physical knowledge of a building energy system, similar to a white-box model, but are simplified versions with shorter development times due to the incorporation of building performance data. Due to their hybrid nature, grey-box models can be used to make comparisons between buildings in a large housing stock but only in relation to the key physical parameters being estimated. [30].

For the development of grey-box models, the Resistance-Capacitance (RC) method is a very widely used approach. An RC model is established based on an electrical analogy and therefore treats heat flow as a current $\dot{Q}$ in a thermal network. As heat flows through a physical element

in a thermal system, its ability to resist the flow of heat to and from its surroundings is modelled by an electrical resistor with a thermal resistance R. Meanwhile, its ability to store heat (if applicable) is modelled by an electrical capacitor with a thermal capacitance C.

The order of a model is an indication of its complexity and is defined by the number of capacitances used in its thermal network; a higher model order indicates a higher resolution (or complexity). Referring to a grey-box model as simplified generally implies the use of a low-order RC model because reducing the model order decreases, in turn, the number of differential heat equations to solve and the number parameters to be identified [33].

The thermal network method often uses a number of approximations and assumptions in order to simplify and solve the otherwise complex differential heat equations representing the thermal processes and properties linked to a building energy system [34]. For example, an RC model, also known as a lumped-parameter model, aggregates physical elements of the thermal system into discrete "lumps" (or groupings); the underlying assumption being that inside each lump the temperature differences and rates of heat transfer between the physical elements are negligible [23]. As heat flows through the lump, the distributed thermal resistances and thermal capacitances of its physical elements are represented in the thermal network by an equivalent resistance and an equivalent capacitance, respectively.

With a long history in BEM [35], RC networks have frequently been used to accurately model the thermal behaviour of buildings. The literature presents vastly different answers to the question of which model type and resolution is most suitable in regards to the data-driven thermal characterization of buildings [36]–[38]. The mixed conclusions could potentially arise due to the fact that each model has a suitable place in real-world applications. For example, the level of detail required to represent a geographical region on a map also varies, according to the map's intended purpose; a simple diagram of Montreal's metro system will not help a tourist to navigate through the city's streets, but its simplicity is useful when identifying where to transfer from one metro line to another. As observed by statistician George Box [39]: "*All models are wrong, but some are useful*". Even the most detailed models are just an approximation of real-world processes; they are built on assumptions and contain some degree of error. In contrast, even rather limited models can provide useful approximations and their structures can help develop a conceptual understanding of how dynamic thermal systems operate. The model resolution (or degree of detail and complexity) required depends on the intended application of the model, the level of accuracy the application requires and the performance data used to identify the model's key parameters [34].

2.2.4 Thermal Time Constant Estimation

In addition to defining the TTC, a useful building characteristic, the next section reviews previous studies that have attempted to estimate and employ it while investigating the transient thermal behavior of buildings.

The Thermal Time Constant

The TTC is a thermal performance indicator that characterizes the transient thermal response of a free-running building zone's indoor space as a result of the outdoor temperature variations [36]. Isolated from the heat gains due to HVAC system, the solar radiation and occupant behaviour, the natural thermal response of each building to outdoor temperature variations is different and varies based on their passive thermal control systems. In its simplest form, the thermal time constant of a thermal system, τ, is defined as the product of its effective thermal resistance R_{eff} and its effective thermal capacitance C_{eff} that is

$$\tau = R_{\text{eff}} C_{\text{eff}} \tag{2.1}$$

This equation displays that TTC has a positive relationship with both R_{eff} and C_{eff}; as these parameters increase in value, the TTC also increases and therefore makes changes in T_{in} slower to achieve.

The R_{eff} represents both the thermal insulation and airtightness of the thermal zone. The thermal insulation of a building enclosure limits the amount of heat that enters or exits the occupied spaces by conduction (which is the predominant heat transfer mechanism through the opaque sections of the building enclosure). In addition, the airtightness of the building enclosure minimizes the heat and moisture flows due to the air leakage across the building enclosure (both infiltration and exfiltration). A number of previous studies have proven the effects of air leakage and ventilation on the τ to be considerable. Air infiltration or exfiltration — whether controlled and uncontrolled — have proven to decrease a building's thermal resistance; its R_{eff} is therefore lower than its apparent (or nominal) thermal resistance and, as a result, its τ is reduced as well [22], [36].

In addition, the C_{eff} represents the zone's thermal mass which includes the building enclosure's interior layer, its internal mass (i.e. internal partitions and furnishings) and its zone air; thermal mass provides thermal stability to a building space by dampening the indoor temperature fluctuations. Apart from the zone air thermal capacitance, the thermal capacitance attributable to the rest of a building zone's thermal mass is important but difficult to measure in

practice, and is therefore an uncertain input parameter in white-box models [40]. As a result, building simulation tool EnergyPlus estimates C_{eff} in the following form:

$$\begin{aligned} C_{\text{eff}} &= C_{\text{air}} \cdot ZCM \\ &= c_{\text{p,air}} \cdot \rho_{\text{air}} \cdot V_{\text{air}} \cdot ZCM \end{aligned} \tag{2.2}$$

where $c_{\text{p,air}}$ is the specific heat of the zone air, ρ_{air} is the density of the zone air, V_{air} is total the zone air volume, and ZCM is the zone air thermal capacitance multiplier. The multiplier ZCM is an equivalent adjustment to the zone air thermal capacitance which required to incorporate the thermal capacitance attributable to the interior layer of building envelope, internal partitions, and furniture.

Consequently, the key passive thermal control features represented by the TTC are as follows:

- Thermal Insulation and Airtightness;
- Thermal Mass.

These building-integrated features all contribute to decreasing heating and cooling loads and providing thermal resilience, and the TTC can serve as an indication of when these features are poorly incorporated.

Previous Studies

The following studies have estimated the TTC for use as a thermal performance indicator and explored a number of important factors that affect it.

First, Antonopoulous and Tzivanidis [41], [42] developed a numerical procedure for calculating the transient indoor temperature; using an implicit finite-difference method, it expresses the indoor energy balance and the transient heat conduction in all the building envelope elements. Using the above procedure, the TTCs of Greek buildings, ranging from 30-3000 m^2 in size, were systematically generated for 21 building types, 18 different wall constructions, and 10 different roof constructions. The results showed that the most important factors affecting the TTC were the wall and roof construction, the size of the building, the surface area exposed to the ambient environment, and the level of ventilation and air leakage. In addition, buildings with different construction characteristics and sizes but with the same TTC were found to have a similar response to the same outdoor temperature variation.

In addition, Antonopoulos and Koronaki[36], using the simulated data from the previous study, split the total TTC into components, examining the contributions of the insulated envelope, the interior partitions and the furnishings. The effects of the internal mass on the TTC of buildings proved to be significant. For example, the indoor partitions of a typical Greek single-story house may increase the TTC up to about 25% and its furnishings would do so up to about 15%. In addition, it was observed that ventilation and air leakage also have considerable effects on the TTC; typically it may decrease it by about 25% except, in some cases, where high ventilation may cause the drop to reach closer to 90%. Moreover, concerning the transient indoor temperature response, the results from the time-consuming and rigorous numerical procedure from the previous study are fitted using a least-squares regression analysis to a simple and flexible first-order RC model. The authors state that the simple model displayed satisfactory accuracy in practice.

Next, Catalina et al. [43] developed regression models to serve as easy and efficient load forecasting tools in the early stages of design for single-family residential buildings in France. The model is obtained by using multiple regression analysis and simulated data generated using TRNSYS. The model inputs included the TTC, the envelope's overall heat loss coefficient, the window-to-floor-area ratio, and the climatic conditions. The results showed a strong relationship between the energy demands linked to space-conditioning (i.e monthly and annual heating demands) and variables such as the building morphology and the TTC.

Moreover, Karlsson et al. [44] developed a conceptual model in order to explore the implications of changing the thermal capacitance of building materials. Programmed in MATLAB, the numerical thermal model represented a wall made up of an exterior wall, an indoor air volume and an internally thermally heavy wall. Six different simulation cases were analyzed over a period of 20 days, varying the external temperature, solar gains and heating. Outputs such as energy consumption, peak power consumption and thermal comfort parameters were analyzed with respect to the thermal capacity and heat loss coefficient of the walls. Using the thermal properties defined for each case, the TTC was calculated for comparison purposes. The TTC grew larger in value with every increase in the volume and/or volumetric heat capacity of the wall. Tested under cold-climate conditions, the results showed the passive storage of sensible heat in the thermally heavy inner wall can significantly change the consumption pattern; this offered significant benefits in shifting thermal loads but not in reducing the total energy consumption.

Finally, Harb et al. [38] developed and compared 4 different RC model structures of varying in complexity (i.e. 1R1C, 3R2C, 4R2C and 8R3C) to forecast the thermal response of occupied buildings. The models are used to reproduce and approximate the thermal behavior of 3 different buildings in normal operation, after estimating their building-specific parameters using historical data. The 3 building case studies, which varied in type, size and heating system,

included: one German office building, one Swedish office building and one Swedish residential building. The inputs for the models consisted of measurement time-series data for the outdoor temperature, solar irradiation, building heat consumption and indoor dry air temperature. The prediction error was quantified using Root-Mean-Square Error (RMSE) as a model performance criteria. The TTC, the heat losses, the internal wall capacity, external wall capacity and heat transfer resistances were all calculated and compared. Although, the second order (4R2C) model performed with the most accuracy (with a mean forecast error of 0.2K), a model comparison revealed that all models gave physically reasonable estimates of the building properties. The temperature simulation provided by the simplest model structure (i.e. 1R1C) delivered a dampened reproduction of the measured temperature fluctuations used as a reference.

In brief, previous studies have estimated the TTC for buildings but mainly using simulated building data and for European locations. Those studies that did use real building data were had small sample sizes and short observation periods. Previous research also confirmed that important factors affecting the TTC include the wall and roof construction, the building size, the surface area exposed to the ambient environment, and the level of ventilation and air leakage. In addition, a strong relationship was observed between the TTC and the energy demands linked to space conditioning. Another pertinent observation was that buildings sharing the same TTC — despite their differences in construction and size — were found to have a similar response to the same outdoor temperature variation. Moreover, it was shown that a first-order model can represent – with sufficient accuracy – the indoor temperature response of a simulated free-running dwelling. Finally, previous research also displayed, under cold-climate conditions, that a building's thermal mass can be used to shift its thermal loads and its energy consumption patterns.

Although the operational data used to calibrate grey-box models can be derived from simulated buildings or real ones, most of the previously mentioned studies relied on simulated building data; until recent years, sizable and quality sources of real building data have been less accessible [17]. Sources of high-volume data from real buildings are becoming more prominent as IoT capabilities become increasingly more affordable and available in the built environment [8]. In particular, a significant increase in the availability of smart thermostat data has been observed in recent years and is expected to continue trending upwards. The arrival and growing presence of the ST can serve as an unprecedented source of data-derived insights for TTC estimation and ZCM estimation in EnergyPlus, and can also change the way we deliver thermal comfort to homes in the future.

2.3 Evolution of Thermostats: From Manual to Smart

A thermostat is a device used to regulate the temperature of a physical system (e.g. a building or a specific space within a building); it activates the heating or cooling equipment in order to maintain the system's temperature near a setpoint (i.e. a desired temperature). For example, in heating mode, if the system's temperature falls below the set point then the thermostat runs the heating equipment to reach and maintain that temperature; alternatively, if the temperature rises above the set point then the thermostat would simply turn off the heating system. The three existing thermostat types are discussed in this section and their features can be seen at a glance in Table 2.1.

Table 2.1: The three thermostat types and their corresponding features at a glance

	Manual Thermostat	Programmable Thermostat	Smart Thermostat
Regulates Temperature	Yes	Yes	Yes
Programs Setpoint Schedule	N/A	Yes	Yes
Controlled Remotely	N/A	N/A	Yes
Considers Web-enabled Data	N/A	N/A	Yes
Considers Occupant Behaviour	N/A	N/A	Yes

The Smart Thermostat (ST) emerged in part as a market response to usability problems of its predecessors — the Manual Thermostat (MT) and the Programmable Thermostat (PT) — but also due to the growing trend of connectivity and automation in buildings [16]. Just over a decade has passed since the introduction of the ST whose automation, connectivity and centralized data collection are inspiring visions of a very different future for the built environment. It is the combination of its automation, programmable set-point schedule, web-enabled features, and consideration of occupant behavior that have made the data collection for this study possible.

2.3.1 Manual Thermostats

Having once been the most common thermostat type available, the manual thermostat (see Figure 2.1) is quickly being replaced and phased out by its more energy-efficient counterparts: the Programmable Thermostat (PT) and the Smart Thermostat (ST) [45].

(a) (b)

Figure 2.1: Two examples of manual thermostats [46], [47].

The Manual Thermostat (MT) can only retain one setpoint at a time and offers very basic control of building HVAC systems. If a different setpoint is required for a specific activity or a period of anticipated vacancy, all adjustments must be made manually by the user. If connected to both the heating and cooling systems, these thermostats must also be manually switched from heating mode to cooling mode, and vice versa. Due to the potential energy savings, some building codes and government programs began promoting the use of more convenient and efficient thermostat types.

2.3.2 Programmable Thermostats

Unlike the MT, the PT functionality enables the user to operate their HVAC systems based on a custom setpoint schedule and independently carries out setpoint adjustments; users can set schedules in advance, and identify a different setpoint for different times of the day and different days of the week [45]. In addition to the convenience and customization, PTs can provide the benefit of energy savings.

PTs were demonstrated to be capable of achieving up to 30% in energy savings through optimized thermostat setpoint schedules [16]. For example, a temperature setback is a simple strategy to help save on utility costs by reducing how often the heating or cooling system operates; this is achieved by lowering setpoints in heating mode or by raising setpoints in cooling mode (see Figure 2.2). Setbacks can be used during periods where a building can be hotter or colder than the normal setpoint without compromising comfort or functionality (i.e. when dwellings are unoccupied or when the occupants are asleep). The larger the setback — meaning the lower (in heating mode) or the higher (in cooling mode) the setpoint is compared to normal — the more the user can potentially save.

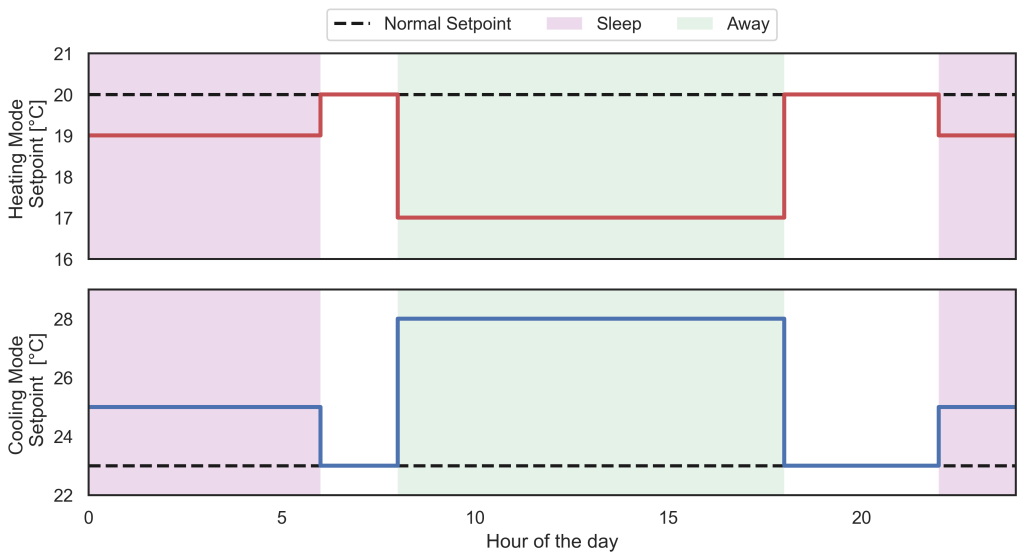


Figure 2.2: Two examples of a thermostat setpoint profile with setbacks; one example for a day in heating mode (top) and another for a day in cooling mode (bottom).

Unfortunately, although many homeowners acknowledge that a PT can save them money, they still use their PT like a MT — by raising and lowering the temperature manually — and/or use it without setbacks therefore missing out on the potential benefits. Frustrated by the process of programming a setpoint schedule, many consumers either misuse, override or abandon the programmable functionality altogether [48].

2.3.3 Smart Thermostats

In an attempt to mitigate the issues with human error involved with PTs, STs have capitalized on advancements in information and communications technologies, and have rapidly expanded their capabilities (see Figure 2.3). Similar to PTs, STs can help users save money, energy and time by controlling HVAC systems through a flexible setpoint schedule. In addition, STs are also web-enabled; they can therefore be controlled and scheduled remotely (via a web portal) using smartphones, tablets and computers.

Figure 2.3: Examples of smart thermostats currently on the market [49]–[53].

As part of the Internet of Things (IoT) market, STs incorporate the use of remote sensors, web-enabled data (e.g. forecast weather from weather stations), algorithms, machine learning, and cloud computing to achieve the optimal balance between comfort and efficiency [16]. Marked by rapid growth and innovation, IoT is a network of internet-connected objects able to collect and exchange data [8]. STs can therefore record and track the building's performance data over time, and send reports regularly informing the user of their energy consumption and savings. Remote sensors are used to monitor a building's temperature, humidity, HVAC equipment, and occupancy. Unlike its predecessors, the ST uses data collection and connectivity to guide its management of set-point schedules and its adjustments to HVAC equipment runtimes. The ST can therefore also identify any consistent modifications made manually to a user-defined setpoint schedule and modify the schedule automatically to incorporate the new patterns of behaviour. As a result, STs have been increasingly used to reduce the electricity demands on the grid in addition to lowering the consumer's energy bills.

Through the use of grey-box models and data mining techniques, the large datasets generated by STs create new opportunities to discover patterns, trends, and correlations related to a building's intrinsic thermal performance. The connectivity and the centralized data collection of an ST permits experimental studies to be performed on a larger scale, for a number of different locations across the globe, and for a far more diverse sample of residential buildings than was previously practical.

2.4 Data Mining: Exploring Large Datasets and Information Extraction

Heavily reliant on information technology, Data Mining (DM) is the systematic extraction of patterns, trends and relationships hidden among large and complex datasets [54], [55]. Over the years, DM has proved useful in many fields including engineering, medicine, finance and marketing. In recent years, DM has been introduced into the field of building science in response to the widespread adoption of building automation systems (BASs) in modern buildings; BASs generate and store vast amounts of data (related to system operation, occupant behavior, power consumption, climatic conditions, etc.) that can be analyzed in order to predict or describe building performance [56]. Regardless of the field, the Cross-Industry Standard Process for Data Mining (CRISP-DM) is a methodology commonly applied to data science projects that has proven to be highly practical, flexible and useful. CRISP-DM consists of six phases placed in an idealized sequence [57], [58]:

1. *Domain understanding (or business understanding)* is the first phase of the process and focuses on building the foundations of the data science project — a clear understanding of the domain being researched, setting the objectives, determining the project requirements, and defining the success criteria [57], [58].

2. *Data Understanding*, the second phase, involves the identification and collection of task-relevant data, and preliminary analysis of the dataset [57], [58]. Once acquired, it is important to develop a familiarity with the data by examining and documenting its current state including quality issues and organization. At this stage, preliminary attempts are made to explore the data using visualization tools and basic statistical methods in order to identify relationships.

3. *Data Preparation* is a very time-consuming process but arguably the most important part of the procedure; this phase readies the dataset for modelling [57], [58]. In an attempt to improve the final modelling results, common practices during this phase are to include new attributes helpful to analysis, to exclude or correct erroneous data, to combine multiple sources or divide the data into subsets, and to reformat as necessary.

4. *Modelling* is a phase that involves the selection of the modelling technique(s) that apply to the set objectives, the development and assessment of the model(s), and the interpretation of the modelling results based on the domain knowledge; this phase may consist of multiple iterations [57], [58]. Well-known DM techniques commonly linked to building science include: regression analysis, clustering analysis, association rule mining (ARM), artificial neural networks (ANN), decision trees, and support vector machines (SVM)

[56]. Larose and Larose as well as Jiawei et al. provide detailed information on a number of commonly used DM techniques including those previously mentioned [55], [58].

5. The *Evaluation* of the modelling results is a review of the tasks accomplished and a summary of all findings; an opportune time to determine whether or not to proceed to deployment, reiterate an old task, or initiate a new task [57], [58].

6. *Deployment* can either be as simple as generating of a report that clearly communicates the final data mining results or as complex as developing a plan to implement the findings in a new project [57], [58]. Anything of note regarding the data mining process such as successes, failures and improvements to be implemented in the future can also be relayed through the report. The knowledge derived from large datasets can be used to inform decision-making and problem-solving tasks linked to the field in question.

In practice, CRISP-DM may involve performing these steps in a different order than prescribed, backtracking to previous steps, and/ or even repeating several iterations of a step.

Chapter 3

Methodology

3.1 Data Mining Approach

The methodology developed for this Data Mining (DM) project can be broken down into the following phases: Domain Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment. Figure 3.1 provides a flow chart of the approach used to guide this study, inspired by the Cross-Industry Standard Process for Data Mining (CRISP-DM).

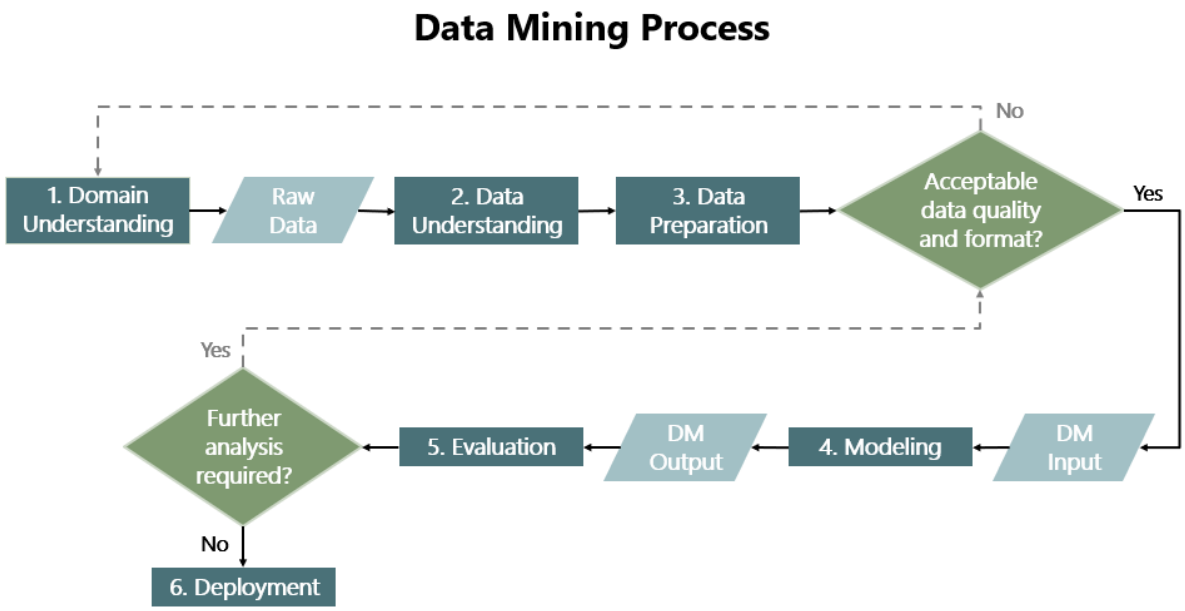


Figure 3.1: A flow chart representing the steps in this study based on CRISP-DM.

The first phase of the DM process, Domain Understanding, can be found in Chapters 1 and 2, and the last phase, Deployment, can be found in Chapter 4. This chapter provides a

description of phases 2 through 5. Section 3.2 describes the original format and content of the large dataset used for the study. Next, Section 3.3 provides a basic overview of how the raw data is cleaned and transformed in order to act as an effective DM input. Finally, Section 3.4 goes on to discuss the DM modelling techniques and the performance criteria used to evaluate them. All of the Python scripts developed to perform the study are accessible through the link provided in section B.1.

3.2 Data Understanding

ecobee, a Canadian home automation company, has established the Donate Your Data (DYD) program for its ST users to donate the operational data from their residential HVAC systems to science. Within two years of the program's 2016 launch, the ecobee database grew to over 70,000 participating dwellings worldwide with up to 36 months of data recorded at 5-minute intervals. The measured data used in this study was gathered during normal occupied operation from the beginning of September 2015 to the end of August 2018.

The scope of the data collected from each registered ecobee thermostat user consists of the measured time series data acquired through the connected ecobee devices (see Figure 3.2) and the user-reported contextual information about the dwelling. The ecobee thermostat is wall-mounted and connects to the dwelling's HVAC system in the same manner as a conventional thermostat [59]. The thermostat's accompanying remote sensors are independent devices that are both wireless and battery operated; their recommended placement is on a flat surface in the dwelling less than 13.7 m from the thermostat and at a height of 1.5 m. The DYD program participant data is pulled from Ecobee's servers, and then anonymized to remove any personally-identifiable information.

Figure 3.2: The ecobee 3 (left), a smart thermostat model, and its remote sensor (right). *Photo courtesy of ecobee inc.*

When the participating thermostat users opt into the DYD program, they are agreeing to make all of the data collected by their thermostat and remote sensors (since the time of installation) available for research purposes. Every month since its installation date, each thermostat saves the data it has recorded (for the month) in an individual CSV file with a consistent and uniquely generated filename. The data files from all registered DYD thermostats worldwide are organized in folders corresponding to the month and year when the data was collected. Listed and described in the Table A.1, the 20 variables provided in each data file are recorded at 5-minute intervals. Each observation in the dataset is stamped with a date and time, beginning from the first instance of use. The thermostat records indoor measurements for temperature, humidity and motion detection. In addition, other recorded variables include HVAC equipment runtimes, scheduled setpoints for temperature (and humidity, if applicable), user-defined descriptors for the setpoints (i.e. sleep, away, etc.), the outdoor temperature and outdoor humidity. The values associated to weather conditions are provided based on the nearest weather stations. Any modifications to the schedule made by the user, by the thermostat based on sensors, or by a third party in a demand response event are also recorded.

Outside the monthly folders, one CSV file titled metadata contains an overview of dwelling characteristics for every participating residence. Listed and described in the Table A.2, there are 19 dwelling characteristics provided in the metadata file. The metadata file includes basic contextual information for all dwellings such as the geographic location (e.g. city, province/state and country), the age, the total floor area, the style, and the number of occupants. For each dwelling, the following additional information is also provided: the thermostat model, the unique filename, the number of floors, the number of installed heat stages, the number of installed cool stages, the number of remote sensors, the fuel type for auxiliary heat, the presence of a heat pump.

Finally, in regards to data access, the metadata file and the monthly folders of data are shared directly with research partners and securely downloaded through a Secure File Transfer Protocol (SFTP), under the understanding that all resulting research findings will be shared publicly for everyone's benefit.

3.3 Data Preparation

All data preparation tasks discussed in this section are accomplished using *Pandas*, an open-source data analysis library developed for the Python programming language [60]. The preliminary data preparation tasks include: addressing duplicated metadata entries, addressing errors and inconsistencies in the metadata, determining the study duration and sample size, and metadata reduction.

Duplicated Metadata Entries

Considering that each thermostat was given a uniquely generated filename, the first step in this study's data preparation process is the removal of any duplicated entries in the metadata and monthly folders. Duplicated entries were only present in the metadata. The total number of registered thermostats, as provided in the metadata file, is 76,003 worldwide. Among the listed filenames, there are two types of duplicated entries:

- The first type of duplicated entry has the identical filename and the identical contextual information of another entry. In this case, the originals were kept and a total of 4,485 duplicated entries were removed from the study.
- The second type of duplicated entry has the identical filename of another entry but with different contextual information. In this case, there is no way to identify the correct corresponding contextual inputs for the filename therefore 152 listed thermostats were removed from the study including the 43 originals.

After the removal of the duplicated entries, the total number of remaining registered thermostats is 71,366.

Errors and Inconsistencies in Metadata

The second step in this data preparation process involves tackling entries in the metadata file with discrepancies in their contextual information. Unlike the data in the monthly folders, the metadata is user-reported and subject to erroneous entries due to human error, efforts to conserve privacy, efforts to incorporate levity, etc. The observed discrepancies include imaginary locations, zero floors, zero occupants, floor areas of zero units, dwellings with more than 4 floors, dwellings with more than 8 occupants, etc. These unrealistic inputs are removed and changed to not available in order to improve the results for any future data analysis involving their associated variables.

In addition, all spelling errors and inconsistencies in city, province/state and country entries are located and corrected in order to simplify the data analysis at later stages in the study.

Determining Study Duration and Sample Size

The third step in the study's data preparation is the determination of the study duration and its sample size. In order to analyze the impact of seasonality on thermal response of a building,

at least one full year will need to be considered, and all dwellings belonging to the sample will be required to have a data file for each month of the selected study duration.

The maximum sample size possible can be determined by examining the number of data files provided by ecobee on a monthly basis. Figure 3.3 shows, from September 2015 to August 2018, the total number of data files provided by ecobee in each monthly folder. As previously mentioned, the data collected was organized by month and year. In a monthly folder, each data file corresponds to an individual dwelling in the DYD program and contains all the data recorded from its ecobee devices over the course of the corresponding month. During this three-year period, Figure 3.3 illustrates a consistent increase in the total number of data files from month to month; therefore the number of participating thermostat users in the DYD program are increasing over time. In the last month of provided data, August 2018, the number of files observed is 64,488 - nearly 7,000 dwellings less than the total registered dwellings in the metadata file. This difference reflects that, over the course of the three years, DYD participants have been lost as well as gained; some thermostat users may have discontinued their participation in the program or their services with the automation company. Consequently, due to either a late registration or a departure from the program, not all dwellings will have a data file present in each monthly folder. With this consideration in mind, the largest possible sample size would be 39,652 dwellings for a study duration of 1 year, 15,246 for 2 years, and 4,594 for 3 years (see Figure 3.3).

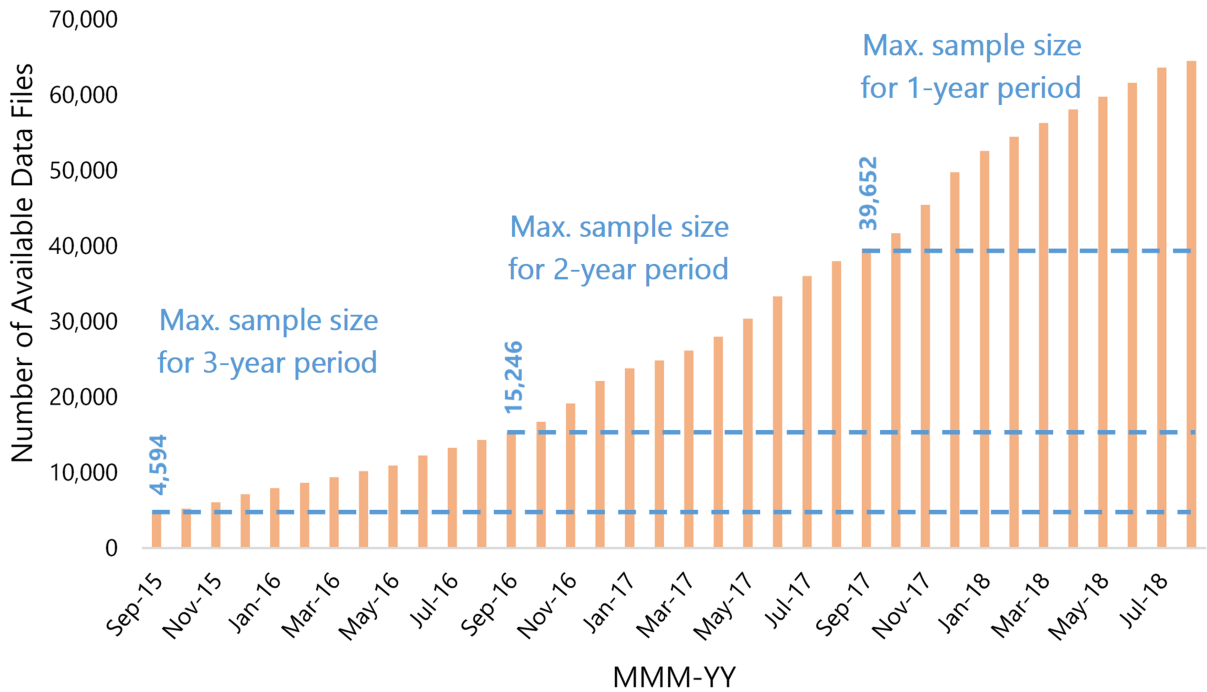


Figure 3.3: The number of data files available per month, from September 2015 to August 2018. Considering that each data file represents one dwelling, the largest possible sample size for a study duration of a 1-year, 2-year and 3-year period are indicated.

Considering the possibility of departures from and late registrations to the DYD program, each study duration was analyzed to determine the filenames common to each month. For a fair comparison of a dwelling's thermal response over the course of a year, the households selected for the study are required to have a data file available for every month in the considered study duration. As a result, Figure 3.4 shows that, among the 71,366 DYD participants worldwide, 35,470 homes consistently have data for a full year, 12,580 homes for two consecutive years and 3,445 homes for three consecutive years. Out of the 35,470 residences that have one year of data, more than 35,000 are located in Canada and the US (see Figure 3.5).

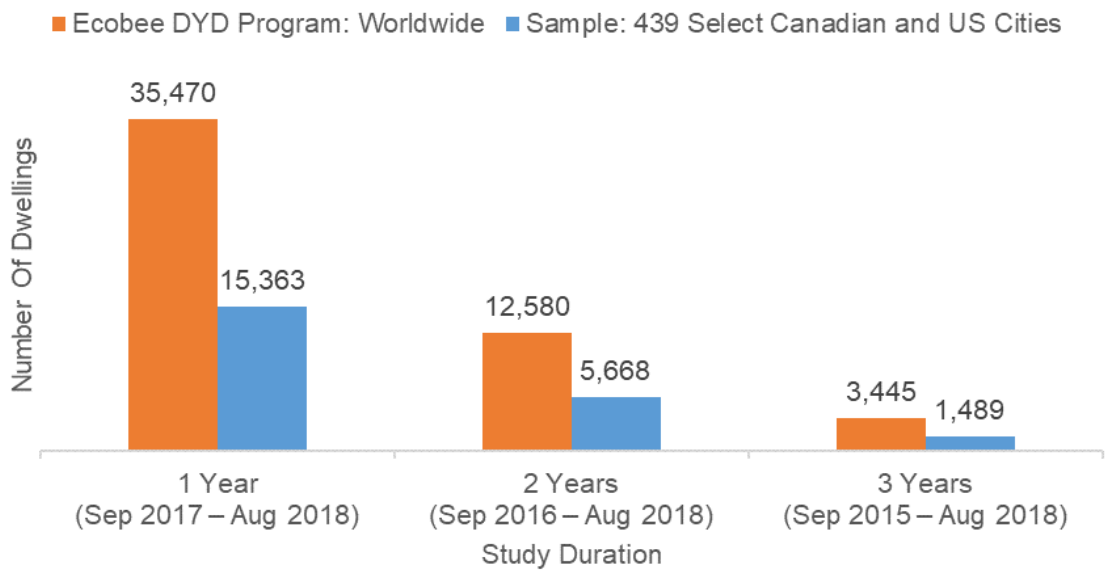


Figure 3.4: The number of DYD dwellings with a full set of monthly data files for a 1-year, 2-year and 3-year sample periods.

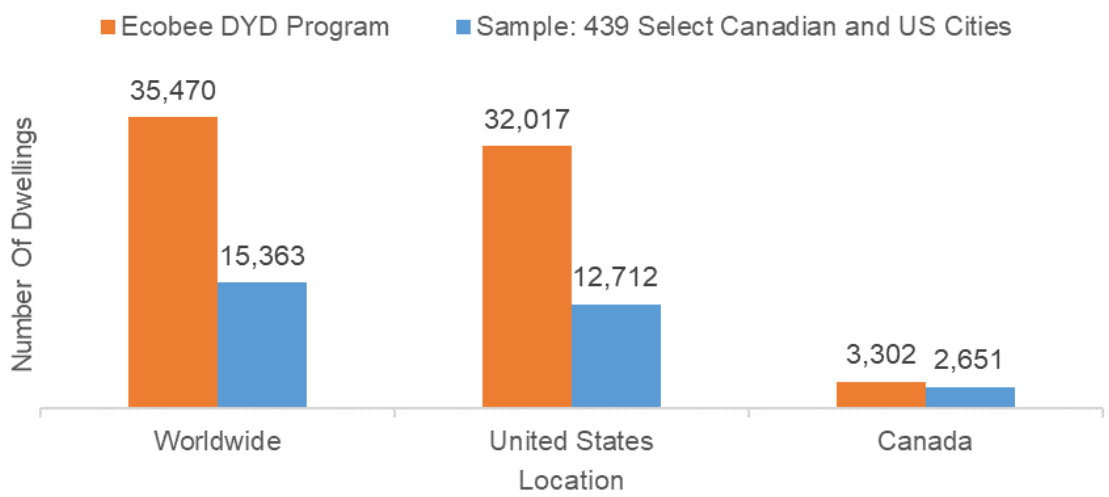


Figure 3.5: The sample considered for the study consists of the 439 cities, with the most DYD thermostat users, out of the total 6,620 cities across Canada and the US

A new dwelling characteristic, climate zone, was assigned using the geographical location of the dwellings. The impact of local construction practices on the time constant will be analyzed by comparing the thermal response of dwellings according to climate zones defined by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). Different building materials and techniques are required in different climate zones in order to ensure the durability and efficiency of a building enclosure. The ASHRAE climate zones displayed in Figure 3.6 are defined based on temperature and moisture, and can be found in Table Annex1-4 of the ASHRAE Standard 90.2 [61].

Table Annex1-4 Thermal Climate Zone Definitions

Thermal Zone	Name	I-P Units	SI Units
0	Extremely hot	10,800 < CDD50°F	6000 < CDD10°C
1	Very hot	9000 < CDD50°F ≤ 10,800	5000 < CDD10°C ≤ 6000
2	Hot	6300 < CDD50°F ≤ 9000	3500 < CDD10°C ≤ 5000
3	Warm	CDD50°F ≤ 6300 and HDD65°F ≤ 3600	CDD10°C < 3500 and HDD18°C ≤ 2000
4	Mixed	CDD50°F ≤ 6300 and 3600 < HDD65°F ≤ 5400	CDD10°C < 3500 and 2000 < HDD18°C ≤ 3000
5	Cool	CDD50°F ≤ 6300 and 5400 < HDD65°F ≤ 7200	CDD10°C ≤ 3500 and 3000 < HDD18°C ≤ 4000
6	Cold	7200 < HDD65°F ≤ 9000	4000 < HDD18°C ≤ 5000
7	Very cold	9000 < HDD65°F ≤ 12600	5000 < HDD18°C ≤ 7000
8	Subarctic/arctic	12600 < HDD65°F	7000 < HDD18°C

Figure 3.6: The ASHRAE international climate zone definitions

The number of each climate zone is an indication of temperature conditions and space conditioning requirements — zone 0 being the warmest and zone 8 being to coldest. The cooling degree days (CDD) and heating deegree days (HDD) are measurements designed to quantify the energy demand needed for space-conditioning. Based on the average cooling and heating degree days in Figure 3.6, climate zones 1 through 3 are typically cooling-dominated and 4 through 8 are typically heating-dominated.

A total of 439 cities across Canada and the US were sorted manually into 8 temperature divisions of the ASHRAE climate zones (see Figure 3.7). The climate zone assignments are based on information provided by Table Annex1-1, Table Annex1-2, Figure Annex1-1 and Figure Annex1-2 of ASHRAE Standard 90.2, as well as an online database developed by the National Renewable Energy Laboratory (NREL) [61], [62]. Due to the time constraints in relation to manually sorting the cities, only 439 cities out of a total of 6,620 were given a climate zone assignment (see Figures 3.4, 3.5 and 3.8). Examples of cities for each climate zone number are provided in Table 3.1.

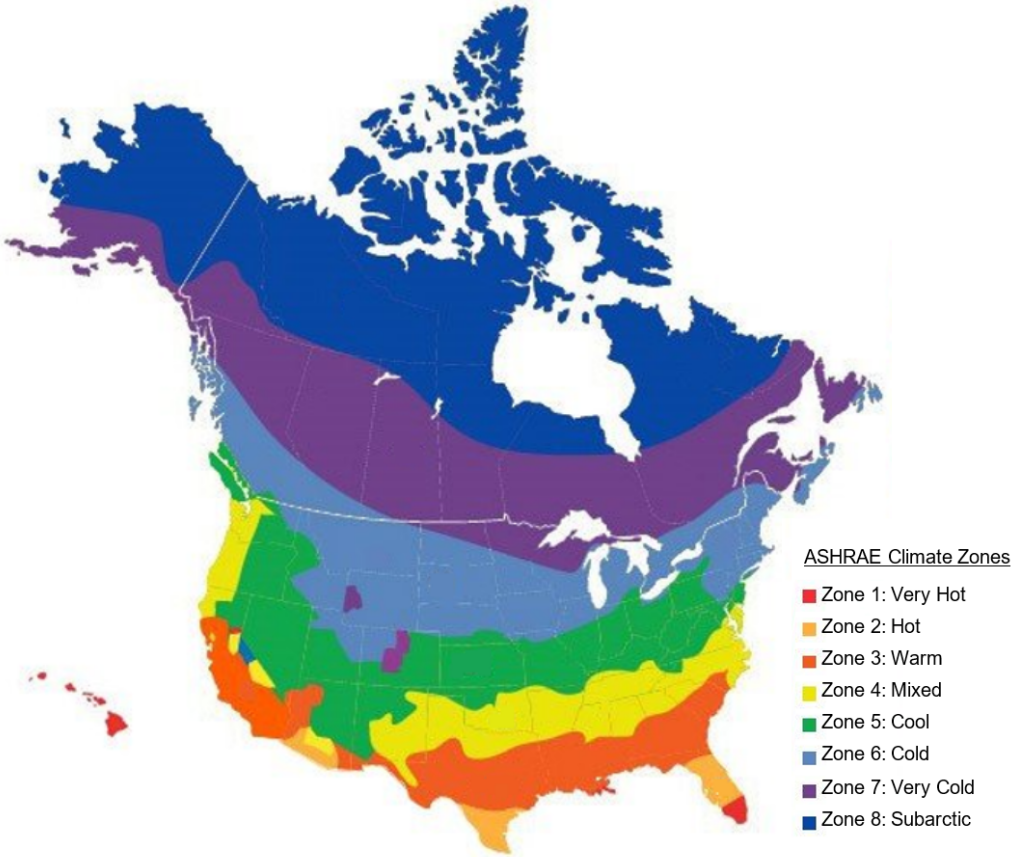


Figure 3.7: A map showing the boundaries of each ASHRAE climate zone found within Canada and the US [63].

Table 3.1: City examples for each climate zone number

Zone Number	United States	Canada
1	Honolulu, HI; Miami, FL	N/A
2	Phoenix, AZ; Orlando, FL	N/A
3	Los Angeles, CA; Atlanta, GA	N/A
4	Seattle, WA; Washington, DC	Victoria, BC
5	Columbus, OH; Chicago, IL	Kelowna, BC; Niagara Falls, ON
6	Minneapolis, MN; Madison, WI	Lethbridge, AB; Toronto, ON; Montreal, QC
7	Anchorage, AK	Edmonton, AB; Trois-Rivières, QC
8	Fairbanks, AK	Yellowknife, NT; Kuujjuaq, QC

In regards to selection, cities with the largest number of residences were given precedence over others in order to provide the largest available sample for the study. Cities belonging

to climate zones with less registered dwellings were also given priority in order to build up a balanced more representative sample. All the missing metadata entries for geographical location were filtered out in the city selection process. For a better understanding of the climate zone distribution, Figure 3.8 shows the number of dwellings according to each climate zone. Climate zones 1, 7, and 8 have significantly less data available which may impact the study's ability to offer a fair comparison for these climates.

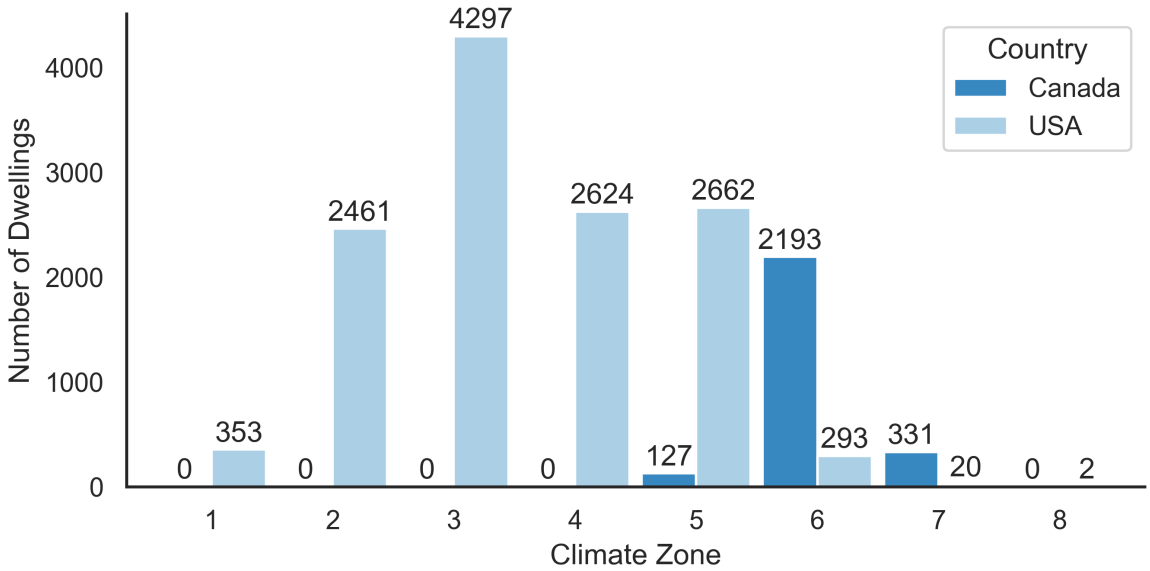


Figure 3.8: The distribution of the DYD dwellings based on ASHRAE climate zone and country.

In order to consider the largest sample of Canadian and US dwellings for this study, the group of 15,363 dwellings is selected and one year of operational data between September 2017 and August 2018 will be analyzed.

Metadata Reduction

Finally, the clean metadata is reduced in volume by eliminating the dwelling characteristics that will not be further investigated. Among the characteristics found in Table A.2, those retained for further analysis include: identifier, country, ASHRAE climate zone, dwelling type, number of floors, number of occupants, construction age and floor area. An exploratory data analysis for the study sample is provided in AppendixA.3.

3.4 Identification of Models through Data Mining

3.4.1 Estimating the Thermal Time Constant of a Dwelling

Building Energy Model

In order to determine the TTC, the first step is to approximate every dwelling as a single-zone thermal space represented by a first-order resistance-capacitance (RC) model (see Figure 3.9). For the purpose of this study, a dwelling is a residential unit, such as a building or part of a building, used by one household as a home. While the first-order model is a very simple model, it can provide a useful characterization of the dynamic thermal behaviour of a dwelling [64].

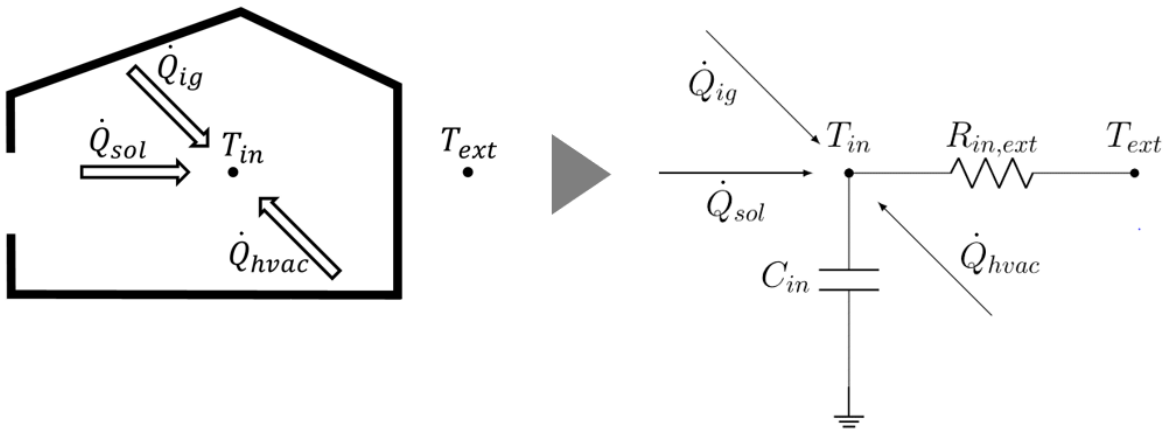


Figure 3.9: A simple diagram of a dwelling approximated as a single thermal zone and the first-order RC thermal network used to represent it.

The first-order model in this study uses a single node to represent the dwelling's indoor environment and summarizes its thermal response using only two parameters: the effective values for its overall thermal resistance ($R_{\text{in,ext}}$) and its overall thermal capacitance (C_{in}). When a temperature difference exists between a dwelling and its surroundings, heat continues to flow between the two until equilibrium is achieved; at which time, the indoor temperature (T_{in}) will assume the same value as the outdoor temperature (T_{ext}). Between the dwelling and the outdoor environment, heat flow is modelled as a current that travels through a resistor with the thermal resistance $R_{\text{in,ext}}$. Heat will either flow out of the dwelling into its colder surroundings (i.e. cool-down mode) or into the dwelling from its warmer surroundings (i.e. warm-up mode); the dwelling's thermal resistance limits these heat exchanges with the surroundings. In cool-down mode, the heat flow travels from the interior to the exterior, discharging the dwelling's thermal mass, modelled as capacitor with the thermal capacitance C_{in}. In contrast, in warm-up mode,

the heat flow travels in the opposite direction, charging the capacitor. With respect to time (t), the relationship between the change in the dwelling's total internal energy (E_{i}) and the change in its indoor temperature is given by:

$$\begin{aligned} \frac{dE_{\mathrm{i}}}{dt} &= \rho c_{\mathrm{p}} V \frac{dT_{\mathrm{in}}}{dt} \\ &= C_{\mathrm{in}} \frac{dT_{\mathrm{in}}}{dt} \end{aligned} \tag{3.1}$$

where ρ is the dwelling's effective density, c_{p} is the dwelling's effective specific heat capacity, V is the dwelling's total volume.

As seen in Figure 3.9, the thermal needs of most dwellings are satisfied by a symbiotic application of both active and passive thermal control systems. With the active systems ($\dot{Q}_{\mathrm{hvac}}$) shut off, a dwelling operates under free-running conditions; the net heat transfer rate ($\dot{Q}_{\mathrm{net}}$) into the dwelling depends on heat gains and losses linked to the indoor-outdoor temperature difference, the sun ($\dot{Q}_{\mathrm{sol}}$), and the indoor environment (i.e the occupants, the appliances and the lighting) ($\dot{Q}_{\mathrm{ig}}$). Under free-running conditions, its indoor temperature, assumed to be uniform throughout the space, is regulated by the performance of the passive thermal control systems represented in the 1R1C model by the $R_{\mathrm{in,ext}}$ and C_{in}. When the dwelling's passive systems are unable to maintain the indoor temperature within an ideal range for the dwelling's needs — whether occupied or unoccupied — the active systems are run to supplement them. Applying the first law of thermodynamics to the node representing the dwelling's indoor environment gives the following relationship between the dwelling's change in internal energy and its heat flow into the system:

$$\begin{aligned} \dot{Q}_{\mathrm{net}} &= \frac{dE_{\mathrm{i}}}{dt} \\ &= C_{\mathrm{in}} \frac{dT_{\mathrm{in}}}{dt} \end{aligned} \tag{3.2}$$

where $\dot{Q}_{\mathrm{net}}$ is net heat transfer rate into the dwelling considering all heat gains (i.e. from the sun, internal gains and indoor-outdoor temperature difference). The sign convention for $\dot{Q}_{\mathrm{net}}$ as heat is flowing into the dwelling is positive.

Developing Equation 3.2 further, the heat dynamics for the 1R1C model in Figure 3.9 are expressed by the following differential equation:

$$\begin{aligned} C_{\text{in}}\frac{dT_{\text{in}}}{dt} &= \frac{A}{R'}(T_{\text{ext}} - T_{\text{in}}) + \frac{dQ_{\text{sol}}}{dt} + \frac{dQ_{\text{ig}}}{dt} + \frac{dQ_{\text{hvac}}}{dt} \\ &= \frac{1}{R_{\text{in,ext}}}(T_{\text{ext}} - T_{\text{in}}) + \dot{Q}_{\text{sol}} + \dot{Q}_{\text{ig}} + \dot{Q}_{\text{hvac}} \end{aligned} \tag{3.3}$$

where A represents the heat transfer surface area, t represents time elapsed in seconds and R' is the effective R-value.

Operating under free-running conditions (i.e. $\dot{Q}_{\text{hvac}} = 0$) at night (i.e $\dot{Q}_{\text{sol}} = 0$) when the internal heat gains at night are assumed to be negligible, the dwelling's natural thermal response can be described using the curves shown in Figure 3.10. The natural thermal response of the free-running dwelling represents the performance of inherent properties (i.e. its insulation, airtightness, and thermal mass) in response to outdoor temperature conditions, and can be expressed by rewriting Equation 3.3 in the following manner:

$$\begin{aligned} \frac{dT_{\text{in}}}{dt} &= \frac{T_{\text{ext}} - T_{\text{in}}(t)}{R_{\text{in,ext}}C_{\text{in}}} \\ &= \frac{T_{\text{ext}} - T_{\text{in}}(t)}{\tau} \end{aligned} \tag{3.4}$$

where τ is the TTC and is the product of $R_{\text{in,ext}}$ and C_{in} (see Equation 2.1).

Using the initial condition $T_{\text{in}}(0) = T_{in,o}$, the solution of Equation 3.4 is as follows:

$$\begin{aligned} T_{\text{in}}(t) &= T_{\text{ext}} + (T_{in,o} - T_{\text{ext}}) \cdot e^{-t/\tau} \\ &= T_{\text{ext}} + \Delta T_{\text{o}} \cdot e^{-t/\tau} \end{aligned} \tag{3.5}$$

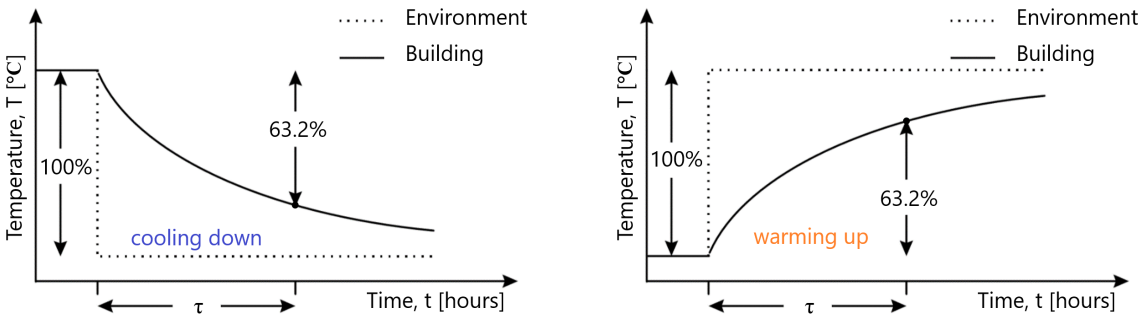
where $T_{in,o}$ is the initial indoor temperature and ΔT_{o} is the initial indoor-outdoor temperature difference. The derivation of this equation can be found in Section B.2.

When $t = \tau$, Equation 3.5 therefore gives:

$$\begin{aligned} T_{\text{in}}(\tau) &= T_{\text{ext}} + (T_{in,o} - T_{\text{ext}}) \cdot e^{-1} \\ &= T_{\text{ext}} + \Delta T_{\text{o}} \cdot \sim 0.368 \end{aligned} \tag{3.6}$$

This equation shows that the indoor-outdoor temperature difference at $t = \tau$ is equal to 36.8% of ΔT_{o}. As depicted in Figure 3.10, the TTC is therefore the time required for the indoor

temperature of a free-running dwelling to move 63.2% (i.e. $1 - e^{-1}$) of the way from its initial temperature towards its final temperature (i.e. the temperature of its surroundings).



(a) Indoor temperature decay curve in response to a colder environment.

(b) Indoor temperature growth curve in response to a warmer environment.

Figure 3.10: Modelled by a 1R1C model, the exponential thermal response of a free-running dwelling to a sudden step change in the outdoor temperature.

Criteria for Selection of Analysis Periods

Since the first-order model is an approximation of reality built upon assumptions, the second step is the identification of suitable time periods for analysis which satisfy these assumptions.

The analysis periods used for the estimation of TTC will be limited to the nighttime; when the sun is down, occupants are less likely to be very active, and equipment (such as appliances and lighting) are less likely to be running. Without the heat losses and gains due to space conditioning and solar, and with negligible internal heat sources, the TTC estimated in the study represents the performance of the insulation, airtightness and thermal mass of a dwelling in response to outdoor temperature conditions.

In this study, analysis periods are therefore selected according to following basic criteria:

1. the analysis periods occur at night (i.e. between sunset and sunrise);
2. the house is under free-running conditions (i.e. when the HVAC system is switched off) for more than one hour;
3. the outdoor temperature remains relatively constant (i.e. the change in outdoor temperature is smaller than or equal to 2°C).

As previously mentioned, the data is organized by month for the study duration. For each location, a monthly sunset and sunrise time are determined based on the longest day of the

month. A reference CSV file of these sunrise and sunset times, according to location and month, is created using the Python function *geopy.geocoders.Nominatim*, *Geopy*, and a Python package for time zone identification, *timezonefinder*. The measured time series data for HVAC equipment runtimes and the outdoor temperature are used to determine if the criteria are met during each analysis period.

Estimation of Non-Linear Model Parameters Using Least Squares Regression

The third step is the estimation of a TTC value for each identified analysis period, assuming the indoor temperature variations would follow the characteristic exponential curve of a first order system. Based on Equation 3.5, the indoor temperature of dwelling with respect to time may be determined using:

$$T_{\text{in}}(t) = (T_{\text{ext}})_{j,k} + (\Delta T_{\text{o}})_{j,k} \cdot e^{-t/\tau_{j,k}} \tag{3.7}$$

where k is the index for each individual dwelling (k= 1, 2, 3, ..., 15363), j is the index for each analysis period (j= 1, 2, 3, ..., n), and n is the total number of analysis periods over the course of one month for dwelling k. Based on the indoor temperature observations from analysis period j, $(T_{\text{ext}})_{j,k}$ is the estimated outdoor temperature, $(\Delta T_{\text{o}})_{j,k}$ is the estimated initial indoor-outdoor temperature difference, $\tau_{j,k}$ represents the estimated TTC value.

Using measured time-series data for indoor temperature and non-linear regression analysis, the following three parameters of the regression equation are calculated with the use of a Python optimization function *scipy.optimize.curve_fit*: (a) $(T_{\text{ext}})_{j,k}$, (b) $(\Delta T_{\text{o}})_{j,k}$, and (c) $\tau_{j,k}$. This function uses non-linear least squares to fit Equation 3.7 to the observed indoor temperature measurements by determining the set of input parameters that minimizes the residuals. The optimization cost function used is as follows:

$$cost_i = Y_i - \hat{Y_i} \tag{3.8}$$

where i is the index for each observation in the analysis period j, Y_i is the predicted value for observation i, and $\hat{Y_i}$ is the actual value for observation i.

The optimization requires the input of initial values for the parameter set as a mere starting points for the optimization process. For $(T_{\text{ext}})_{j,k}$, the initial input is the average of the recorded outdoor temperature values over the analysis period *j*. For $(\Delta T_{\text{o}})_{j,k}$, the initial input is the difference between T_{in_0} and and the initial input for T_{ext}. For $\tau_{j,k}$, a arbitrary value of 50 is used

as the initial input and it is based on previous experimentation with the Python function with a smaller sample of data [65].

Figures 3.11 and 3.12 illustrate examples of curve fitting performed using Equation 3.7 on data during a winter analysis period and a summer analysis period, respectively.

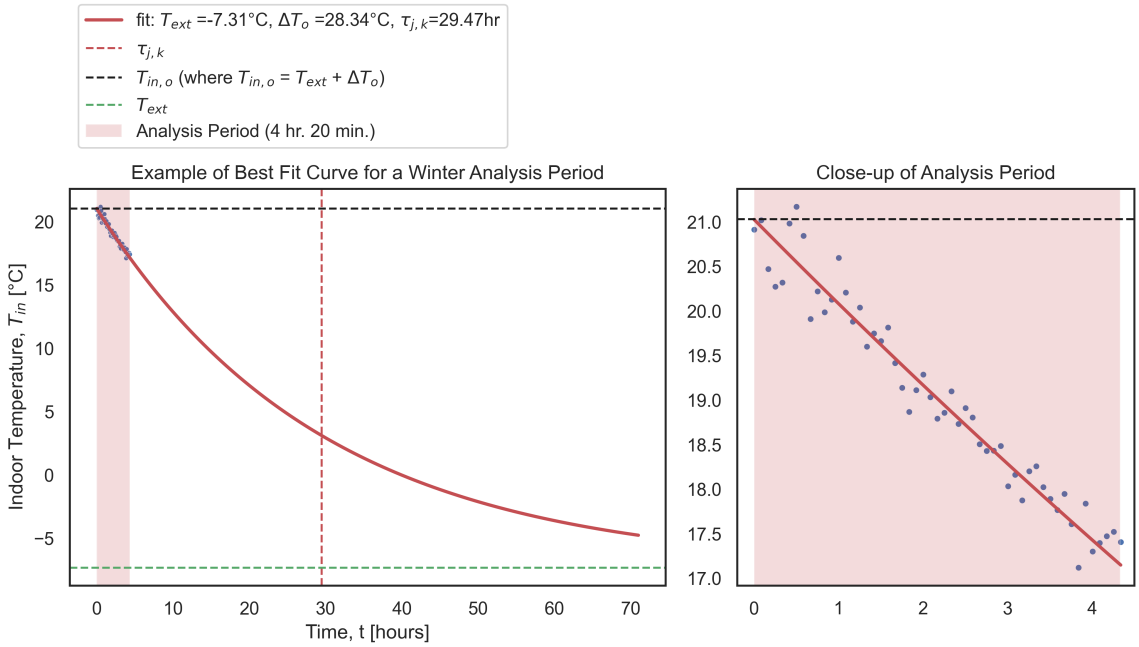


Figure 3.11: A curve fitting example for free-running indoor temperature data from an analysis period in winter.

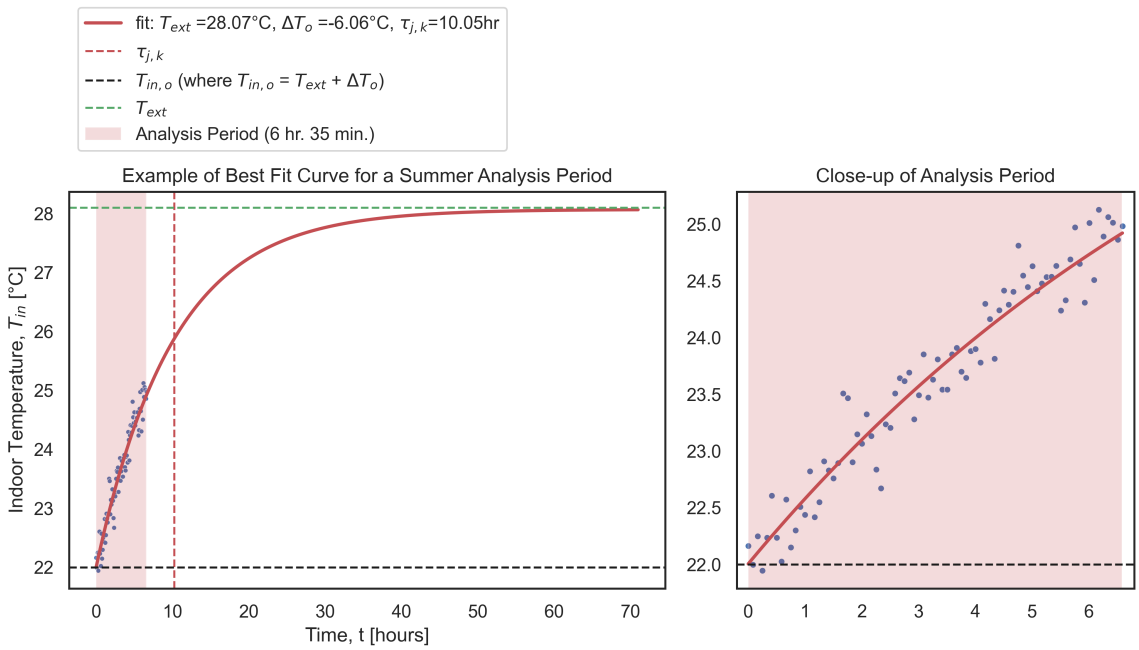


Figure 3.12: A curve fitting example for free-running indoor temperature data from an analysis period in summer.

Model Performance Criteria

The fourth step is to detect and eliminate the estimated $\tau_{j,k}$ values that may be considered inaccurate and/or outliers. The inaccurate $\tau_{j,k}$ values are detected according to their corresponding Root-Mean-Square Error $RMSE_j$ value and corresponding calculated $(T_{\text{ext}})_{j,k}$ value. The outliers are found using the Interquartile Range (IQR) method.

The RMSE is a criterion commonly used to evaluate the fit of regression models using least squares. For every estimation of the $\tau_{j,k}$, the $RMSE_j$ is a measure of how close the regression model's predicted values are to the empirical indoor temperature values. The $RMSE_j$ is calculated using the following equation:

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}} \tag{3.9}$$

where j is the index for each analysis period, i is the index of each observation (i= 1, 2, 3, ..., n), n is the total number of observations in the analysis period j, Y_i is the predicted value for observation i, and $\hat{Y}_i$ is the actual value for observation i. The lower the value of the $RMSE_j$ the better the fit of the model to the indoor temperature measurements. The estimated $\tau_{j,k}$ values retained are those with Root-Mean-Square Error (RMSE) values equal to and lower than 0.50°C. The selected maximum $RMSE_j$ value of 0.50°C was determined with the use of data visualizations and based on the accuracy of the smart thermostats. Next, the calculated $(T_{\text{ext}})_{j,k}$ value of each $\tau_{j,k}$ value are monitored to ensure that unexpected internal gains are not driving the indoor temperature fluctuations. If the calculated $(T_{\text{ext}})_{j,k}$ value lies more than 5 °C away from the initial input of $(T_{\text{ext}})_{j,k}$ value, its associated $\tau_{j,k}$ value is removed from the dataset.

The remaining estimated $\tau_{j,k}$ values are inspected to determine whether any of them are outliers. Within the context of a dataset, an outlier is a extreme value that differs significantly from the other observations. According to the IQR method, a value that lies further than 1.5 times IQR away from the mean (i.e. about 3 standard deviations) on either side can be classified as an outlier and eliminated from the dataset [54].

3.4.2 Identification of Typical Thermal Time Constants

Monthly Thermal Time Constants

In section 3.4.1, multiple analysis periods are found for each dwelling k over the course of one month. For each month mth, a weighted average is calculated for each dwelling k using

their corresponding $\tau_{j,k}$ values; the weighted average, $\tau_{mth,k}$ is calculated using the following equation:

$$\tau_{mth,k} = \frac{\sum_{j=1}^{n} w_j \tau_{j,k}}{\sum_{j=1}^{n} w_j} \tag{3.10}$$

where mth is the index for each month of the year (mth= Jan, Feb, ..., Dec), k is the index for each individual dwelling (k= 1, 2, 3,..., 15363), j is the index for each analysis period (j= 1, 2, 3, ..., n), n is the total number of analysis periods in month mth for dwelling k, and weighting factor w_j is simply the reciprocal of $RMSE_j$. The use of w_j gives more weight to $\tau_{j,k}$ values that are more reliable. For each residence, potentially 12 $\tau_{mth,k}$ values can be calculated.

Distribution Fitting

Distribution fitting is used to determine whether typical $\tau_{mth,k}$ values for a dwelling can be identified based on the season and its ASHRAE climate zone. For the purposes of the study, the seasons start on first day of the months that include the equinoxes and solstices:

- Fall runs from September 1 to November 30;
- Winter runs from December 1 to February 28 (or 29);
- Spring runs from March 1 to May 31;
- Summer runs from June 1 to August 31.

Fitting a Probability Density Function (PDF) to the data offers a simpler way of approximating how many dwellings would have a given value of $\tau_{mth,k}$ using only a few parameters. The PDF is a statistical distribution often represented by a line graph that displays probability versus the variable. The probability represented in the PDF is what percentage of the sample observations for a variable, such as $\tau_{mth,k}$, are expected to occur at a given value of the variable. Statistical distributions are often used as theoretical models of real-world data; they do not necessarily fit the data perfectly but they provide a reasonable representation of data.

First, for each combination of climate zone and season with more than 60 $\tau_{mth,k}$ values [54], a range of different statistical distributions are fit to the observed distribution of values using *scipy.stats.rv_continuous.fit*, a function from the Python's SciPy library which performs a Maximum Likelihood Estimate.

Second, a summary of statistics for each climate-season combination is calculated including the median, standard deviation and a 95% confidence interval range for the mean, $\tau_{ssn,cl}$. The ssn is the index for each season (ssn= fal, win, spr, sum) and the cl is the index for the ASHRAE climate zone (cl= 1, 2,... , 8).

Model Performance Criteria

Next, when fitting a statistical model to observed data, the chi-squared value (χ^2) is a measure used to compare the goodness of fit of each theoretical distribution to the observed distribution [54]. In order to calculate χ^2 the data for each climate-season combination are binned into 20 bins based on percentiles so that each bin contains approximately an equal number of data observations. The χ^2 is calculated as follows:

$$\chi^2_{ssn,cl} = \sum_{h=1}^{n} \frac{(S_h - \hat{S}_h)^2}{S_h} \tag{3.11}$$

where h is the index of each bin (h= 1, 2, 3,..., n), n is the total number of bins for each climate-season combination, S_h is the cumulative sum of predicted frequencies for bin range 1 through h, and $\hat{S}_h$ is the cumulative sum of observed frequencies for bin range 1 through h.

In addition, the Kolmogorov-Smirnov (KS) test is performed using the function *scipy.stats.kstest* from the Python's SciPy library [66]. The KS test is used to evaluate the suitability of the theoretical distribution and determine if there are any significant differences between observed and fitted distribution. Ideally, a p-value of greater than 0.05 is obtained which means that the fitted distribution is not significantly different to the observed distribution of the data.

Chapter 4

Results and Discussion

In this section, the results of the parameter estimation using regression analysis are explored and the effect of the following variables on a dwelling's thermal response is investigated: ASHRAE climate zone and season. Using data mining and statistical methods, patterns and trends are identified in the thermal responses of 15,363 Canadian and U.S. dwellings monitored over the course of one year. Data visualizations are used to display the distribution of the estimated monthly Thermal Time Constant (TTC)s ($\tau_{mth,k}$) and the sample subset that yielded them, in relation to the previously mentioned variables.

4.1 Exploratory Data Analysis: Estimated Thermal Time Constants for Dwellings

Respecting the criteria set for the regression model and the performance criteria (see Section 3.4.1), the $\tau_{j,k}$ was calculated for only night periods in order to diminish the influence of solar gains and occupant behaviour on the dwelling's thermal response to outdoor temperature fluctuations. With over 15,000 dwellings being considered, the algorithm that estimated the $\tau_{j,k}$ values for each month was computationally intense. For one full year of data, $\tau_{j,k}$ was calculated for 235,024 analysis periods. The average duration of an analysis period is 7.14 hours and Figure 4.1 shows that the analysis periods vary in length from 1.00 to 17.92 hours. For each analysis period, the indoor temperature response curve of the free-running dwelling is represented by a first-order model and its characteristic exponential curve (see Equation 3.7) is fit to the measured indoor temperature data. Each regression analysis performed has a corresponding $RMSE_j$ value; the average of this performance indicator is 0.11°C and its distribution is presented in Figure 4.2.

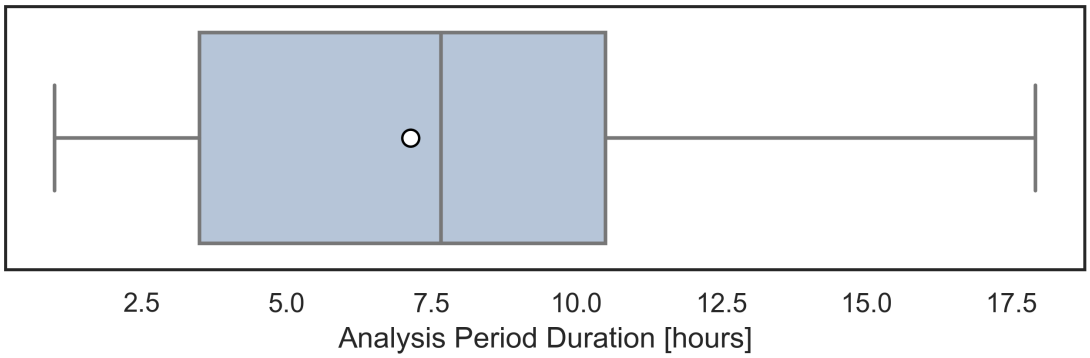


Figure 4.1: The distribution of duration for over 235,000 analysis periods.

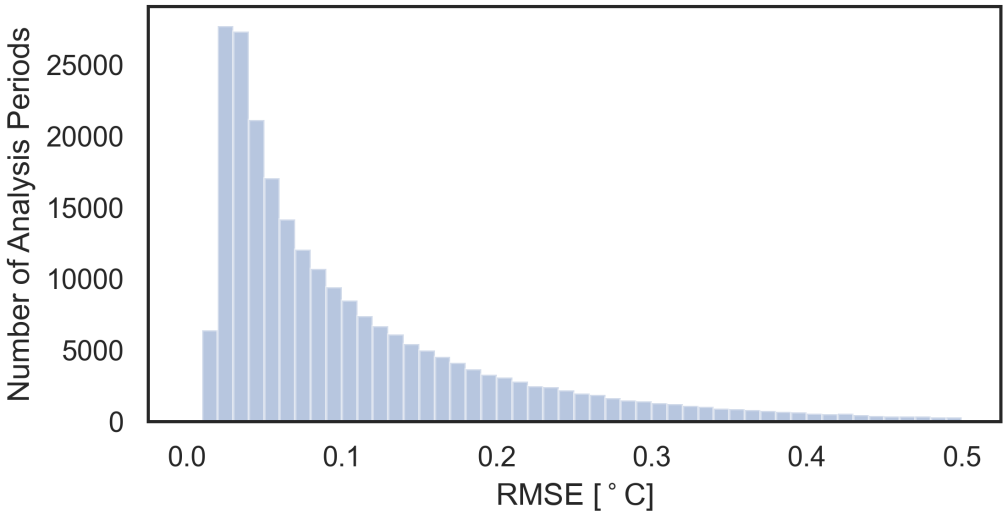


Figure 4.2: The distribution of $RMSE_j$ which is a performance criterion provided for each analysis period j.

In addition, Figure 4.3 demonstrates that the distribution of the estimated $\tau_{j,k}$ values is positively skewed with values ranging from 3.00 to 239.85 hours; the majority of the values fall toward the lower side of the scale, with an average value of 18.40 hours and a median of 7.67 hours. The TTC value significantly depends on season and climate; these relationships will be further explained later in this chapter.

For each dwelling, a monthly weighted average of the $\tau_{j,k}$ values ($\tau_{mth,k}$) was calculated for the months where analysis periods were found. The 41,289 estimated $\tau_{mth,k}$ values represent 11,740 dwellings from the original sample and their distribution is also positively-skewed as seen in Figure 4.4. The analysis of these values can be found in the following section.

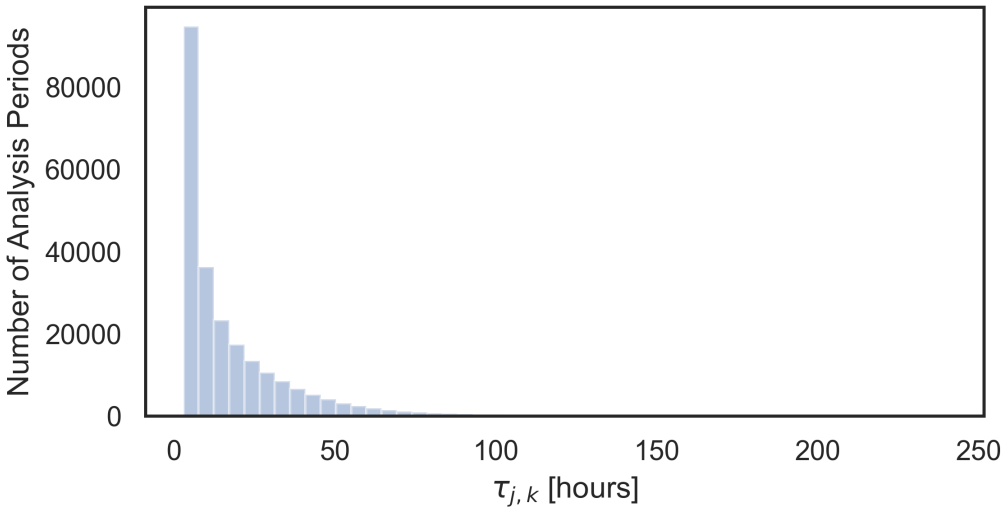


Figure 4.3: The distribution of the $\tau_{j,k}$ values resulting from the regression analyses.

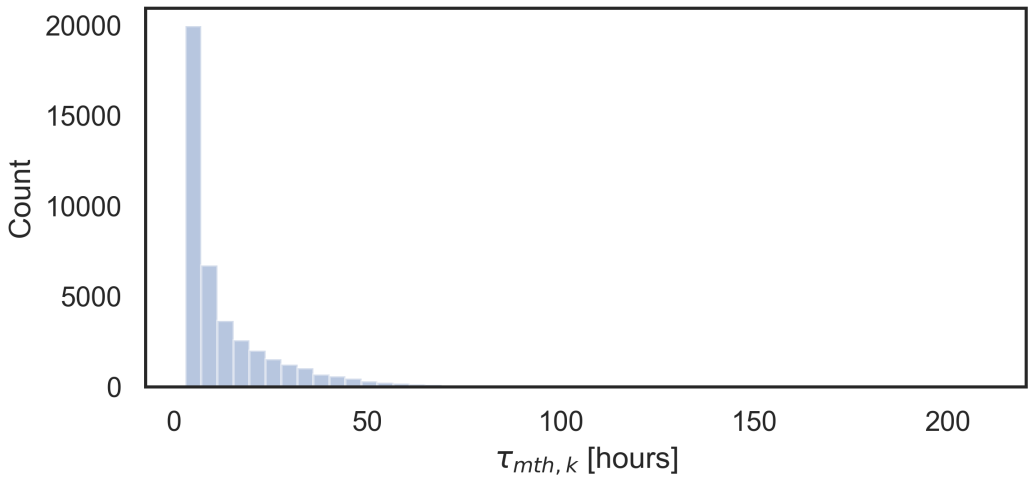


Figure 4.4: The distribution of over 41,000 monthly TTC values, $\tau_{mth,k}$, estimated over the course of one year for over 11,000 dwellings.

4.1.1 The Effects of Location and Seasonality

This section covers the analysis of the monthly TTC values, $\tau_{mth,k}$, in relation to climate zone and season. A number of insights have been found related to the impact of location and seasonality on the passive thermal performance of the Canadian and US dwellings.

Data Distribution across Climate Zones

First, it is important to note that the number of dwellings that provided data is not evenly distributed across all climate zones. As seen previously in Figure 3.8, Climate zones 1, 7, and 8 make up less than 5% of the dwellings yielding $\tau_{mth,k}$ values and consequently these zones have significantly less data available than the other zones. As a result, climate zone 8 will not be considered when analyzing the impact of season and climate on the TTC values. In regards to climate zones 1 and 7, the study's ability to offer a fair representation of the average dwelling's thermal response may be limited. Tables C.1, C.2, through C.5 offer a summary of key statistics for the $\tau_{mth,k}$ values organized according to climate zone, for full year as all for each individual season.

Likelihood of Finding Free-running Conditions

Second, the results suggest that the likelihood of a dwelling entering free-running mode can be linked to its climate zone and season. Figure 4.5 compares the percentage of dwellings from the original sample (i.e. 15,363), according to climate zone, that yielded at least one $\tau_{mth,k}$ value. For cooling-dominated zones, the percentage of dwellings increases (from about 83% to 85%) from zone 1 to 3. For the heating-dominated zones, this percentage steadily decreases (from about 80% to 50%) from climate zone 4 to 7. This observed pattern reasonably suggests that the greater the temperature differences occurring between the indoor and outdoor environment the lower the chances of finding a dwelling under free-running conditions.

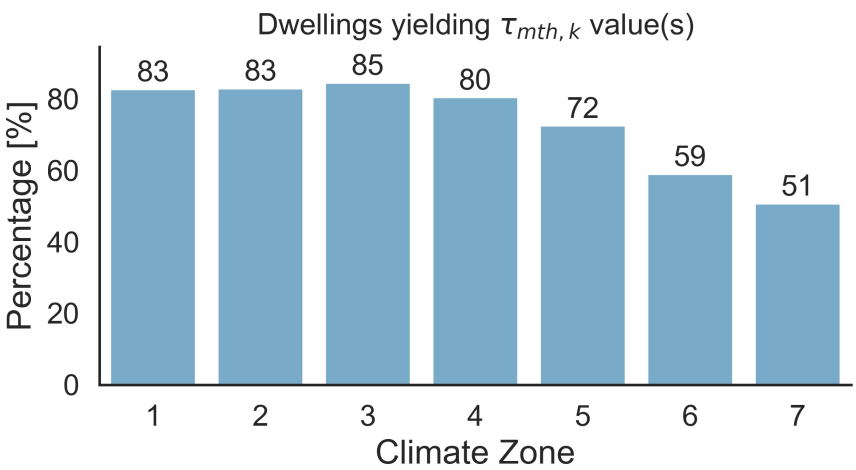


Figure 4.5: A comparison of the dwelling count of the sample subset that yielded monthly TTC values ($\tau_{mth,k}$) to the original sample, organized according to climate zone.

For each climate zone, Figure 4.6 provides, the percentage of the $\tau_{mth,k}$ values estimated in each season. For each climate zone, the season providing the highest percentage of $\tau_{mth,k}$ values reflects the most favorable time of year for finding appropriate analysis periods, and the season providing the lowest percentage of $\tau_{mth,k}$ values reflects the season dominated by the use of the HVAC system.

Concerning the summer months, the percentage of $\tau_{mth,k}$ values extracted during this season consistently increases from zone 1 to zone 6. In contrast, the percentage of $\tau_{mth,k}$ values extracted from the winter months shows decreasing pattern from zone 1 to zone 6. In regards to fall and spring, the cooling-dominated zones show no significant differences between the two seasons, however, for a reason unknown to the author, the heating-dominated zones consistently have a higher percentage of $\tau_{mth,k}$ values in the fall. As expected, finding free-running conditions for cooling-dominated zones is least likely in the summer months and for heating-dominated zones least likely in the winter months.

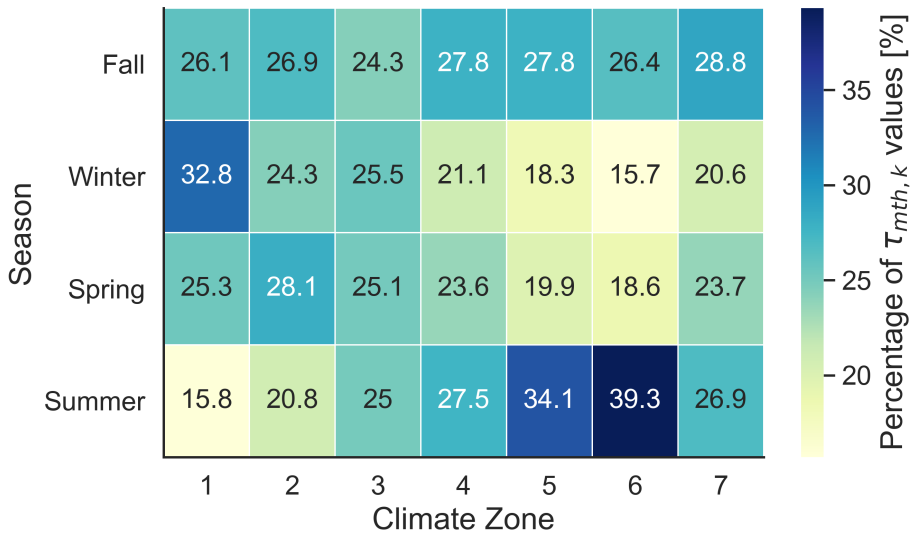


Figure 4.6: A count of monthly TTC values, $\tau_{mth,k}$, according to climate zone and season.

Climate-based and Seasonal Patterns in Thermal Time Constants

Third, a climate-based pattern is observed in the $\tau_{mth,k}$ values themselves that reflects the expected differences in insulation requirements based on location. Figure 4.7 displays an increasing trend in both the annual mean and median values of $\tau_{mth,k}$ from zone 1 to zone 7 (i.e. from warmer to colder climates). The dwellings built in colder climates therefore generally have longer estimated TTC than those built in warmer climates; these results reflect that the dwellings built in colder climates are usually prescribed higher thermal resistances.

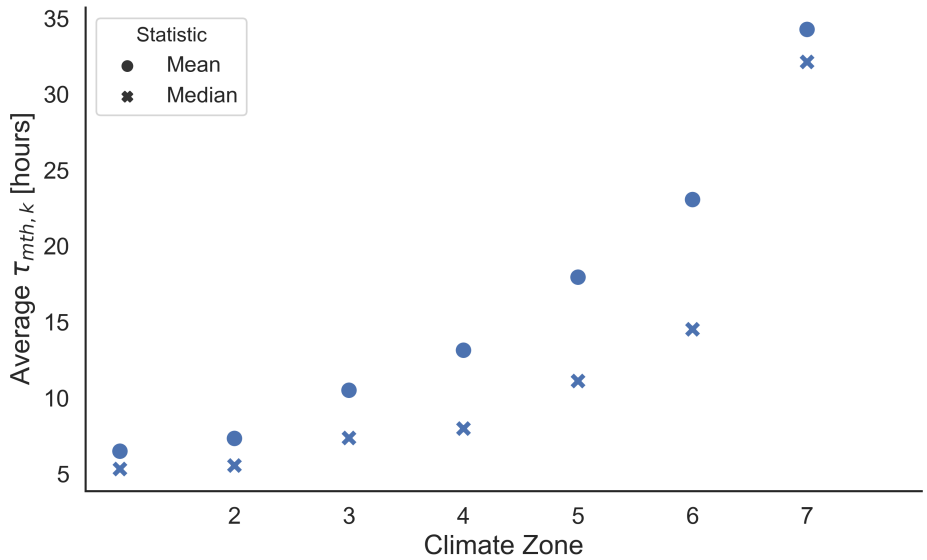


Figure 4.7: The annual mean and median values of the monthly TTCs ($\tau_{mth,k}$) with respect to ASHRAE climate zone (zones 1 through 7).

Next, Figure 4.8 provides both the mean and median values of $\tau_{mth,k}$ based on the season in which they were estimated in. From fall to summer, the mean and median values for $\tau_{mth,k}$ both follow an inverted U-shape pattern peaking in winter with the two low limits -one in fall and the other in summer. The lowest seasonal mean and median value for $\tau_{mth,k}$ occur in summer. The fact that the TTC of one building varies this significantly throughout the year is expected; this seasonal change in the value of $\tau_{mth,k}$ can be linked to the behavioural patterns of occupants in regards to opening their windows. During the summer months, it is reasonable to expect occupants to have their windows open at night and to leave them open for long periods of time due to more favorable weather conditions. Open windows increase the infiltration (or exfiltration) of air, decrease a building's ability to resist thermal changes in its surroundings, and thus reduces its overall effective thermal resistance ($R_{\text{in,ext}}$). The TTC is the product of the effective thermal resistance and the thermal capacitance of the dwelling (see Equation 2.1); the lower the $R_{\text{in,ext}}$ the faster heat can cross the boundary between the indoor and outdoor spaces, and the smaller the TTC will be. This is a seasonal pattern that is reflected not only in the overall sample but in each individual climate zone.

Finally, visualizing the combined effects of climate zone and season are even more telling; the influence of location and seasonality can be seen clearly in Figures 4.9 and 4.10 where the winter and summer values for $\tau_{mth,k}$ are compared across climate zones 1 through 7. For each zone, the winter and summer months – when the coldest or hottest temperatures occur, respectively – show significant differences between their resulting $\tau_{mth,k}$ distributions; the mean summer values range from 26% to 55% of their corresponding winter means.

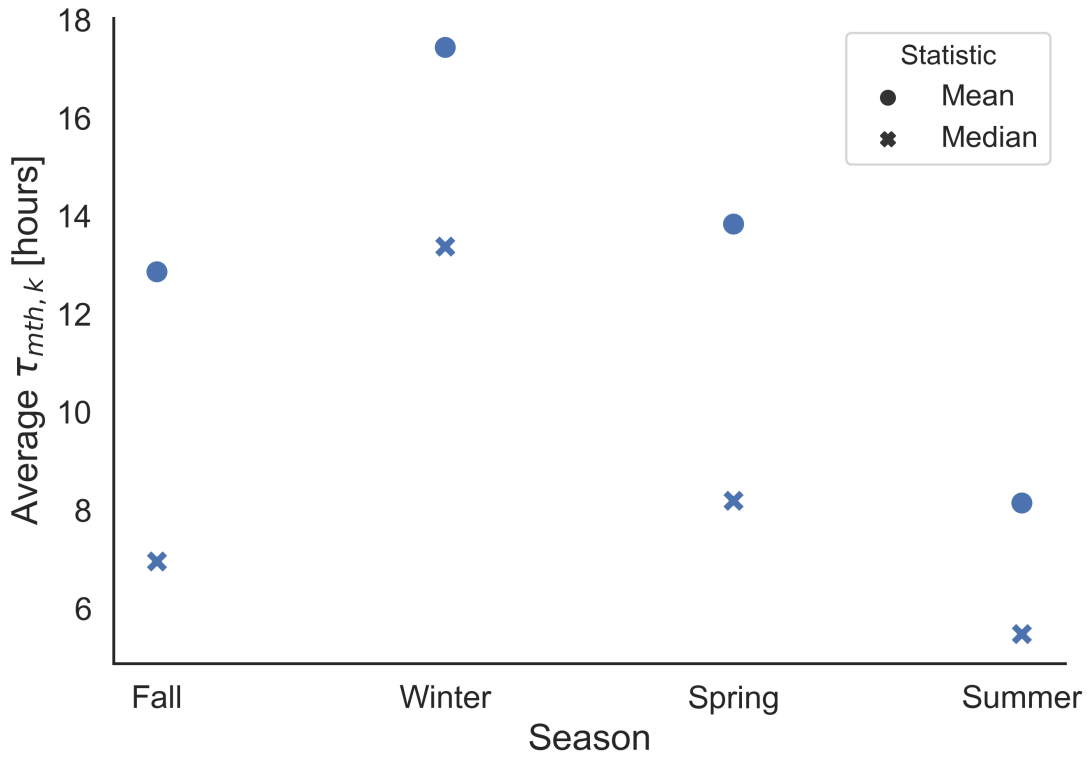


Figure 4.8: The mean and median values of the monthly TTCs ($\tau_{mth,k}$) with respect to season.

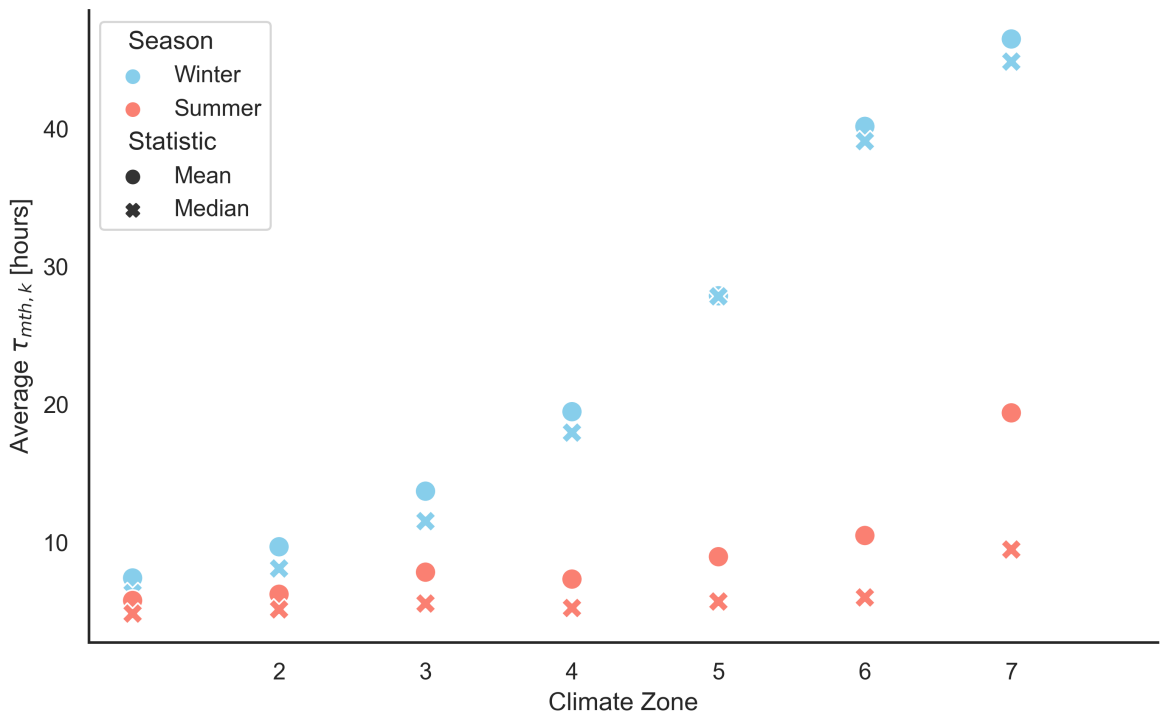


Figure 4.9: The annual mean and median values of the monthly TTCs ($\tau_{mth,k}$) for the winter and summer seasons with respect to climate zone.

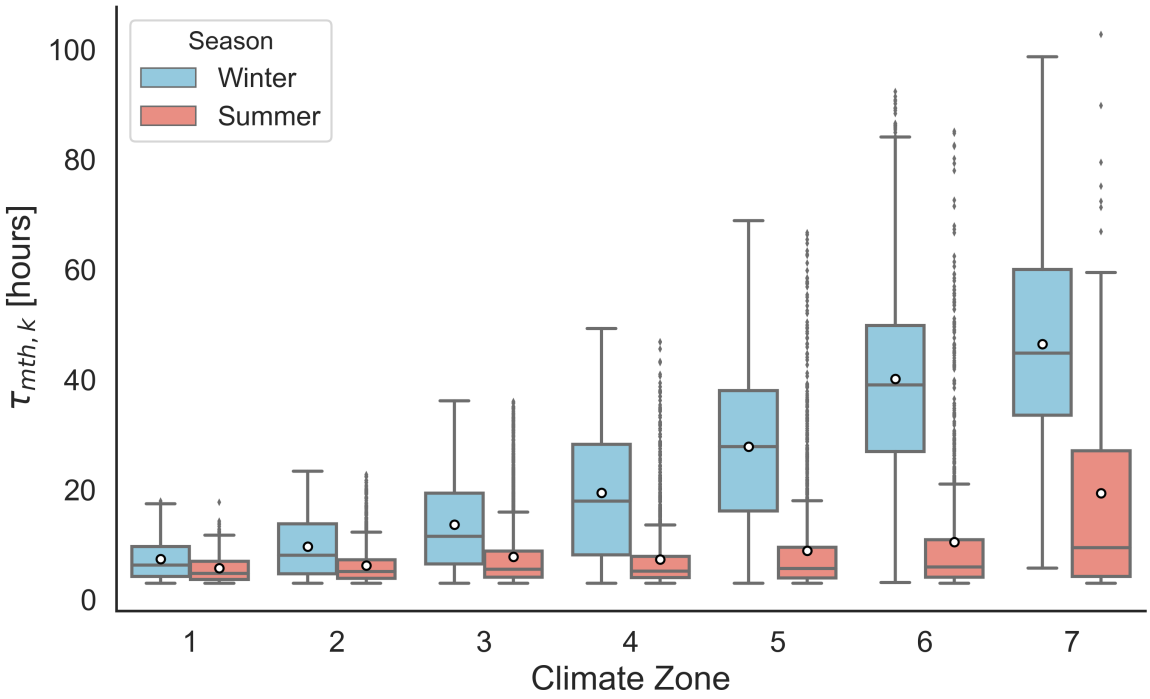


Figure 4.10: Comparing the winter and summer distributions of monthly TTCs, $\tau_{mth,k}$, according to climate zone.

The comparison of the distributions of $\tau_{mth,k}$ for all four seasons across climate zones 1 through 7 is also available (see Figure C.1). While the seasonal pattern of $\tau_{mth,k}$ is consistent in all climate zones, the cause of this pattern (i.e. occupants tending to have their windows open in the summer months) can only be speculated. Through the data provided by ecobee, it is not possible to test this hypothesis.

4.2 Typical Thermal Time Constants of Dwellings based on Climate and Season

Based on the data available and the observations made in section 4.1, statistical distributions may be used as a means of providing a snapshot of the current TTCs in the Canadian and US residential sectors; this thermal characterization of dwellings will be provided across climate zones 1 through 7 for each season. A total of 28 different climate-season combinations are represented in this study and the large amounts of $\tau_{mth,k}$ data associated to them have been simplified into a few parameters using distribution fitting.

For each climate-season combination, the best fit distribution for the data from the follow-

ing statistical models were found and compared: Normal, Lognormal, Johnson SB, Gamma, Exponential, Beta, Chi-Square, F, and Weibull. Although 10 out of the 28 combinations had p-values lower than 0.05, the Johnson SB distribution provides the most acceptable fit for each climate-season combination, among the distributions considered (see Table C.6 and Figures C.2 through C.11).

Johnson SB is a lognormal distribution which is often used to represent continuous variables with a non-negative nature whose distributions are skewed [54]. This specific lognormal distribution has 4 parameters making it extremely flexible and therefore capable of fitting a wide range of distribution shapes. The PDF of the Johnson SB distribution is defined by the following equation [67]:

$$f(x; a, b) = \frac{b}{x(1-x)} \cdot \phi \cdot (a + b \cdot log\frac{x}{1-x}) \tag{4.1}$$

where x is a value between 0 and 1 that represents the quantile (i.e. one of the cut points dividing the observations of the non-negative independent variable (e.g.$\tau_{mth,k}$) into continuous intervals of the same size). The a and b (a non-negative value) are the two shape parameters, and ϕ is the normal PDF. The two other parameters, loc and $scale$ are used to shift and scale the distribution, respectively. In order to use loc and $scale$, the standard form of the Johnson SB PDF is transformed by replacing x with the following variable m:

$$m = \frac{(x - loc)}{scale} \tag{4.2}$$

In this section, the seasonal variations for climate zone 6 will be presented whereas the results for the remaining climate zones may be found in section C.2; two of the most populated cities in Canada, Montreal and Toronto, are located in climate zone 6. Figures 4.11 through 4.13 display the seasonal variations in climate zone 6 for the $\tau_{mth,k}$ values. The mean ($\tau_{ssn,cl}$) and median value associated to each best fit distribution are provided in Table 4.1. Based on the fitted distribution of summer TTCs, the 95% confidence interval for the mean ($\tau_{sum,6}$) is 10.43± 0.58 hours - which is the lowest of all 4 seasons. The mean of winter TTCs ($\tau_{win,6}$) is 40.17± 2.89 hours and nearly 4 times the value of $\tau_{sum,6}$. Finally, the seasonal mean values calculated for fall and spring ($\tau_{fal,6}$ and $\tau_{spr,6}$) are 25.54± 2.90 hours and 30.42± 3.36 hours, respectively — approximately 2.5-3 times the value of $\tau_{sum,6}$.

The fitted distributions for fall, spring and summer are significantly more skewed than that of winter. In the summer, 50% of the values are below 10.43 hours, but the distribution spreads widely to higher TTCs. In the fall and spring, the values shift to the right with higher medians of about 16.11 and 24.01 hours, respectively. There is a dramatic difference is observed between

the winter distribution and the other seasonal distributions because it is significantly less skewed and bears a higher resemblance to a normal distribution. From summer to winter, a similar shift occurs across all climate zones towards higher TTC values.

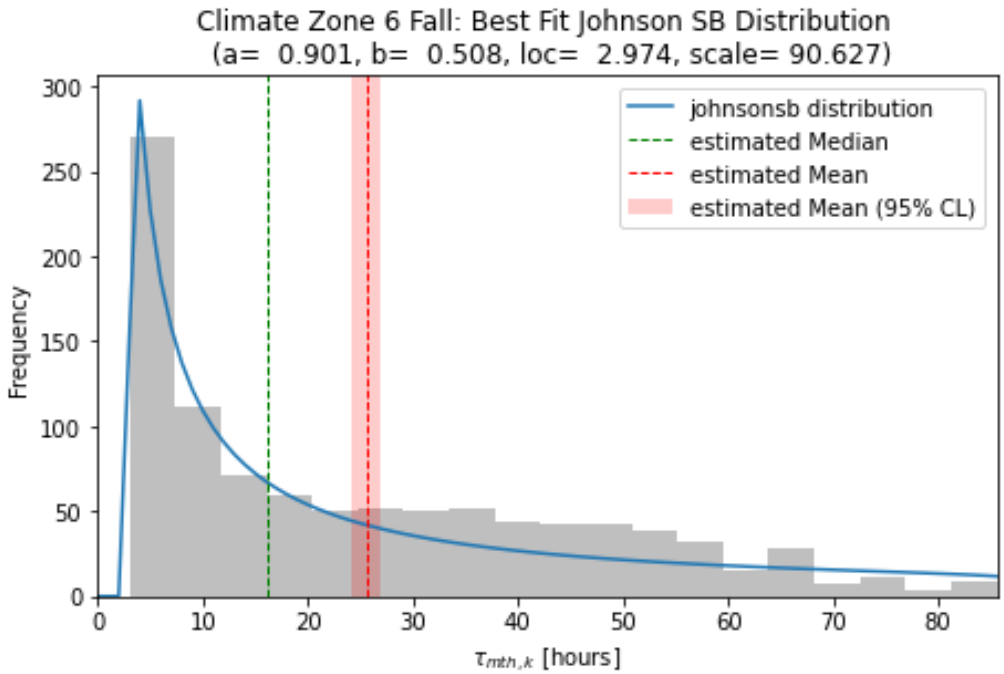


Figure 4.11: For climate zone 6 dwellings in fall, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

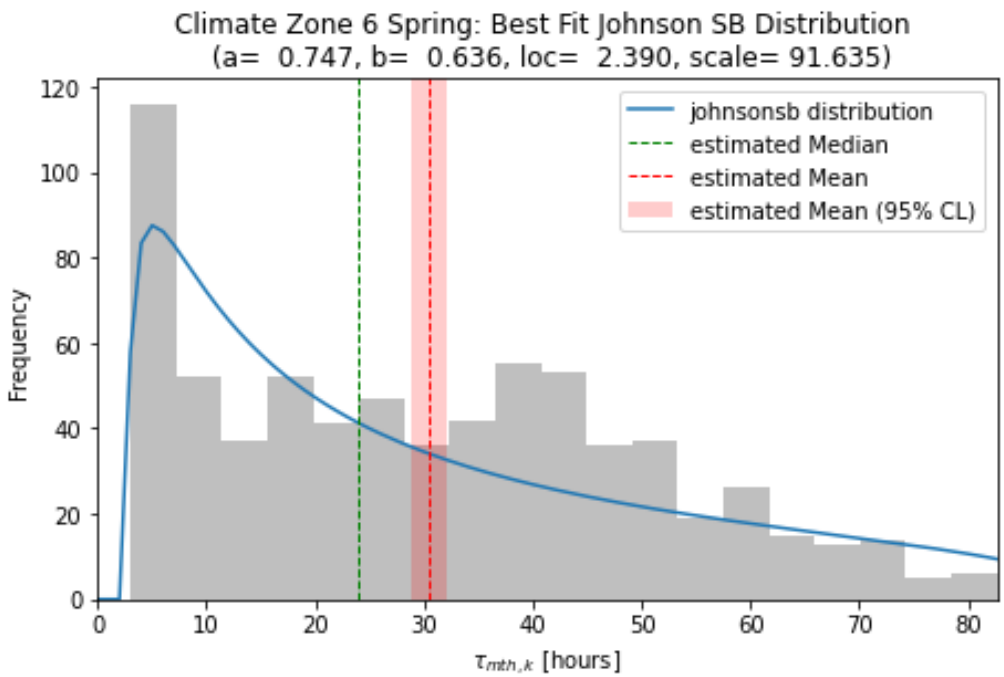


Figure 4.12: For climate zone 6 dwellings in spring, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

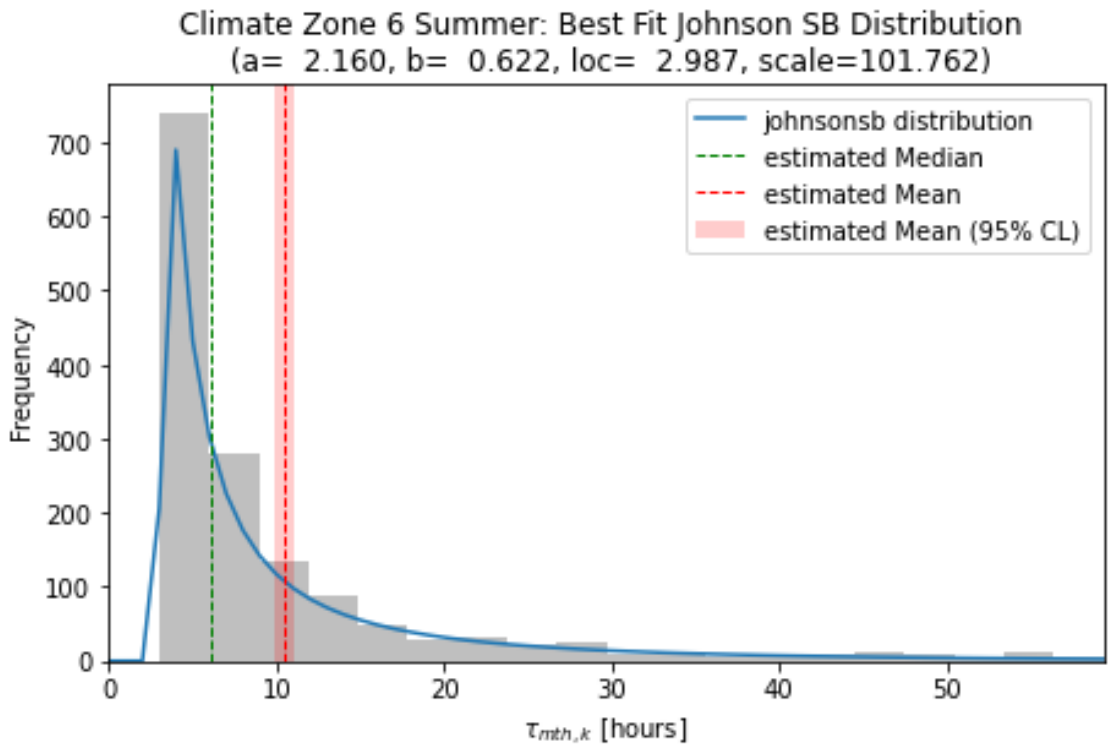


Figure 4.13: For climate zone 6 dwellings in summer, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

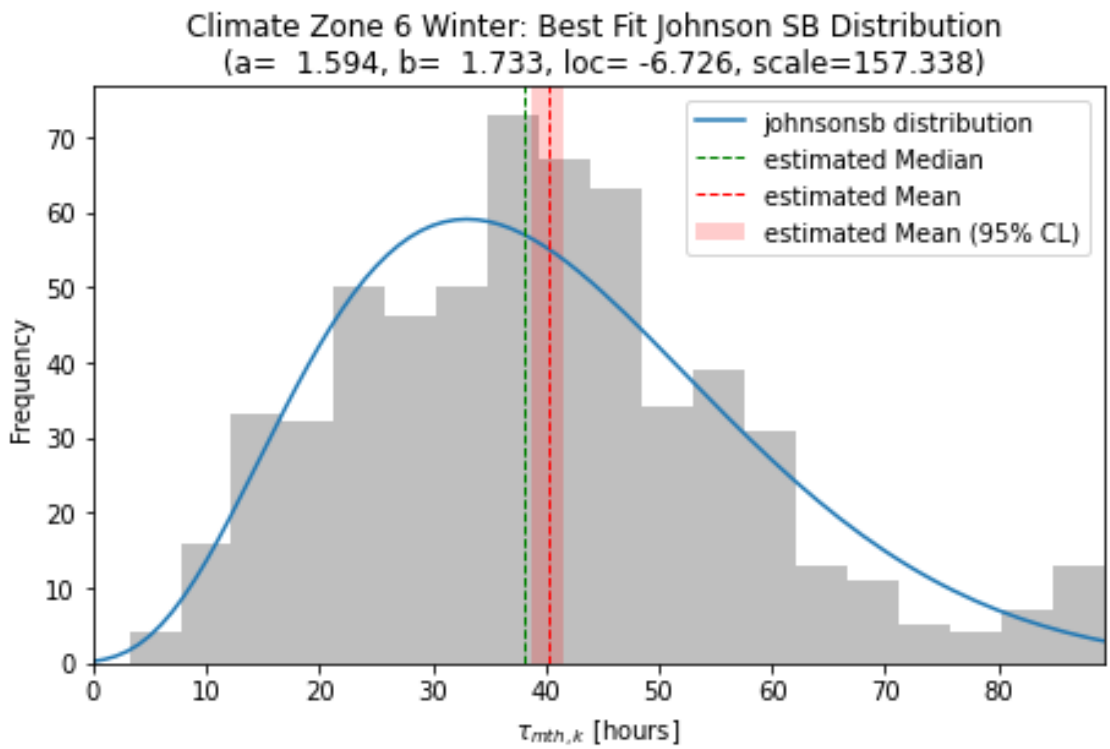


Figure 4.14: For climate zone 6 dwellings in winter, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

Table 4.1: Climate zone 6: Statistics for the best fit climate-season distributions.

					95% confidence interval for the mean		
Zone	Season	Mean	Median	Std	Lower Limit	Upper Limit	Length
6	Fall	25.54	16.11	23.42	24.09	26.99	2.90
6	Winter	40.17	38.11	18.01	38.72	41.62	2.90
6	Spring	30.42	24.01	22.77	28.75	32.10	3.35
6	Summer	10.43	6.04	11.38	9.85	11.01	1.16

4.2.1 Indicators of Thermal Performance and Resilience

The TTC can be used as an indication of both thermal performance and thermal resilience for a dwelling. The interval data collected from thousands of residential smart thermostats helped to develop a simplified and scalable method for estimating the TTC of a dwelling. In turn, the estimated TTCs can also help to characterize the thermal response of a free-running Canadian or American dwelling, providing insight into the passive thermal performance provided its thermal insulation, its airtightness, and its thermal mass. As seen in Figure 4.15, the TTC can also inform decision-making regarding Demand Side Management (DSM) strategies (e.g. building retrofits and Demand Response (DR)) related to the existing housing stock. Being aware of a dwelling's TTC offers building shareholders an idea of how long a free-running dwelling can passively maintain a habitable indoor temperature for occupants or how long before the dwelling would be vulnerable to a number of temperature-related durability issues.

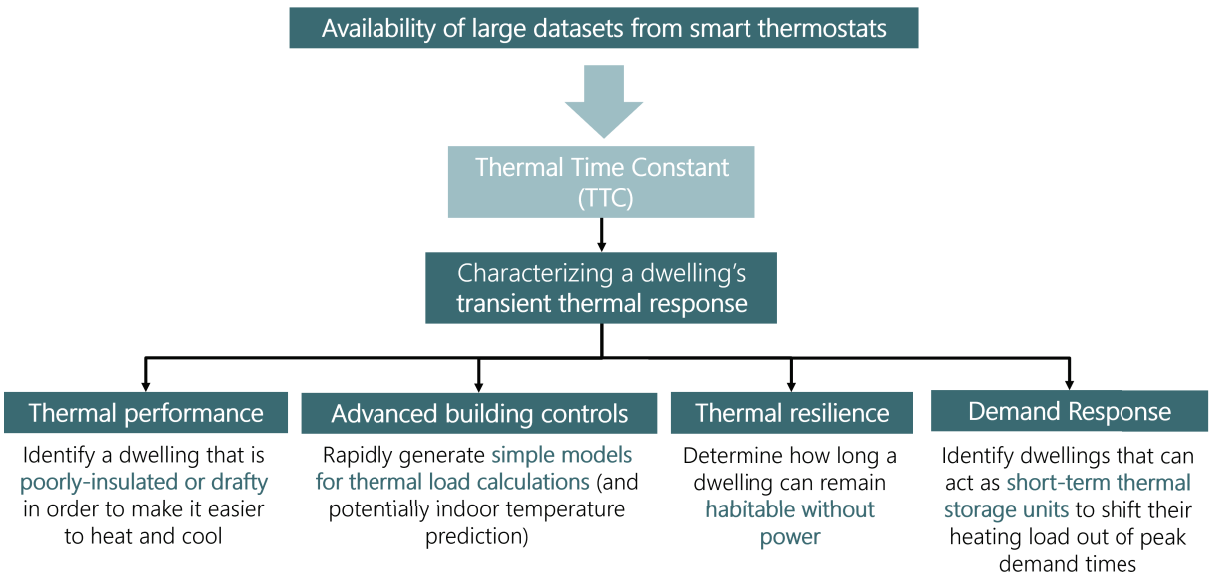


Figure 4.15: The TTC can offer valuable insights regarding thermal performance and resilience.

To the benefit of utilities and homeowners alike, the TTC characterizes a building's capacity to store and transfer heat. By providing short-term sensible heat storage, a building's thermal inertia could help optimize building energy management [68]. The TTC also offers valuable insights into the thermal flexibility provided by the dwelling's construction, and where electrical heating is used, electrical flexibility. These insights can contribute to two important and intervention approaches: DSM and Model-based Predictive Control (MPC).

Applications Towards Grid Operation

The first approach, DSM, consists of optimal operating strategies designed to efficiently manage on-site energy consumption with the aim of reducing energy costs, improving grid reliability and maintaining occupant comfort [69]. These strategies include matching demand load profiles to electricity generation profiles, reducing or interrupting consumption, and shifting consumption away to different time periods.

For example, in both Canada and the US, the cost of delivering power often depends on the time of day it is used. Utilities have a responsibility to meet the varying energy demands of their consumers, at any instant in time. Sharp rises in energy demand at peak times make electricity more costly to deliver; utilities must either operate additional and often less efficient power plants or purchase the missing electricity from the power market [15]. When power is not readily available, load shedding is required which often results in power outages.

Smart grids will create opportunities for new solutions to energy challenges not only on the supply side but also on the demand side. Being able to identify a typical TTC value from operational data can provide utilities and home automation companies with a better understanding of a home's thermal resilience. This understanding can inform decisions about their selection and use of DSM strategies for existing dwellings. Moreover, power outages due to emergency load reduction, extreme weather events and physical failures in the energy infrastructure have left large numbers of homes without power in the past. For example, most recently, a severe winter storm left millions of Texans without power, heat and running water for several days [70]. During power outages, access to TTC values also inform utilities on how long a typical North American house could go without power, under different outdoor temperature conditions, before compromising thermal comfort and safety levels.

Applications Towards Building Operation

The second approach, MPC, uses a building energy model to select the HVAC control inputs that result in the best predicted energy performance; the optimal performance is determined

by minimizing the energy costs and maintaining occupant comfort [17]. The emergence of large volumes of residential data and the modernization of the energy infrastructure provide an opportunity to consider the use of more advanced building control strategies in the residential sector.

North America represents a number of different climate zones and a widely diverse housing stock. Traditional detailed investigations of each home's thermal performance would be impractical due to cost and realistic time expectations. With IoT capabilities being more affordable and available, optimizing building operation through the use of building control technologies, like smart thermostats, has become a convenient and non-invasive solution for homeowners to reduce their energy costs. Using the smart thermostat data available, simple low-order Resistance-Capacitance (RC) thermal networks could be rapidly generated to represent the thermal performance of a dwelling and improve thermostat performance. With smart thermostats gaining in popularity, their presence in the home and the data they provide creates, in turn, an opportunity to rapidly calculate thermal loads, and potentially introduce a more advanced building control strategies to the residential sector.

For more than two decades, grey-box modelling using low-order RC-models has been used successfully in advanced building control strategies such as Model-based Predictive Control (MPC) in commercial buildings [17], [71]. Due to time and cost constraints, the application of MPC in the residential sector has previously been limited. With a more advanced building control strategy in the residential sector, the run-time of HVAC homes could be optimized to reduce energy costs without comprising comfort. In the future, the introduction of MPC could help flatten the energy use of a neighborhood during peak periods by coordinating the run-times of several HVAC systems. Derived from smart thermostat data and a simplified grey-box model, a TTC estimated in this study could be paired with knowledge of the total heat gains to test its accuracy when representing the thermal dynamics of North American dwellings for use in MPC; testing the TTC in this manner is outside the scope of this study.

Estimating the Thermal Delay

As a thermal performance and resilience indicator, one disadvantage of the TTC is that it does not readily communicate to building stakeholders the time until a dwelling becomes uncomfortable to live in. The TTC is a mathematically convenient measure for expressing the dwelling's thermal inertia but its physical meaning can be difficult to grasp for homeowners despite being a time-based metric. Figure 3.10 offers a visual example of how the TTC relates to the indoor temperature response curve of a dwelling when exposed to both a colder and a warmer surrounding environment. A more accessible measure is the thermal delay (t_{d}); this measure can be defined as the estimated time required for the indoor temperature to fall (or

rise) by a specific temperature difference (ΔT_{in}) after the dwelling enters into free-running conditions [36]. The thermal delay can offer for insight regarding thermal comfort or durability issues after a power outage or the failure of an HVAC system (e.g. how long the dwelling will remain habitable or how long before the pipes may freeze).

Due to time constraints and the lack of information available regarding the solar and internal gains, this study limits the exploration of the thermal delay to cold-weather conditions. In warm-weather conditions, the fluctuations of the solar or internal heat gains throughout the day are very important factors in addition to the TTC for determining how quickly a dwelling may become uncomfortable. The focus will therefore be on the link between the TTC and cold-weather thermal resilience (i.e. how long a free-running dwellings in cool-down mode can remain habitable).

The thermal delay can be useful in providing building stakeholders with an idea of how quickly the indoor temperature of a dwelling can drop, in cold-weather conditions, to any temperature of interest (e.g. the acceptable minimum for occupant comfort). The difference between the indoor temperature at the beginning of the free-running period ($T_{in,o}$) and the selected temperature of interest ($T_{\text{in}}(t_{\text{d}})$) can be expressed as follows:

$$\Delta T_{\text{in}} = T_{\text{in}}(t_{\text{d}}) - T_{\text{in,o}} \tag{4.3}$$

Using Equation 3.7 where t= t_{d}, $T_{\text{in}}(t_{\text{d}})$ may be defined as:

$$T_{\text{in}}(t_{\text{d}}) = T_{\text{ext}} + (T_{in,o} - T_{\text{ext}})e^{-t_{\text{d}}/\tau} \tag{4.4}$$

The substitution of $T_{\text{in}}(t_{\text{d}})$ from Equation 4.4 into Equation 4.3 therefore yields:

$$\Delta T_{\text{in}} = T_{\text{ext}} + (T_{\text{in,o}} - T_{\text{ext}})e^{-t_{\text{d}}/\tau} - T_{in,o} \tag{4.5}$$

and t_{d} can therefore be represented as a function of ΔT_{in}, the initial indoor-outdoor temperature difference ΔT_{o} and τ:

$$\begin{aligned} t_{\text{d}} &= f(\Delta T_{\text{in}}, \Delta T_{\text{o}}, \tau) \\ &= -\tau \cdot ln(1 - \frac{\Delta T_{\text{in}}}{\Delta T_{\text{o}}}) \end{aligned} \tag{4.6}$$

Figures 4.16 and 4.17 illustrate the times expected for typical dwellings to reach its minimum acceptable temperature, assuming an initial indoor temperature ($T_{in,o}$) of 21°C and a T_{ext}

of -5°C. In each figure, the equation 3.7 is used to plot the indoor temperature response of the typical dwellings considered. The comfort ranges applied in the figures are based those used by Henao et al. (in accordance with the Canadian Center for Occupational Health and Safety, and the ASHRAE Standard 55) [72].

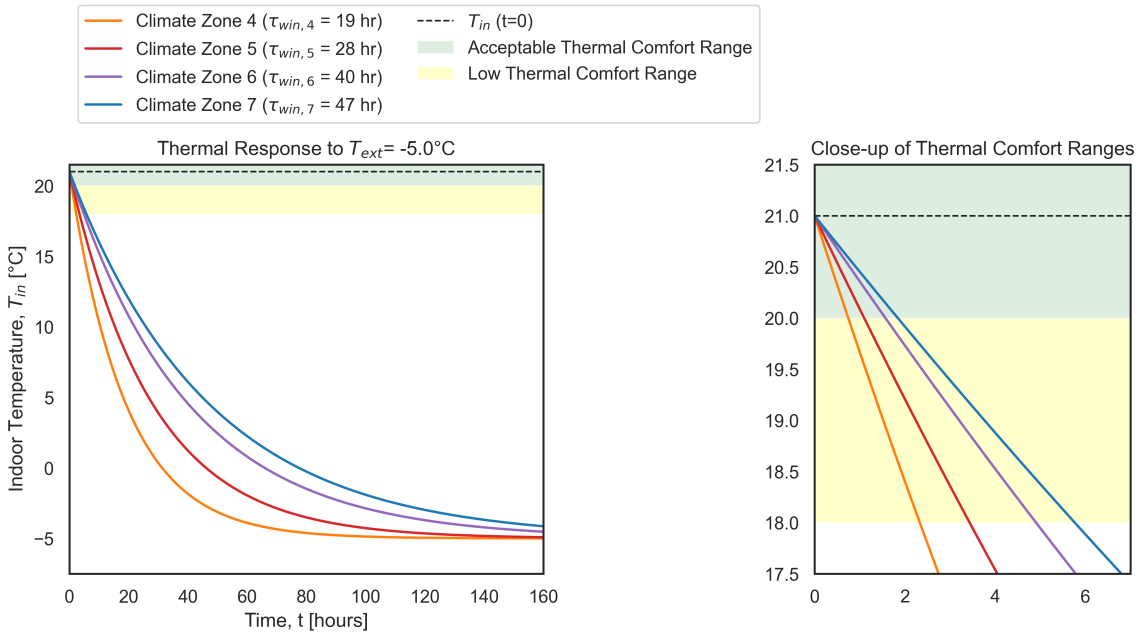


Figure 4.16: T_{in} vs. t where T_{ext}=-5°C and different time constants (i.e. typical dwellings in climate zones 4 through 7).

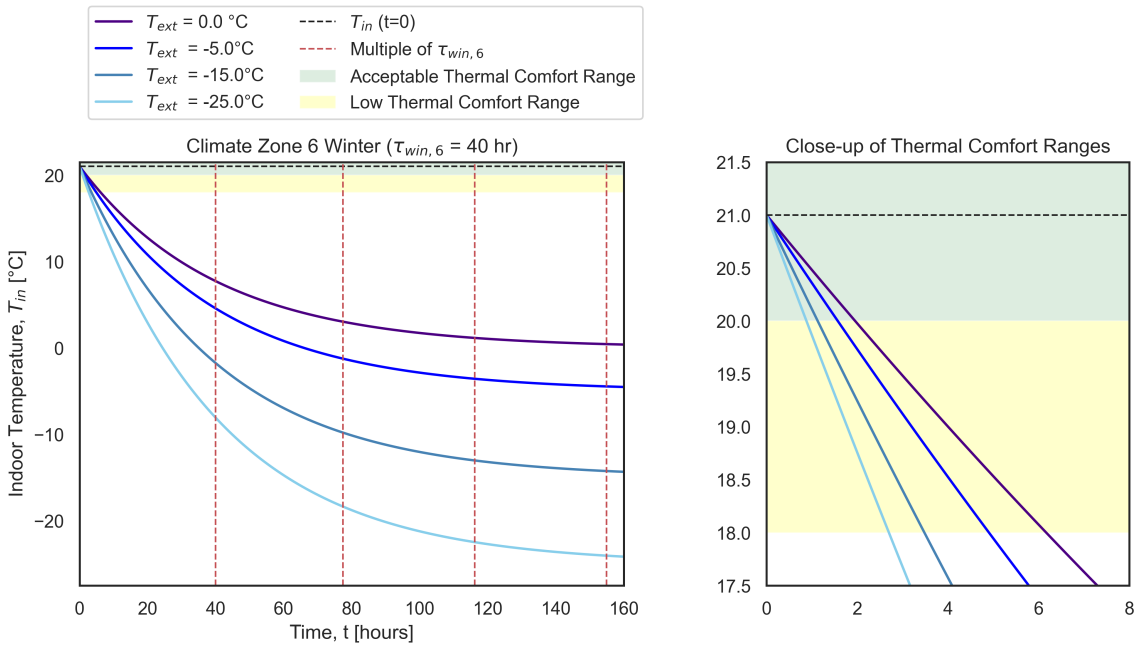


Figure 4.17: T_{in} vs. t for a typical climate zone 6 dwelling (τ = 40 hr.) where T_{in}= 21°C, and different T_{ext} varying from 0 to -25°C.

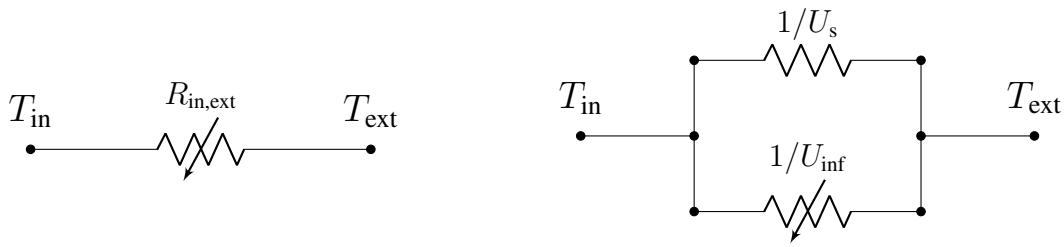
In Figure 4.16, the cooling response of free-running dwellings in different heating-dominated zones is displayed using $\tau_{win,4}$, $\tau_{win,5}$, $\tau_{win,6}$, and $\tau_{win,7}$, assuming an outdoor temperature (T_{ext}) of -5°C. Comparing the thermal response of different climate zones, the curve estimating $T_{\text{in}}(t)$ becomes steeper as the the τ decreases in value and the time until the dwelling becomes uncomfortable is shortened.

In Figure 4.17, the focus shifts to the cooling response of a typical climate zone 6 dwelling under different outdoor temperature conditions between 0 and -25°C, assuming a $\tau_{win,6}$ of 40 hours and a $T_{in,o}$ of 21°C. As the indoor-outdoor temperature difference increases in value, the curve estimating $T_{\text{in}}(t)$ becomes steeper.

Effect of Natural Ventilation and Air Leakage

Acceptable thermal comfort and indoor air quality is achieved in part by considering air exchange between the outdoor environment and indoor environment; air exchange can be classified as either ventilation if intentional or air leakage (e.g infiltration and exfiltration) if unintentional [73].

For a free-running dwelling, the combined effect of natural ventilation, infiltration, and exfiltration have a significant impact on monthly and seasonal τ values, and therefore its indoor temperature response. These three types of air exchanges can be hard to predict because they depend on a number of factors including weather conditions, building construction, occupants and maintenance [73]. Figure 4.18 shows how the $R_{\text{in,ext}}$ of a free-running dwelling is variable and can broken down into two parallel resistors. One resistor represents the heat transmission through the physical components of the building envelope (i.e. the walls, windows, doors, floors and ceilings), and the other, a variable resistor, represents the heat transfer through its openings (i.e. doors, windows, cracks and gaps); the thermal resistances for each resistor is equal to the reciprocal of its corresponding thermal conductance (U_{s} and U_{inf}, respectively).



(a) Overall thermal resistance (b) Equivalent pair of parallel thermal resistances

Figure 4.18: Overall thermal resistance broken down to represent heat transfer across the building envelope's physical components and through its openings.

The $R_{\text{in,ext}}$ of a free-running dwelling can therefore be expressed in the following form by adding the parallel resistors:

$$\begin{aligned} R_{\text{in,ext}} &= \frac{1}{U_{\text{s}} + U_{\text{inf}}} \\ &= \frac{1}{U_{\text{s}} + (c_{\text{p,air}} \cdot \rho_{\text{air}} \cdot ACH \cdot \frac{1hr}{3600s} \cdot V_{\text{air}})} \end{aligned} \tag{4.7}$$

where $c_{\text{p,air}}$ is the specific heat of air, ρ_{air} is the density of air, ACH is the number of air changes per hour (i.e. how many times the total volume of interior air is replaced per hour), and V_{air} is total air volume in the dwelling.

A building with a high U_{inf} value from occupants opening their windows or poor construction can result in a reduced $R_{\text{in,ext}}$ value; this can be demonstrated using Equation 4.7 in the following example. Consider a dwelling located in Quebec, Canada for which Date et al. developed a 1R1C model and identified its model parameters using training data from winter; the dwelling has a TTC of 31 hours, a $R_{\text{in,ext}}$ of 0.0064 K/W, and a C_{in} of 17.2 MJ/K [74]. For the dwelling in question, the U_{inf} and ACH are estimated in Table 4.2 for the case of both winter and summer; the calculations are carried out assuming that there is a negligible difference for C_{in} between the two cases, and that the dwelling has a summer TTC value of 6.2 hours, a total surface area (A) of 350 m^2, a V_{air} of 500 m^3 and an enclosure R-value (R') of 3.5 m^2K/W.

Table 4.2: Example of the seasonal effects related to natural ventilation and air leakage

		Winter (i.e. when windows are mainly closed)	**Summer** (i.e. when windows are opened for longer periods)
$R_{\text{in,ext}}$	[K · W^{-1}]	**0.0064**	**0.0013**
C_{in}	[MJ · K^{-1}]	17.2	17.2
$\tau = R_{\text{in,ext}}C_{\text{in}}$	[hr]	**31**	**6.2**
R'	[m^2 · K · W^{-1}]	3.5	3.5
A	[m^2]	350	350
$R_{\text{s}} = R'/A$	[K · W^{-1}]	**0.010**	**0.010**
$U_{\text{s}} = 1/R_{\text{s}}$	[W · K^{-1}]	98	98
$U_{\text{inf}} = (1/R_{\text{in,ext}}) - U_{\text{s}}$	[W · K^{-1}]	**62**	**702**
V_{air}	[m^3]	500	500
$c_{\text{p,air}}$	[J · kg^{-1} · K^{-1}]	1.2	1.2
ρ_{air}	[kg · m^{-3}]	1000	1000
ACH		**0.37**	**4.2**

As a result, the example shows the maximum $R_{\text{in,ext}}$ of 0.0064 K/W would occur in winter

(i.e. when the windows are mainly closed) when the U_{inf} and ACH are low. In contrast, in the summer, the windows are more likely to be open during free-running conditions; the U_{inf} and ACH are therefore higher in value, and the $R_{\text{in,ext}}$ is reduced to 0.0013 K/W. The relationship displayed between the estimated TTC, U_{inf} and ACH could play an important role in identifying a building's potential for summer peak load reductions; pre-cooling a building at night through natural ventilation can help shift thermal loads related to cooling [75], [76].

Chapter 5

Conclusion

As a result of the increased popularity of Internet of Things (IoT) devices, the research in the field of building sciences is adapting to take advantage of the growing availability of building performance datasets. When dealing with large datasets, Data Mining (DM) is an effective, scalable and flexible method for discovering valuable patterns and trends, pertaining to many aspects of a building's operation including its thermal behaviour. Based on increasingly available smart thermostat data, this thesis proposes a framework for using DM and a Reduced-Order Model (ROM) to estimate the effective Thermal Time Constant (TTC) of a dwelling [77]. The TTC is an indicator of the passive thermal performance of a building; with no influence from a heating or cooling system, it provides an indication of the time required for a building space to cool down or heat up in response to a thermal change in its surroundings. The TTC represents the combined influence of effective thermal insulation, airtightness and thermal capacitance provided by a building's enclosure and internal mass on its thermal response.

The proposed methodology permits the thermal performance of a building to be described and compared in terms of the TTC; a thermal performance indicator which is not as popular or as readily available as the R-value. At present, the R-value is often the only information provided to reflect the thermal performance of a building however relying on the R-value alone fails to encompass the effects of thermal bridging, building enclosure defects, thermal mass, and air leakage. Estimated using real building performance data, the TTC can provide insight into the effective thermal capacitance of the dwelling, however it is, unfortunately, a building property whose value is not commonly known and difficult to determine when limited in time, cost and information. Where previous studies related to the TTC may have been restricted by a lack of available field data due to practical constraints, the methodology described in this thesis uses smart thermostat data to provide efficiency in time and effort; this method can estimate the TTC of thousands of dwellings across North America at once and is not limited by study duration. Building professionals and stakeholders could therefore easily use the TTC, in addition to the

widely accepted R-value, to provide a better indication of the thermal performance and thermal resiliency of a building.

The estimation of the TTC is achieved by using building operational data to calibrate simple Resistance-Capacitance (RC) models. Collected over a one year period from September 2017 to August 2018, the smart thermostat data used in this study includes measured time series data recorded at 5-minute intervals (i.e. the indoor temperature readings, the outdoor temperature readings and the HVAC equipment runtimes) and background information for over 15,000 Canadian and US dwellings. The indoor temperature response of the dwellings was tracked only during periods when (a) solar and internal gains could be assumed negligible (i.e. between sunset and sunrise), (b) the house is under free-running conditions (i.e. when the Heating, Ventilation and Air-Conditioning (HVAC) system is switched off) for more than one hour, and (c) the outdoor temperature remains relatively constant (i.e. the change in outdoor temperature is smaller than or equal to 2°C). Next, the individual TTCs are estimated by fitting a temperature response curve of a first-order model to building-specific measured data. Finally, a statistical analysis is applied to identify a typical range of effective TTC values in the North American housing stock, according to climate zone (i.e location) and season; this additional step provides provides building stakeholders with a reference for comparing the natural thermal response between buildings and the foundations for developing standards in relation to TTC values.

The results of the study suggest that TTC values are affected by location-based construction practices and seasonality. As expected, an increasing trend was observed in the TTC values from warmer to colder climate zones. The ASHRAE climate zones with sufficient data to yield accurate results are zones 1 through 7. The dwellings in colder climates have longer TTCs which reflect the higher thermal resistances and heavier constructions than those found in warmer climates whose TTCs are shorter. There are also significant differences between estimated values for the summer and winter months across all climate zones, which may be attributed to the interaction of occupants with the building envelope. In winter, the mean TTC according to climate zone ranges from about 7 to 47 hours. In contrast, the summer values vary over a smaller and lower range of 6 to 19 hours presumably due to occupants opening windows in more favourable weather. The combined effects of climate and seasonality are significant and show that seasonality should be considered in the future considerations of the TTC as a building parameter.

The estimated TTC can be applied in the following three ways to support a future towards more resilient and sustainable buildings as a:

1. **thermal performance indicator** permitting the rapid identification of homes in need of building envelope retrofits;

2. **parameter in simple RC models**, providing a less costly and less timely way of introducing of a large -scale application of model-based energy load estimation and management to the existing housing stock;

3. **thermal resilience indicator** providing insight into the suitability of certain Demand Side Management (DSM) strategies.

Moreover, future studies may expand upon the presented methodology by focusing on:

- The incorporation of solar gains and internal gains into the simplified model;
- Further investigation into the TTC's dependency on other building-specific variables (e.g. age, floor area);
- The determination of when and how much windows are opened by analyzing the measured indoor temperature data and measured relative humidity data;
- The identification of threshold TTC values associated with poor and high performance buildings;
- The applied use of the TTC in building control strategies and the prediction of indoor temperature;
- The applied use of the TTC in energy management strategies and shifting energy demand profiles.

In brief, this thesis presents a novel approach that uses smart thermostat data and a ROM to estimate the TTC - a lesser known dynamic building characteristic. The method and Python code developed for estimating the TTC is more efficient in time and cost than the typical detailed building simulation and applicable particularly when little building information is available. The ease and flexibility provided by the data mining approach makes the estimation of the TTC possible for thousands of real North American residential buildings. The analysis of the TTC values shed light on important seasonal and geographic differences regarding a building's thermal behaviour under transient heat flow conditions; these insights are helpful in the assessing thermal comfort, energy efficiency and thermal resilience. The scalable methodology can be expanded to millions of buildings in the North American housing stock, and exploited towards the larger objective of more resilient and sustainable buildings and communities.

References

[1] J. A. Leech, W. C. Nelson, R. T. Burnett, S. Aaron, and M. E. Raizenne, "It's about time: A comparison of Canadian and American time-activity patterns," *Journal of Exposure Analysis and Environmental Epidemiology*, vol. 12, no. 6, pp. 427–432, 2002.

[2] Office of Energy Efficiency, "Energy Use Data Handbook Tables – 1990 to 2017," National Resources Canada, Tech. Rep., 2019. [Online]. Available: https://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/menus/trends/handbook/tables.cfm.

[3] U.S. Energy Information Administration, "April 2019 Monthly Energy Review," U.S. Department of Energy, Tech. Rep., 2019. [Online]. Available: https://www.eia.gov/totalenergy/data/monthly/.

[4] U.S. Environmental Protection Agency, *Greenhouse Gas Emissions : Sources of Greenhouse Gas Emissions*, 2020. [Online]. Available: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions (visited on 07/01/2020).

[5] U.S. Energy Information Administration, "2015 Residential Energy Consumption Survey," U.S. Department of Energy, Tech. Rep., 2018. [Online]. Available: https://www.eia.gov/consumption/residential/data/2015/.

[6] Canada Energy Regulator, "Canada's Energy Future 2019 Energy Supply and Demand Projections to 2040," Canada Energy Regulator, Tech. Rep., 2019. [Online]. Available: https://www.cer-rec.gc.ca/en/data-analysis/canada-energy-future/2019/index.html.

[7] U.S. Energy Information Administration, "Annual Energy Outlook 2019 with projections to 2050," U.S. Department of Energy, Tech. Rep., 2019. [Online]. Available: https://www.eia.gov/outlooks/aeo/pdf/0383(2017).pdf.

[8] J. Manyika, M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon, "The Internet of Things: Mapping the value beyond the hype," Tech. Rep. June, 2015, p. 144. [Online]. Available: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world.

[9] Statistics Canada, *Table 11-10-0222-01: Household spending, Canada, regions and provinces*, 2019. [Online]. Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110022201.

[10] Canada Energy Regulator, *Market Snapshot: Fuel poverty across Canada – lower energy efficiency in lower income households*, 2017. [Online]. Available: https://www.cer-rec.gc.ca/nrg/ntgrtd/mrkt/snpsht/2017/08-05flpvrt-eng.html?=undefined%7B%5C&%7Dwbdisable=true (visited on 07/01/2018).

[11] D. J. Bednar and T. G. Reames, "Recognition of and response to energy poverty in the United States," *Nature Energy*, vol. 5, no. June, pp. 432–439, 2020.

[12] S. Jessel, S. Sawyer, and D. Hernández, "Energy, Poverty, and Health in Climate Change: A Comprehensive Review of an Emerging Literature," *Frontiers in Public Health*, vol. 7, no. December, 2019.

[13] A. T. D. Perera, V. M. Nik, D. Chen, J.-L. Scartezzini, and T. Hong, "Quantifying the impacts of climate change and extreme climate events on energy systems," *Nature Energy*, vol. 5, no. 2, pp. 150–159, 2020.

[14] W. Maclay and Maclay Architects, *The New Net Zero: Leading-Edge Design and Construction of Homes and Buildings for a Renewable Energy Future*. White River Junction, VT: Chelsea Green Publishing, 2014.

[15] J. Zheng, D. W. Gao, and L. Lin, "Smart meters in smart grid: An overview," in *IEEE Green Technologies Conference*, Denver, CO, United states, 2013, pp. 57–64.

[16] J. Rotondo, R. Johnson, N. Gonzalez, A. Waranowski, C. Badger, N. Lange, E. Goldman, and R. Foster, "Overview of Existing and Future Residential Use Cases for Connected Thermostats," U.S. Department of Energy, Tech. Rep., 2016. [Online]. Available: https://www.energy.gov/sites/prod/files/2016/12/f34/Overview%20of%20Existing%20Future%20Residential%20Use%20Cases%20for%20CT%7B%5C_%7D2016-12-16.pdf.

[17] A. Gonzalez-Vidal, A. P. Ramallo-Gonzalez, F. Terroso-Saenz, and A. Skarmeta, "Data driven modeling for energy consumption prediction in smart buildings," in *2017 IEEE International Conference on Big Data (Big Data)*, Piscataway, NJ, USA, 2017, pp. 4562–9.

[18] G. Baasch, P. Westermann, and R. Evins, "Identifying whole-building heat loss coefficient from heterogeneous sensor data: An empirical survey of gray and black box approaches," *Energy and Buildings*, vol. 241, p. 110 889, 2021.

[19] C. Miller, Z. Nagy, and A. Schlueter, "A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings," *Renewable and Sustainable Energy Reviews*, pp. 1–13, 2017.

[20] A. Kavousian, R. Rajagopal, and M. Fischer, "Determinants of residential electricity consumption: using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior," *Energy (UK)*, vol. 55, pp. 184–94, 2013.

[21] S. Verbeke and A. Audenaert, "Thermal inertia in buildings: A review of impacts across climate and building use," *Renewable and Sustainable Energy Reviews*, vol. 82, no. August 2017, pp. 2300–2318, 2018.

[22] Building Science Corporation, "Thermal Metric Summary Report," Tech. Rep., 2013, p. 153. [Online]. Available: `https://www.buildingscience.com/documents/special/thermal-metric-documents/thermal-metric-summary-report`.

[23] A. J. Ghajar and Y. A. Cengel, *Heat and Mass Transfer: Fundamentals and Applications*. McGraw-Hill Education, 2014.

[24] J. Straube, *Thermal Control in Buildings*, 2006. [Online]. Available: `https://www.buildingscience.com/documents/digests/bsd-011-thermal-control-in-buildings` (visited on 04/01/2018).

[25] T. Kesik, L. O'Brien, A. Ozkan, and A. Chong, "Thermal Resilience Design Guide," University of Toronto, Tech. Rep., 2019. [Online]. Available: `https://pbs.daniels.utoronto.ca/faculty/kesik%7B%5C_%7Dt/PBS/Kesik-Resources/Thermal-Resilience-Guide-v1.0-May2019.pdf`.

[26] P. Gass, D. Echeverría, and A. Asadollahi, "Cities and Smart Grids in Canada," International Institute for Sustainable Development, Tech. Rep. September, 2017, p. 36. [Online]. Available: `https://www.iisd.org/system/files/publications/cities-smart-grids-canada.pdf`.

[27] W. A. Beckman, L. Broman, A. Fiksel, S. A. Klein, E. Lindberg, M. Schuler, and J. Thornton, "TRNSYS. Most complete solar energy system modeling and simulation software," *Renewable energy*, vol. 5, no. 1 -4 pt 1, pp. 486–488, 1994.

[28] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte, and J. Glazer, "EnergyPlus: Creating a new-generation building energy simulation program," *Energy and Buildings*, vol. 33, no. 4, pp. 319–331, 2001.

[29] James J. Hirsh and Associates, *The Home of DOE-2 based Building Energy Use and Cost Analysis Software*. [Online]. Available: `https://doe2.com/` (visited on 06/01/2019).

[30] D. Coakley, P. Raftery, and M. Keane, *A review of methods to match building energy simulation models to measured data*, 2014.

[31] M. R. Braun, H. Altan, and S. B. Beck, "Using regression analysis to predict the future energy consumption of a supermarket in the UK," *Applied Energy*, vol. 130, pp. 305–313, 2014.

[32] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 102–109, 2014.

[33] R. Kramer, J. van Schijndel, and H. Schellen, "Simplified thermal and hygric building models: A literature review," *Frontiers of Architectural Research*, pp. 318–325, 2012.

[34] A. Athienitis and W. O'Brien, *Modeling, design, and optimization of net-zero energy buildings*. 2015.

[35] G. Burnand, "The study of the thermal behaviour of structures by electrical analogy," *British Journal of Applied Physics*, vol. 3, no. 2, pp. 50–53, 1952.

[36] K. A. Antonopoulos and E. P. Koronaki, "Effect of indoor mass on the time constant and thermal delay of buildings," *International Journal of Energy Research*, vol. 24, no. 5, pp. 391–402, 2000.

[37] G. Reynders, J. Diriken, and D. Saelens, "Quality of grey-box models and identified parameters as function of the accuracy of input and observation signals," *Energy Build. (Netherlands)*, vol. 82, pp. 263–74, 2014.

[38] H. Harb, N. Boyanov, L. Hernandez, R. Streblow, and D. Müller, "Development and validation of grey-box models for forecasting the thermal response of occupied buildings," *Energy and Buildings*, vol. 117, pp. 199–207, 2016.

[39] G. E. Box, "Science and statistics," *Journal of the American Statistical Association*, vol. 71, no. 356, pp. 791–799, 1976.

[40] T. Hong and S. H. Lee, "Integrating physics-based models with sensor data: An inverse modeling approach," *Building and Environment*, vol. 154, no. March, pp. 23–31, 2019.

[41] K. A. Antonopoulos and C. Tzivanidis, "Time constant of Greek buildings," *Energy (Oxford)*, vol. 20, no. 8, p. 785, 1995.

[42] ——, "Finite-difference prediction of transient indoor temperature and related correlation based on the building time constant," *International Journal of Energy Research*, vol. 20, no. 6, pp. 507–520, 1996.

[43] T. Catalina, J. Virgone, and E. Blanco, "Development and validation of regression models to predict monthly heating demand for residential buildings," *Energy and Buildings*, vol. 40, no. 10, pp. 1825–1832, 2008.

[44] J. Karlsson, L. Wadso, and M. Oberg, "A conceptual model that simulates the influence of high thermal inertia in building structures," in *fib Symposium 2012: Concrete Structures for Sustainable Community - Proceedings*, Stockholm, Sweden, 2012, pp. 631–634.

[45] I. S. Walker and A. K. Meier, "Residential Thermostats : Comfort Controls in California Homes," Lawrence Berkeley National Laboratory, Tech. Rep., 2008.

[46] Honeywell International Inc, *Non-Programmable Thermostats*. [Online]. Available: https://www.honeywellhome.com/us/en/products/air/thermostats/non-programmable-thermostats/manual-4-wire-premium-baseboard-thermostat-ct410b1017-e1/ (visited on 05/01/2021).

[47] ——, *Thermostat*. [Online]. Available: https://www.honeywellhome.com/us/en/products/air/thermostats/non-programmable-thermostats/the-round-mercury-free-thermostat-manual-changeover-single-stage-ct87n1001-u1/ (visited on 05/01/2021).

[48] 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.

[49] ecobee Inc, *Smart Thermostat with voice control and Smart Sensor*. [Online]. Available: https://www.ecobee.com/en-us/smart-thermostats/smart-wifi-thermostat-with-voice-control/ (visited on 05/01/2021).

[50] Sensibo, *Sensibo: Smart Air Sensing & Control*. [Online]. Available: https://sensibo.com/products/sensibo-air-bundle (visited on 05/01/2021).

[51] Honeywell International Inc, *T5+ Smart Home Thermostat*. [Online]. Available: https://www.honeywellhome.com/us/en/products/air/thermostats/wifi-thermostats/t5-smart-thermostat-with-c-wire-adapter-rcht8612wf2005-u/ (visited on 05/01/2021).

[52] Mysa Smart Thermostats, *Mysa for Electric Baseboard Heaters*. [Online]. Available: https://shop.getmysa.com/products/mysa-baseboard (visited on 05/01/2021).

[53] Google Inc, *Nest thermostats*. [Online]. Available: https://store.google.com/ca/magazine/compare%7B%5C_%7Dthermostats?hl=en-GB (visited on 05/01/2021).

[54] T. A. Reddy, *Applied Data Analysis and Modeling for Energy Engineers and Scientists*. 2011.

[55] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.

[56] Z. Yu, F. Haghighat, and B. C. Fung, "Advances and challenges in building engineering and data mining applications for energy-efficient communities," *Sustainable Cities and Society*, vol. 25, pp. 33–38, 2016.

[57] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "Step-by-step data mining guide," SPSS inc, Tech. Rep., 2000, pp. 1–78. [Online]. Available: `http://www.crisp-dm.org/CRISPWP-0800.pdf`.

[58] D. T. Larose and C. D. Larose, *Data mining and predictive analytics*, English, Hoboken, New Jersey, 2015.

[59] ecobee Inc, *Installation Guide ecobee 3*, 2014. [Online]. Available: `https://www.ecobee.com/wp-content/uploads/2013/12/ecobee3%7B%5C_%7DInstallationGu` `pdf`.

[60] NumFocus, *Pandas - Python Data Analysis Library*. [Online]. Available: `https://pandas.pydata.org/` (visited on 12/01/2020).

[61] American Society of Heating Refrigerating and Air-Conditioning Engineers, *ANSI/ASHRAE Standard 90.2-2018: Energy-Efficient Design of Low-Rise Residential Buildings*. Atlanta, GA, 2018.

[62] National Renewable Energy Laboratory, *Ashrae Climate Zones*, 2011. [Online]. Available: `https://openei.org/wiki/ASHRAE%7B%5C_%7DClimate%7B%5C_%7DZones` (visited on 10/01/2018).

[63] R. Talami, C. Karmann, F. Bauman, S. Schiavon, and P. Raftery, "Recent trends in radiant system technology in North America," Tech. Rep., 2017.

[64] J. Vivian, A. Zarrella, G. Emmi, and M. De Carli, "An evaluation of the suitability of lumped-capacitance models in calculating energy needs and thermal behaviour of buildings," *Energy and Buildings*, vol. 150, pp. 447–465, 2017.

[65] C. John, C. Vallianos, J. Candanedo, and A. Athienitis, "Estimating time constants for over 10,000 residential buildings in North America: towards a statistical characterization of thermal dynamics," in *Proceedings of the 7th International Building Physics Conference - Healthy, Intelligent and Resilient Buildings and Urban Environments*, 2018.

[66] The SciPy Community, *scipy.stats.kstest - Scipy v1.6.1 Reference Guide*. [Online]. Available: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html` (visited on 03/01/2020).

[67] ——, *scipy.stats.johnsonsb - SciPy v1.6.1 Reference Guide*. [Online]. Available: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.johnsonsb.html` (visited on 03/01/2021).

[68] M. J. Oliveira Panão, N. M. Mateus, and G. Carrilho da Graça, "Measured and modeled performance of internal mass as a thermal energy battery for energy flexible residential buildings," *Applied Energy*, vol. 239, no. October 2018, pp. 252–267, 2019.

[69] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381–388, 2011.

[70] E. Chung, *What caused the deadly power outages in Texas and how Canada's grid compares*, 2021. [Online]. Available: `https://www.cbc.ca/news/technology/power-outages-texas-canada-1.5920833` (visited on 02/01/2021).

[71] J. A. Candanedo, V. R. Dehkordi, and P. Lopez, "A multi-level architecture to facilitate MPC implementation in commercial buildings : basic principles and case study," no. May, 2013.

[72] N. Henao, M. Fournier, S. Kelouwani, and T. De, "Characterizing Smart Thermostats Operation in Residential Zoned Heating Systems and its Impact on Energy Saving Metrics," in *Proceedings of eSim 2018, the 10 conference of IBPSA-Canada*, 2018, pp. 17–25.

[73] American Society of Heating Refrigerating and Air-Conditioning Engineers, *2017 ASHRAE Handbook: Fundamentals*. Atlanta, GA, 2017.

[74] J. Date, J. A. Candanedo, and A. K. Athienitis, "Control-Oriented Modelling of Thermal Zones in a House : A Multi-Level Approach," in *International High Performance Buildings Conference*, 2016, pp. 1–10.

[75] Y. Chen, Z. Tong, W. Wu, H. Samuelson, A. Malkawi, and L. Norford, "Achieving natural ventilation potential in practice: Control schemes and levels of automation," *Applied Energy*, vol. 235, no. November 2018, pp. 1141–1152, 2019.

[76] A. Rabl and L. K. Norford, "Peak load reduction by preconditioning buildings at night," *International Journal of Energy Research*, vol. 15, no. 9, pp. 781–98, 1991.

[77] C. John, "Grey-box models: toward a systematic approach for design and operation of thermal zones," *Countribution to foundation paper by J. A. Candanedo, A. K. Athienitis et al., in preparation*,

[78] ecobee Inc, *How accurate is the temperature sensor in the ecobee?* [Online]. Available: `https://support.ecobee.com/hc/en-us/articles/227869067-How-accurate-is-the-temperature-sensor-in-the-ecobee-` (visited on 07/01/2018).

Appendix A

Additional Dataset Information

A.1 Data File Description

Table A.1: A list of the variables contained in each data file.

Field Name	Variable	Data Type
DateTime	Date and time (YYYY-MM-DD HH:MM) that the reading was taken	Metric
HvacMode	Indicates whether space conditioning system is set to heat, cool, auto or off	Categorical
Event	Anything that modifies the schedule (e.g. temperature hold, demand response, Vacation, SmartRecovery)	Categorical
Schedule	User-defined descriptors for desired set points against activity/behaviour (e.g. Vacation, Sleep, Away, Nap)	Categorical
T_ctrl	Average indoor temperature (in Fahrenheit) based on readings from relevant sensors (as defined by the schedule or mode the user); This temperature value used to guide the operation of the HVAC system.	Metric
T_stp_cool	Indoor cool setpoint (in Fahrenheit)	Metric
T_stp_heat	Indoor heat setpoint (in Fahrenheit)	Metric
Humidity	Indoor humidity (in RH%)	Metric
HumidityExpectedLow	Setpoint (for users who have a Humidifier) (in RH%)	Metric
HumidityExpectedHigh	Setpoint (for users who have a Humidifier) (in RH%)	Metric
auxHeat1,2,3	Runtime (seconds) for any heat source other than a heat pump (where 1,2,3 are the stages of the equipment)	Metric
compCool1,2,3	Runtime (seconds) for any cooling (where 1,2,3 are the stages of the equipment)	Metric
compHeat1,2,3	Runtime (seconds) for heat-pumps used in heating	Metric
fan	Runtime (seconds) for fan	Metric

Continued on next page

Table A.1 – continued from previous page

Field Name —	Variable	Data Type
Thermostat_Temperature	Indoor temperature measurement (in Fahrenheit) at the thermostat (i.e. not remote sensor); The ecobee temperature sensor has a +/- 1.0°C accuracy [78].	Metric
Thermostat_Motion	Detects motion at that date/time	Binary
Remote_Sensor_1,2,3_Temperature	Indoor temperature measurement (in Fahrenheit) at the remote sensor (where 1,2,3 denotes different sensors); The ecobee temperature sensor has a +/- 1.0°C accuracy [78].	Metric
Remote_Sensor_1,2,3_Motion	Detects motion at that date/time at the remote sensor (where 1,2,3 denotes different sensors)	Binary
T_out	Outdoor temperature, for the specific location of the dwelling, estimated using data from the closest weather stations	Metric
RH_out	Outdoor humidity (in RH%), for the specific location of the dwelling, estimated using data from the closest weather stations	Metric

A.2 Metadata File Description

Table A.2: A list of the dwelling characteristics contained in the metadata file.

Field Name	Dwelling Characteristic	Data Type
Identifier	Label assigned to thermostat (unique, alphanumeric, 40 characters long)	Categorical
Model	Label indicating the available thermostat features	Categorical
UserID	Label assigned to thermostat user (unique, alphanumeric, 40 characters long)	Categorical
Country	Country where thermostat installed	Categorical
ProvinceState	Province or state where thermostat installed	Categorical
City	City where thermostat installed	Categorical
Floor Area [ft2]	Floor area of dwelling	Metric
Style	Type of dwelling (e.g. detached, row-house, apartment, etc.)	Categorical
Number of Floors	Number of floors in the dwelling	Count
Age of Home [years]	Age of the dwelling	Count
Number of Occupants	Number of occupants in the dwelling	Count
installedCoolStages	Levels of cooling output available (up to 4 based on equipment)	Count
installedHeatStages	Levels of heating output available (up to 4 based on equipment)	Count
allowCompWithAux	Indicates whether the heat pump and auxiliary heat may run at the same time	Binary
Has Electric	Indicates if space conditioning equipment powered by electricity	Binary
Has a Heat Pump	Indicates if heat pump is installed	Binary
AuxilliaryHeatFuelType	Indicates fuel type used to power auxilliary heat	Categorical
Number of Remote Sensors	Number of remote sensors connected to the thermostat	Count
filename	A filename used for the monthly CSV data files consisting of the thermostat identifier	Categorical

A.3 Study Sample

A.3.1 Initial Data Analysis

As previously mentioned, for the purpose of the study, the 15,363 Canadian and U.S. dwellings will be analyzed over the course of one year in order to identify patterns and trends in their thermal response. Upon the completion of the data preparation, data visualizations are used to display the distribution of the dwellings in relation to the following dwelling characteristics: country, climate zone, dwelling type, number of floors, number of occupants, age and floor area. In this section, the goal is to become familiar with how the selected dwelling characteristics are represented in the study sample.

Considering the study sample, a summary of statistics for number of floors, floor area, age, and number of occupants is provided in Table A.3.

Table A.3: Study Sample - Summary of statistics for values of dwelling characteristics

	No. of Floors	Floor Area [m^2]	Age [yr]	No. of Occupants
Count	14611	14364	15363	8304
Average	2.07	248.50	26.29	2.93
Standard deviation	0.79	114.86	27.79	1.27
Mode	2	N/A	0	2
Minimum	1	37.16	0	1
Lower quartile (25%)	2	185.81	5	2
Median (50%)	2	232.26	20	3
Upper quartile (75%)	3	297.29	40	4
Maximum	4	1393.55	120	7

Dwelling Type

In regards to dwelling type, Figures A.1 and A.2 depict the distribution of single-family and multi-family residential in the sample, respectively. Approximately 13% of the study sample - 1,998 dwellings exactly- have no available entry for dwelling type.

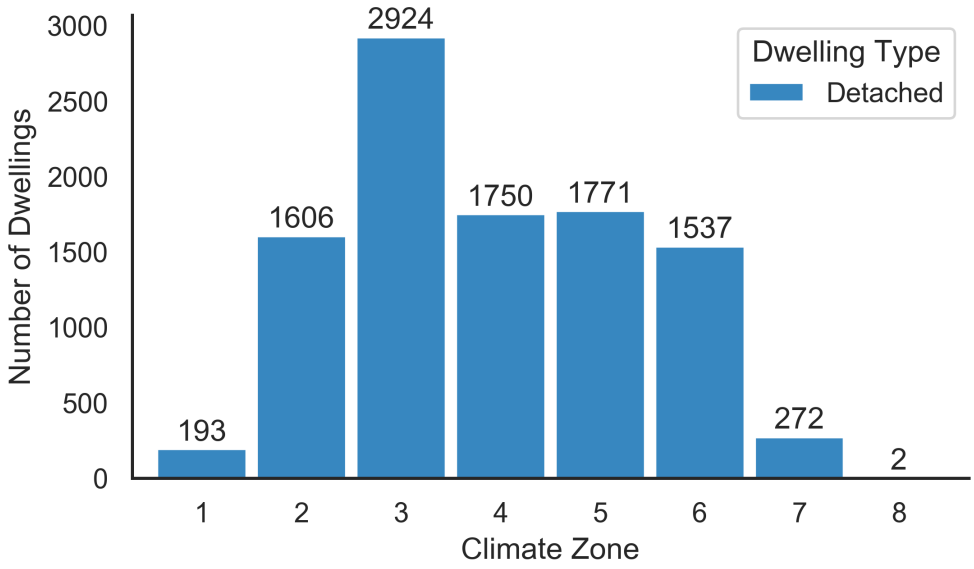


Figure A.1: The distribution of the single-family dwellings in the study sample, based on dwelling type and climate zone.

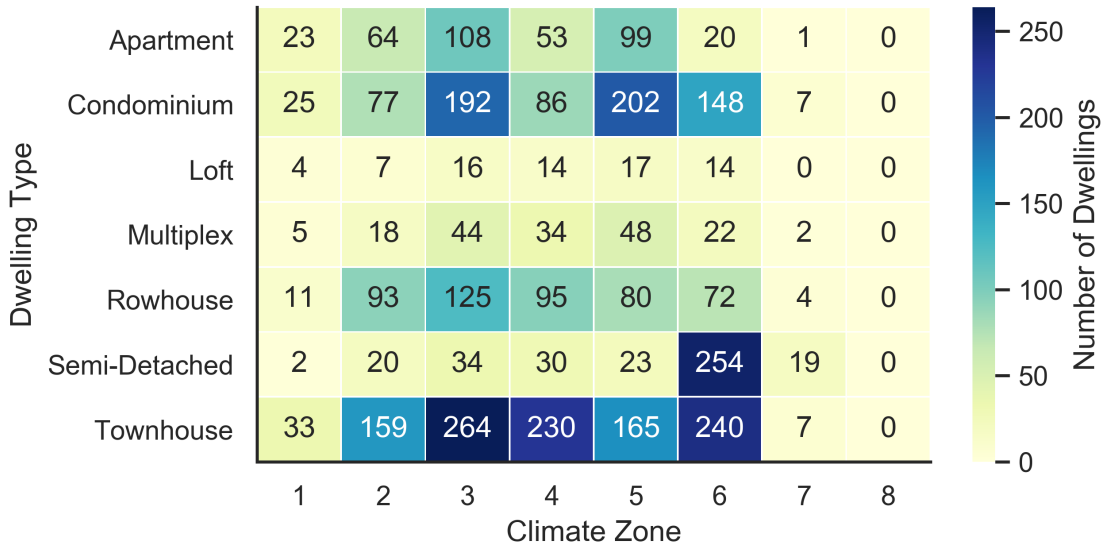


Figure A.2: The distribution of the multi-family residential in the sample study, based on dwelling type and climate zone.

With over 10,000 detached dwellings, there is significantly more dwellings in the study that can be classed as single-family residential in comparison to the multi-family residential. The dwelling types with the highest percentage of dwellings are detached houses(65%), townhouses (7%) and condominiums (5%).

Number of Floors

The number of floors based on dwelling type can be observed in Figures A.3 and A.4. The number of floors range between 1 and 4, with the most common entry being 2 floors. There are only 752 dwellings with no available entry for the number of floors - which is about 5% of the sample study.

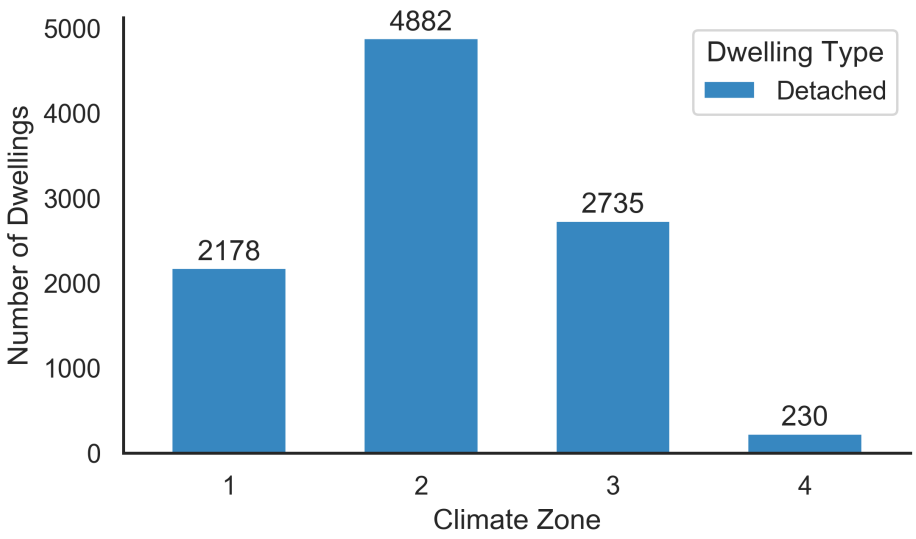


Figure A.3: The distribution of the single-family dwellings in the study sample, based on dwelling type and the number of floors.

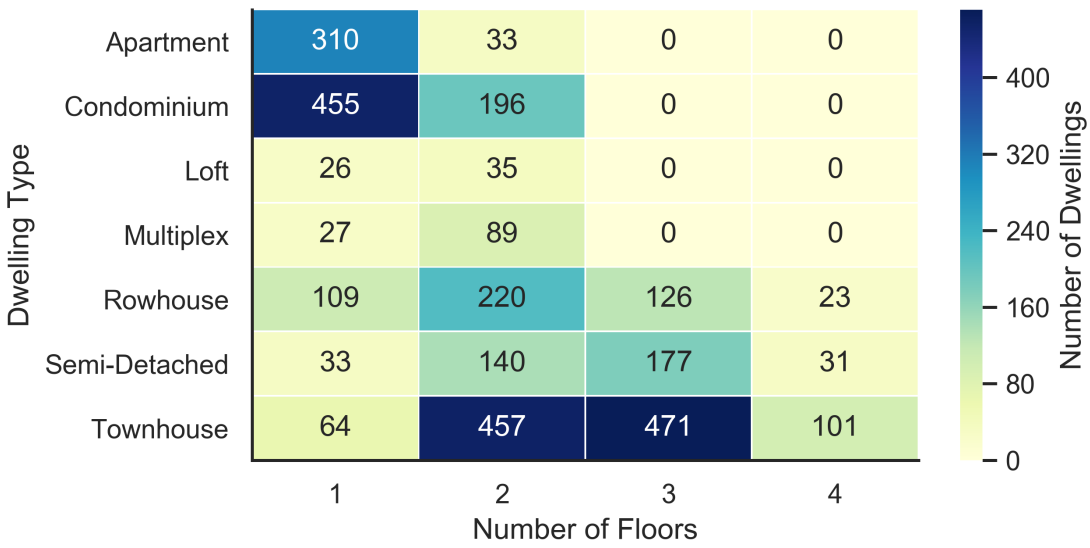


Figure A.4: The distribution of the multi-family dwellings in the study sample, based on dwelling type and the number of floors.

Floor Area

In relation to floor area, a dwelling in the study sample can fall into one of the following 6 intervals: 0-150 m^2, 150-200 m^2, 200-250 m^2, 300-350 m^2 and larger than 350 m^2. The intervals were chosen as such in order to create a relatively balanced distribution across all intervals. Figure A.5 shows the number of dwellings by climate zone for every floor-area interval. The floor-area intervals with the highest percentage of dwellings are 0-150 m^2 (20%), 150-200 m^2 (19%) and 200-250 m^2 (18%), respectively- showing that the majority of dwellings are smaller than 250 m^2. All missing entries for floor area account for about 4% of the study sample with a count of 603 dwellings.

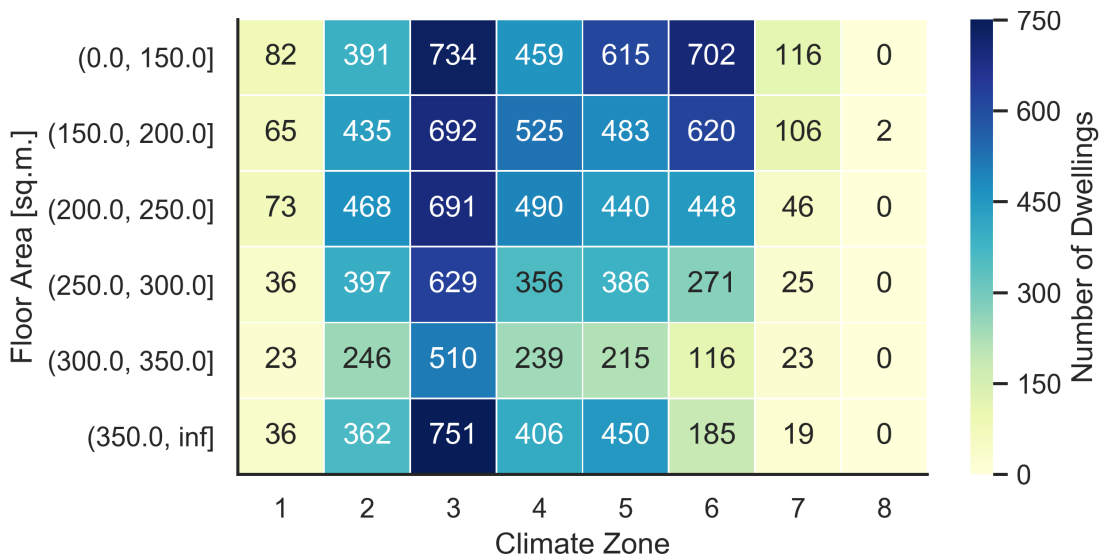


Figure A.5: The distribution of dwellings in the study sample, based on floor area and climate zone.

Year of Construction

Concerning year of construction, a dwelling in the study sample can belong to one of following 8 intervals: 2016-2018 (0-2 years), 2011-2015 (3-7 years), 2006-2010 (8-12 years), 2001-2005 (13-17 years), 1991-2000 (18-27 years), 1981-1990 (28-37 years), 1961-1980 (38-57 years) and before 1960 (older than 58 years). The year-of-construction intervals were selected as such in order to get a relatively balanced distribution across all intervals. Figure A.6 gives a dwelling count for each climate zone per year-of-construction interval. No missing entries were observed for age. Although there are dwellings present in the study sample that date back to 1900, more than 50% of the dwellings were constructed after 2000 and are therefore less than 18 years old.

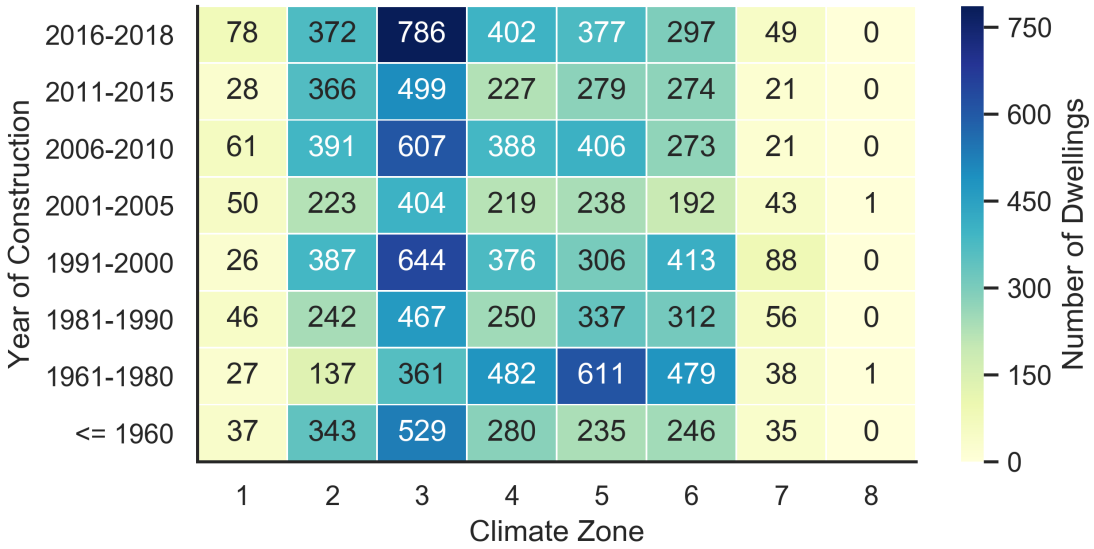


Figure A.6: The distribution of dwellings in the study sample, based on year of construction and climate zone.

Number of Occupants

With over 7,000 missing entries for number for occupants, it is concluded that this dwelling characteristic may lack the data required to produce useful or accurate findings.

Appendix B

Additional Methodology Information

B.1 Access to Python Code

In this study, data manipulations and visualizations are performed through the use of Python, a programming language, and Spyder, a scientific Python development environment commonly used in computer programming. All of the files required for the study are accessible using the following url : `https://github.com/camjjohn/CJohn_Thesis.git`. This section provides an account of all the python scripts provided and their purposes.

The data preparation is performed using the code from the following scripts:

- *assignLatLongTz.py*;
- *assignSuntime.py*;
- *functions_assignSuntime.py*;
- *metadataDataframe.py*;
- *functions_metadataDataframe.py*.

In addition, the estimation of the Thermal Time Constant values is achieved using the code from:

- *timeCstEstimation.py*;
- *functions_timeCstEstimation.py*.

Moreover, the filtering of the inaccurate values and the outliers as well as the creation of all figures are accomplished using the following scripts:

- *analysisDataframeNb.py*;
- *functions_analysisDataframeNb.py*;
- *dataAnalysisNb.py*;
- *functions_dataAnalysisNb.py*;
- *plotDataPresentation.py*.

Finally, the identification of the typical time constant values and the creation of their related figures are performed using the scripts beginning with *distFitData*.

B.2 Derivation of Equation 3.5

Solve

$$\frac{dT_{in}}{dt} = \frac{T_{ext} - T_{in}(t)}{\tau} \tag{B.1}$$

such that $T_{in}(0) = T_{in,o}$.

First, rewrite Equation B.1.

$$\frac{dT_{in}}{dt} + \frac{T_{in}(t)}{\tau} = \frac{T_{ext}}{\tau} \tag{B.2}$$

Let $\mu(t) = e^{\int 1/\tau \, dt} = e^{t/\tau}$, and multiply both sides by $\mu(t)$:

$$e^{t/\tau} \cdot \frac{dT_{in}}{dt} + \frac{e^{t/\tau}}{\tau} \cdot T_{in}(t) = \frac{T_{ext} \cdot e^{t/\tau}}{\tau} \tag{B.3}$$

Substitute $e^{t/\tau} = \frac{d}{dt}(e^{t/\tau})$:

$$e^{t/\tau} \cdot \frac{dT_{in}}{dt} + \frac{d}{dt}(e^{t/\tau}) \cdot T_{in}(t) = \frac{T_{ext} \cdot e^{t/\tau}}{\tau} \tag{B.4}$$

Apply the reverse product rule to the left-hand side:

$$\frac{d}{dt}(e^{t/\tau} \cdot T_{in}(t)) = \frac{T_{ext} \cdot e^{t/\tau}}{\tau} \tag{B.5}$$

Integrate both sides with respect to t:

$$\begin{aligned} \int \frac{d}{dt}(e^{t/\tau} \cdot T_{in}(t)) &= \int \frac{T_{ext} \cdot e^{t/\tau}}{\tau} \\ e^{t/\tau} T_{in}(t) &= T_{ext} \cdot e^{t/\tau} + c_1 \end{aligned} \tag{B.6}$$

where c_1 is an arbitrary constant.

Divide both sides by $\mu(t) = e^{t/\tau}$:

$$T_{in}(t) = T_{ext} + c_1 \cdot e^{-t/\tau} \tag{B.7}$$

Solve for c_1 by substituting $T_{in}(0) = T_{in,o}$:

$$\begin{aligned} T_{in,o} &= T_{ext} + c_1 \cdot e^0 \\ T_{in,o} &= T_{ext} + c_1 \cdot 1 \\ c_1 &= T_{in,o} - T_{ext} \end{aligned} \tag{B.8}$$

The solution to Equation B.1, using the initial conditions $T_{in}(0) = T_{in,o}$ is as follows:

$$\begin{aligned} T_{in}(t) &= T_{ext} + (T_{in,o} - T_{ext}) \cdot e^{-t/\tau} \\ &= T_{ext} + \Delta T_o \cdot e^{-t/\tau} \end{aligned} \tag{B.9}$$

Appendix C

Additional Results

C.1 The Effects of Location and Seasonality

Table C.1: Summary of statistics for one full year of monthly TTC values based on climate zone.

	Monthly Thermal Time Constant, $\tau_{mth,k}$ [hr]								
Climate Zone	All	1	2	3	4	5	6	7	8
Count	41,289	1,403	7,672	13,502	8,155	6,331	3,807	413	6
Arithmetic Mean	12.85	6.50	7.34	10.51	13.15	17.96	23.07	34.27	74.84
Standard deviation	12.74	3.36	4.60	7.88	11.35	15.73	20.80	24.25	83.73
Minimum	3.00	3.00	3.00	3.00	3.00	3.00	3.01	3.01	3.60
Lower Whisker	3.00	3.00	3.00	3.00	3.00	3.00	3.01	3.01	3.60
Lower quartile (25%)	4.51	3.95	4.05	4.55	4.62	5.07	5.43	12.10	10.52
Median (50%)	7.44	5.33	5.56	7.37	7.99	11.12	14.53	32.14	44.76
Upper quartile (75%)	16.40	8.16	8.97	14.16	18.85	27.88	37.49	49.14	121.34
Upper Whisker	34.23	14.44	16.33	28.58	40.17	61.96	85.41	102.79	209.97
Maximum	209.97	18.01	23.38	36.20	49.41	68.94	93.13	111.26	209.97
Interquartile Range (IQR)	11.89	4.21	4.92	9.61	14.22	22.81	32.06	37.04	110.82

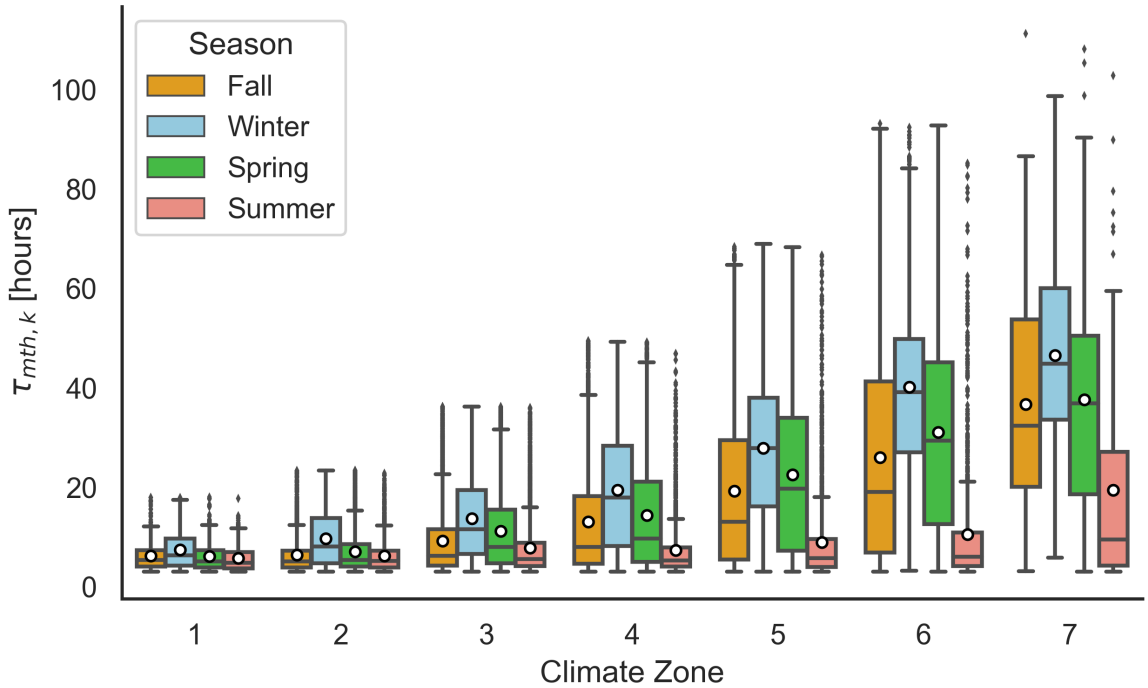


Figure C.1: The distributions of the monthly Thermal Time Constants based on season with respect to ASHRAE climate zone (zones 1 through 7).

Table C.2: Summary of statistics for monthly TTC values ($\tau_{mth,k}$) for fall according to on climate zone.

	Monthly Thermal Time Constant, $\tau_{mth,k}$ [hr]								
Climate Zone	All	1	2	3	4	5	6	7	8
Count	10,856	366	2,060	3,284	2,266	1,757	1,004	119	N/A
Arithmetic Mean	12.87	6.20	6.38	9.24	13.07	19.26	26.00	36.68	N/A
Standard deviation	13.42	3.09	3.79	7.21	11.45	16.33	21.62	22.49	N/A
Minimum	3.00	3.00	3.00	3.00	3.00	3.01	3.02	3.11	N/A
Lower Whisker	3.00	3.00	3.00	3.00	3.00	3.01	3.02	3.11	N/A
Lower quartile (25%)	4.34	4.03	3.85	4.20	4.61	5.46	6.91	20.05	N/A
Median (50%)	6.96	5.34	5.04	6.18	8.01	13.01	19.04	32.34	N/A
Upper quartile (75%)	15.95	7.37	7.28	11.55	18.20	29.49	41.26	53.73	N/A
Upper Whisker	33.35	12.13	12.41	22.58	38.57	64.75	92.14	86.57	N/A
Maximum	111.26	17.97	23.33	36.19	49.41	68.37	93.13	111.26	N/A
Interquartile Range (IQR)	11.60	3.34	3.43	7.36	13.58	24.04	34.35	33.68	N/A

Table C.3: Summary of statistics for monthly TTC values ($\tau_{mth,k}$) for winter according to on climate zone.

	Monthly Thermal Time Constant, $\tau_{mth,k}$ [hr]								
Climate Zone	All	1	2	3	4	5	6	7	8
Count	9,332	460	1,862	3,449	1,719	1,157	597	85	3
Arithmetic Mean	17.44	7.41	9.68	13.71	19.47	27.88	40.17	46.51	140.06
Standard deviation	14.03	3.82	5.62	8.5	12.32	14.95	17.99	19.68	68.70
Minimum	3.00	3.01	3.00	3.00	3.00	3.02	3.16	5.81	72.63
Lower Whisker	3.00	3.01	3.00	3.00	3.00	3.02	3.16	5.81	72.63
Lower quartile (25%)	6.65	4.25	4.75	6.55	8.21	16.18	26.99	33.54	105.10
Median (50%)	13.38	6.32	8.10	11.53	17.96	27.84	39.09	44.86	137.58
Upper quartile (75%)	24.21	9.67	13.84	19.41	28.30	38.01	49.85	60.06	173.77
Upper Whisker	50.48	17.44	23.37	36.20	49.29	68.94	84.11	98.69	209.97
Maximum	209.97	17.94	23.37	36.20	49.29	68.94	92.35	98.69	209.97
Interquartile Range (IQR)	17.57	5.43	9.09	12.85	20.09	21.83	22.86	26.51	68.67

Table C.4: Summary of statistics for monthly TTC values ($\tau_{mth,k}$) for spring according to on climate zone.

	Estimated Thermal Time Constant, $\tau_{mth,k}$ [hr]								
Climate Zone	All	1	2	3	4	5	6	7	8
Count	9,894	355	2,156	3,391	1,927	1,257	710	98	N/A
Arithmetic Mean	13.84	6.07	7.07	11.16	14.37	22.49	31.05	37.61	N/A
Standard deviation	13.31	3.06	4.39	8.22	11.75	16.51	20.82	24.76	N/A
Minimum	3.00	3.01	3.00	3.01	3.00	3.00	3.01	3.01	N/A
Lower Whisker	3.00	3.01	3.00	3.01	3.00	3.00	3.01	3.01	N/A
Lower quartile (25%)	4.65	3.82	4.01	4.72	4.95	7.28	12.61	18.61	N/A
Median (50%)	8.20	5.05	5.39	8.01	9.66	19.69	29.33	36.90	N/A
Upper quartile (75%)	18.50	7.37	8.54	15.48	21.08	34.01	45.12	50.47	N/A
Upper Whisker	39.25	12.41	15.27	31.60	45.13	68.26	92.79	90.31	N/A
Maximum	108.13	18.01	23.38	36.20	49.17	68.26	92.79	108.13	N/A
Interquartile Range (IQR)	13.85	3.55	4.53	10.77	16.13	26.73	32.51	31.86	N/A

Table C.5: Summary of statistics for monthly TTC values ($\tau_{mth,k}$) for summer according to on climate zone.

	Estimated Thermal Time Constant, $\tau_{mth,k}$ [hr]								
Climate Zone	All	1	2	3	4	5	6	7	8
Count	11,207	222	1,594	3,378	2,243	2,160	1,496	111	3
Arithmetic Mean	8.15	5.78	6.23	7.83	7.33	8.96	10.49	19.39	9.62
Standard deviation	7.90	2.80	3.40	5.97	5.96	9.15	11.94	21.64	6.73
Minimum	3.00	3.01	3.00	3.00	3.00	3.00	3.01	3.03	3.60
Lower Whisker	3.00	3.01	3.00	3.00	3.00	3.00	3.01	3.03	3.60
Lower quartile (25%)	4.03	3.71	3.89	4.11	4.06	3.98	4.10	4.25	6.00
Median (50%)	5.48	4.83	5.13	5.56	5.24	5.70	6.00	9.48	8.39
Upper quartile (75%)	8.69	6.99	7.28	8.86	7.93	24.259.58	10.94	27.09	12.64
Upper Whisker	15.65	11.96	12.29	15.98	13.59	17.98	21.07	59.46	16.88
Maximum	102.79	17.75	22.79	36.06	46.89	66.73	85.18	102.79	16.88
Interquartile Range (IQR)	4.66	3.28	3.39	4.75	3.87	5.60	6.84	22.85	6.64

C.2 Typical Thermal Time Constants Dwellings based on Climate and Season

Table C.6: Performance Criteria for Best Fit Johnson SB PDF curves

Zone	Season	χ^2	p-value
1	Fall	7.80	0.520
1	Winter	1.56	0.984
1	Spring	1.40	0.966
1	Summer	1.70	0.986
2	Fall	19.4	0.148
2	Winter	2.28	0.942
2	Spring	18.4	0.172
2	Summer	10.5	0.302
3	Fall	18.0	0.243
3	Winter	38.5	0.0129
3	Spring	7.70	0.410
3	Summer	24.3	0.0372
4	Fall	15.1	0.317
4	Winter	57.2	0.00017
4	Spring	5.36	0.499
4	Summer	39.2	0.0279
5	Fall	20.4	0.00291
5	Winter	70.2	0.00106
5	Spring	91.2	0.00
5	Summer	25.2	0.115
6	Fall	52.5	0.00014
6	Winter	9.30	0.129
6	Spring	95.1	0.00014
6	Summer	8.78	0.442
7	Fall	11.16	0.831
7	Winter	1.40	0.947
7	Spring	29.5	0.00425
7	Summer	4.51	0.734

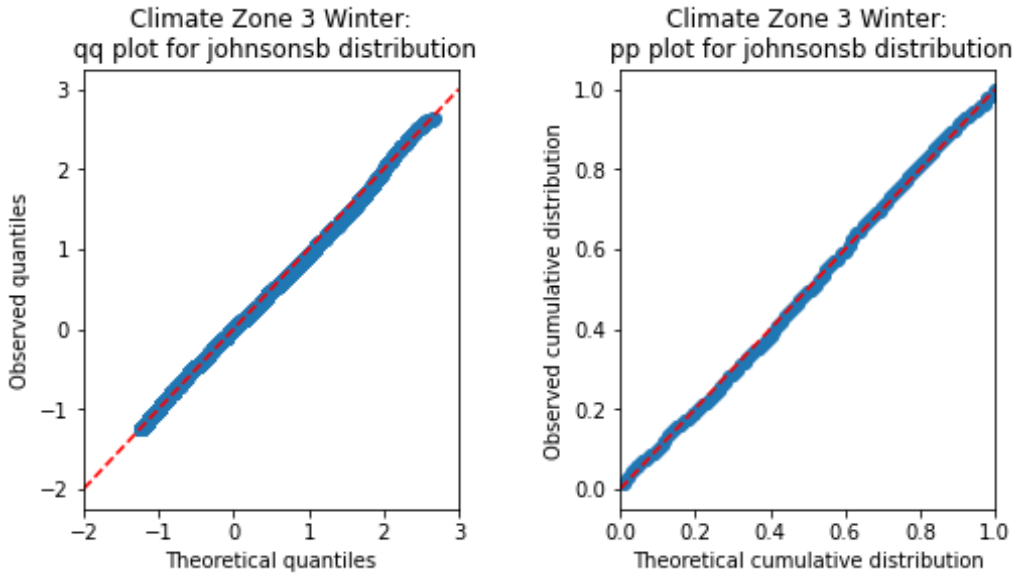


Figure C.2: For climate zone 3 $\tau_{mth,k}$ values dwellings in winter, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

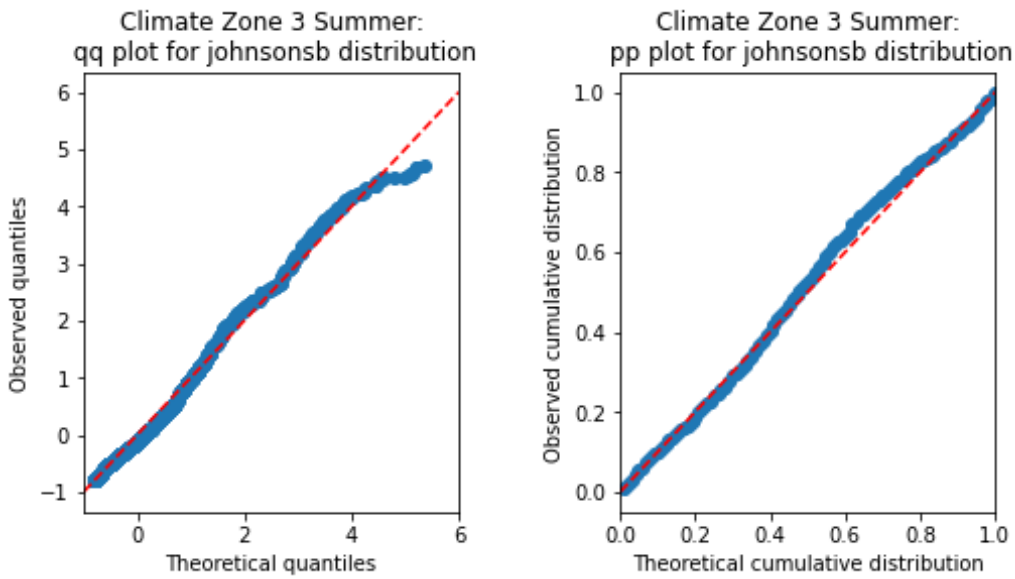


Figure C.3: For climate zone 3 $\tau_{mth,k}$ values dwellings in summer, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

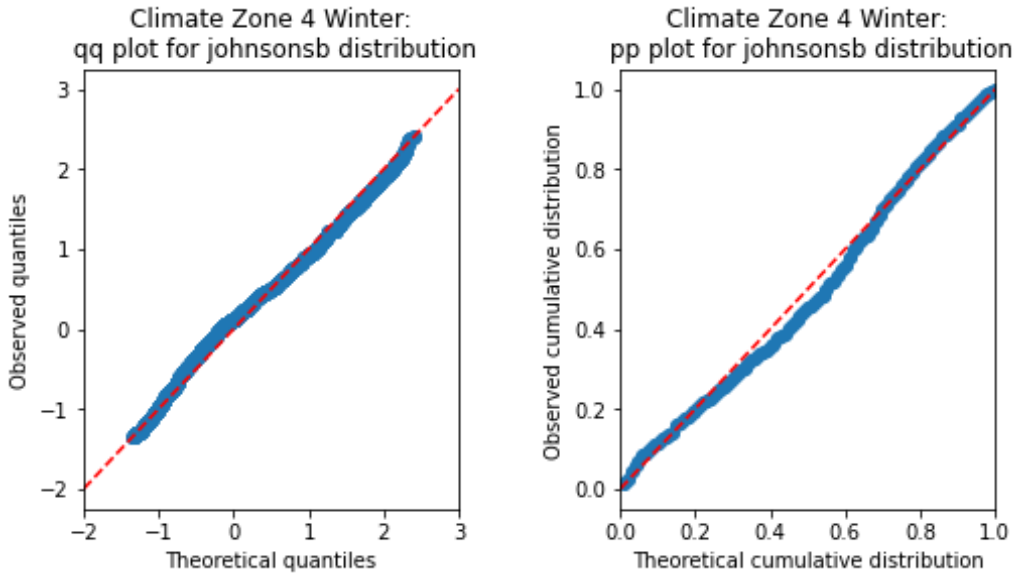


Figure C.4: For climate zone 4 $\tau_{mth,k}$ values dwellings in winter, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

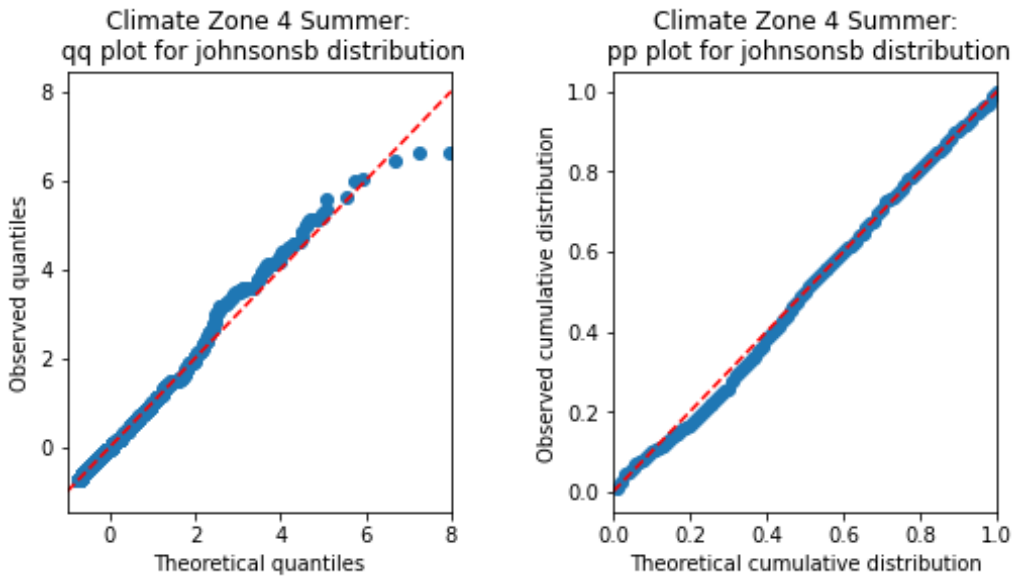


Figure C.5: For climate zone 4 $\tau_{mth,k}$ values dwellings in summer, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

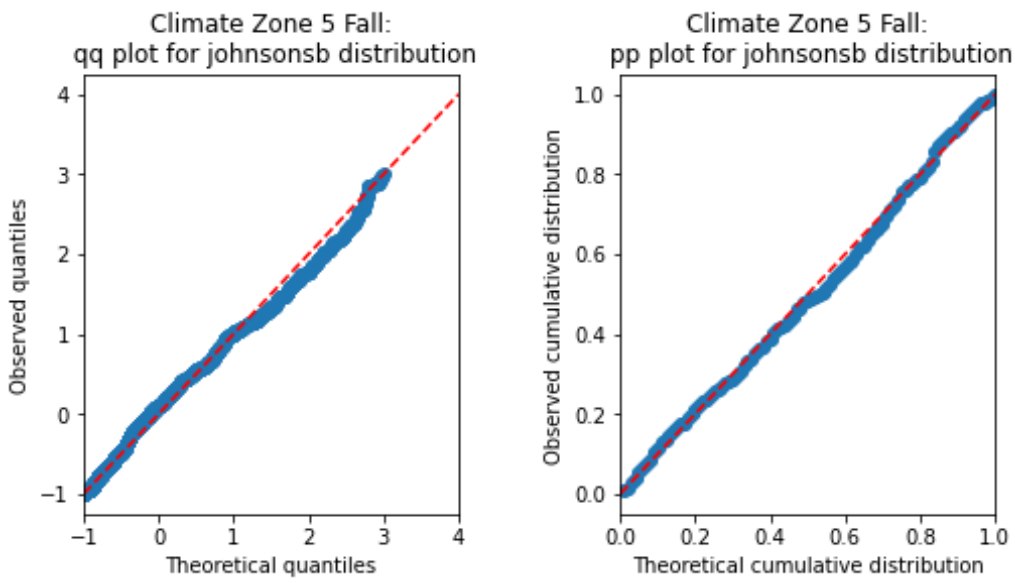


Figure C.6: For climate zone 5 $\tau_{mth,k}$ values dwellings in Fall, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

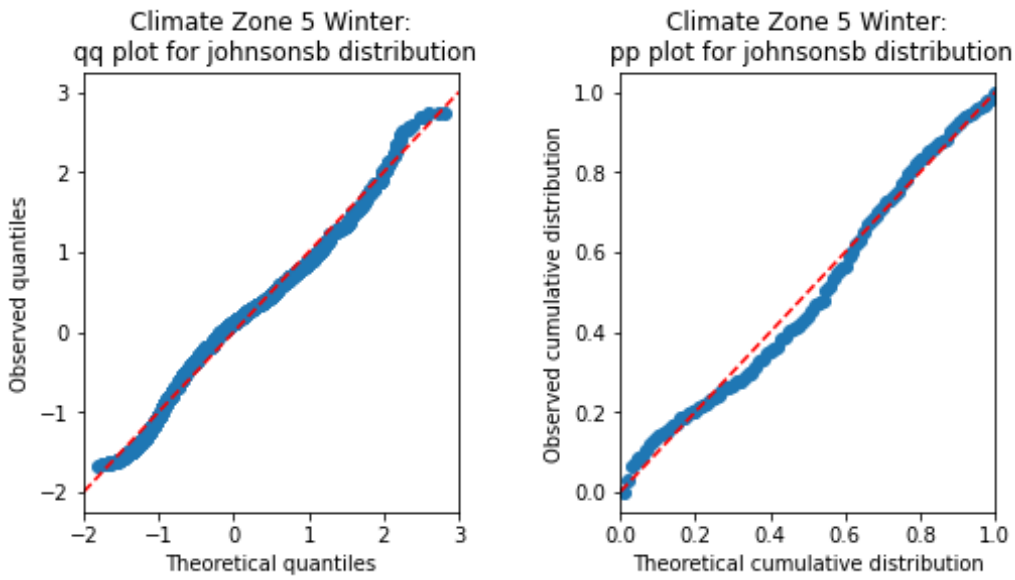


Figure C.7: For climate zone 5 $\tau_{mth,k}$ values dwellings in winter, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

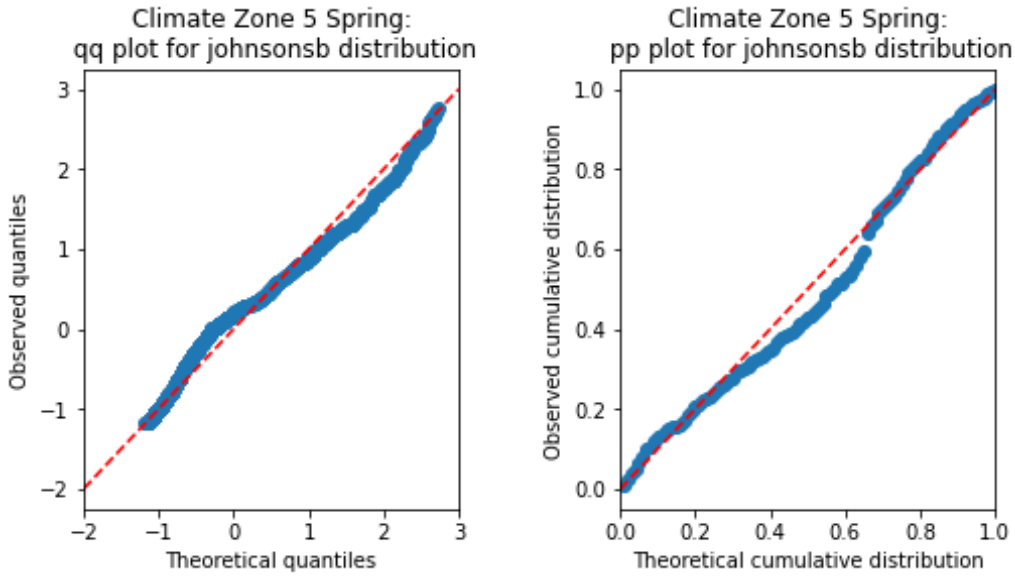


Figure C.8: For climate zone 5 $\tau_{mth,k}$ values dwellings in spring, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

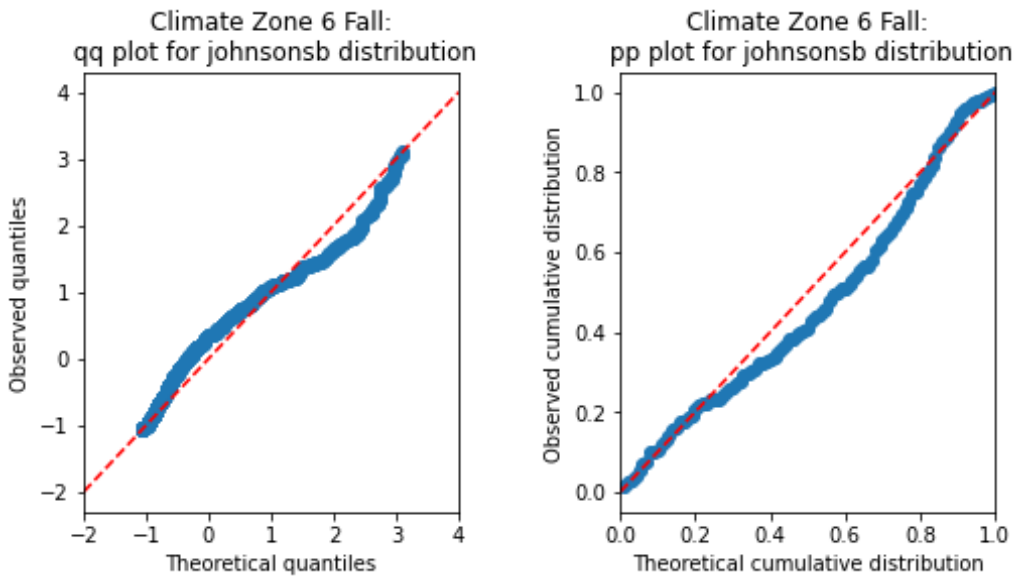


Figure C.9: For climate zone 6 $\tau_{mth,k}$ values dwellings in Fall, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

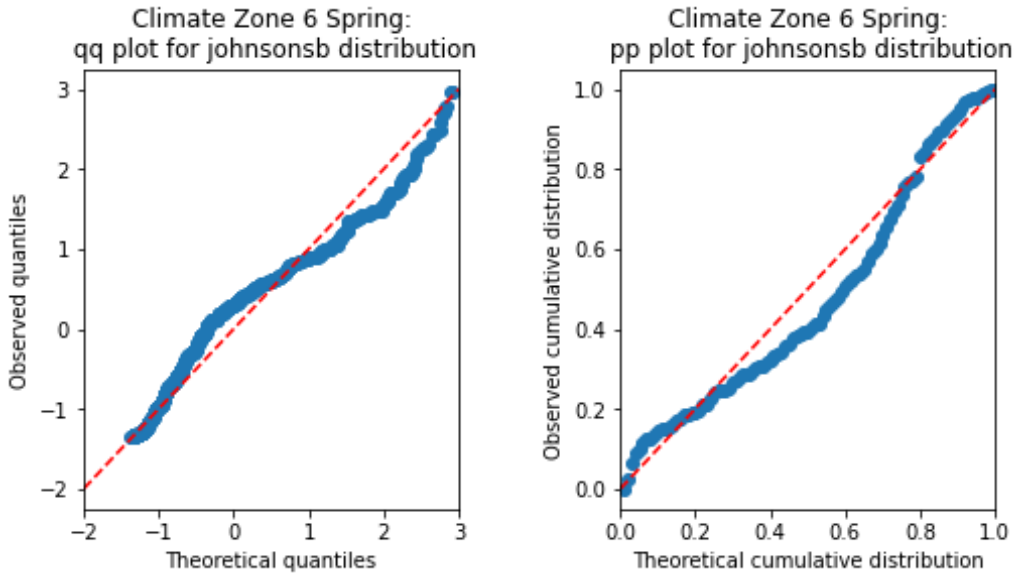


Figure C.10: For climate zone 6 $\tau_{mth,k}$ values dwellings in spring, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

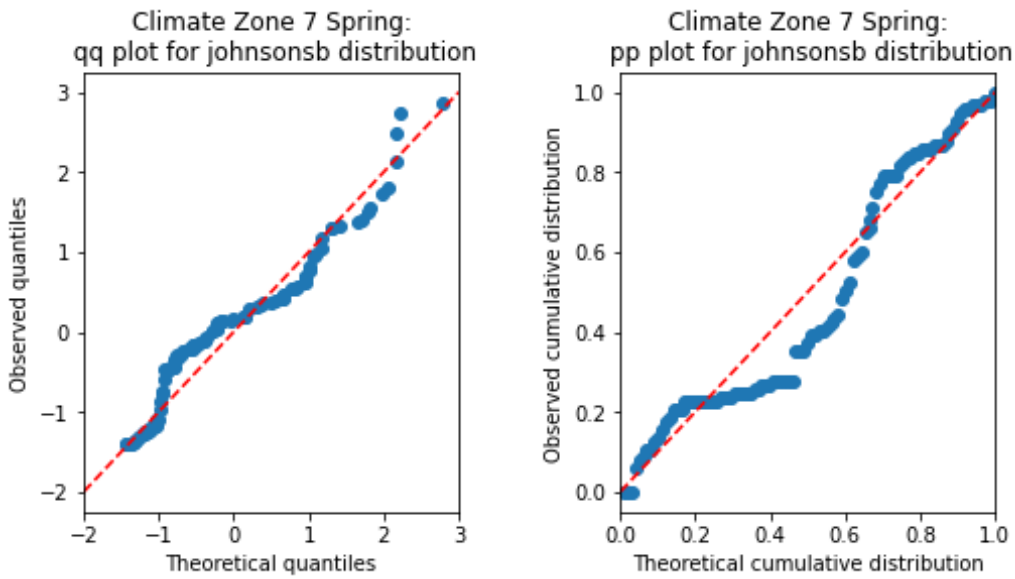


Figure C.11: For climate zone 7 $\tau_{mth,k}$ values dwellings in spring, a Quantile-Quantile (Q-Q) plot and a Probability-Probability (P-P) comparing the observed and theoretical distributions.

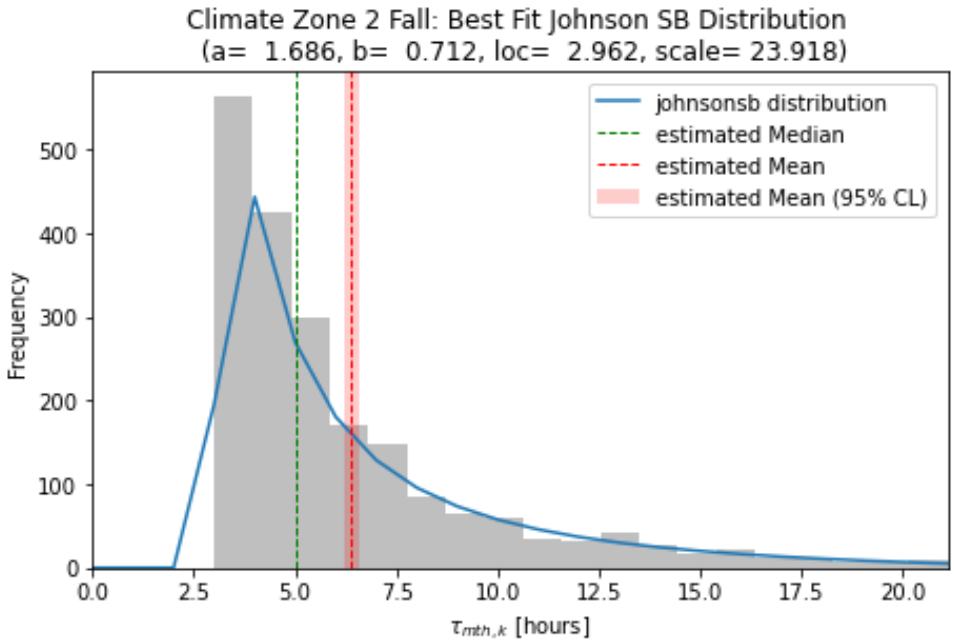


Figure C.12: For climate zone 2 dwellings in fall, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

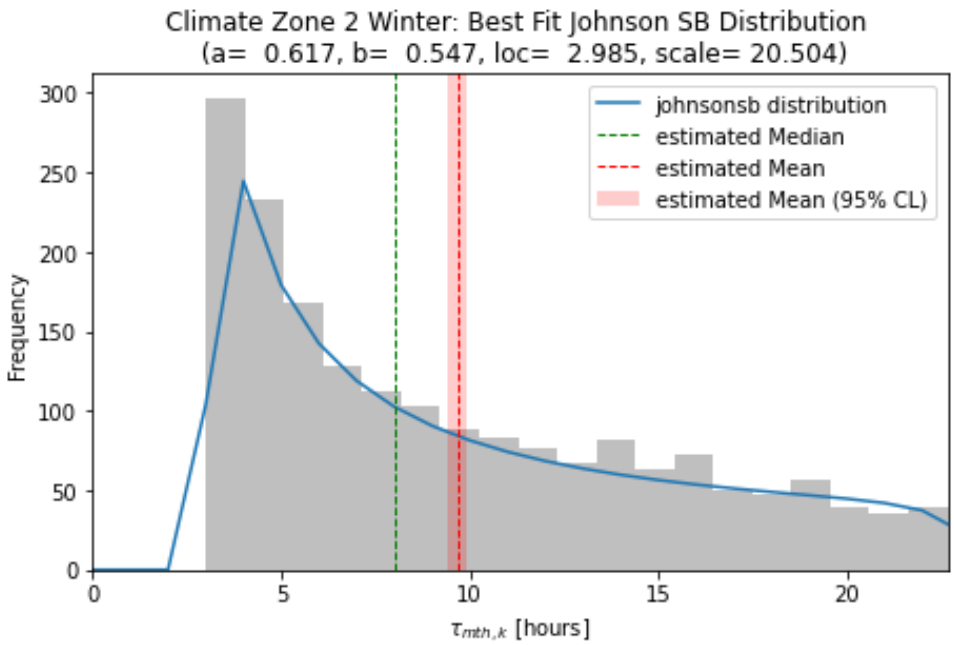


Figure C.13: For climate zone 2 dwellings in winter, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

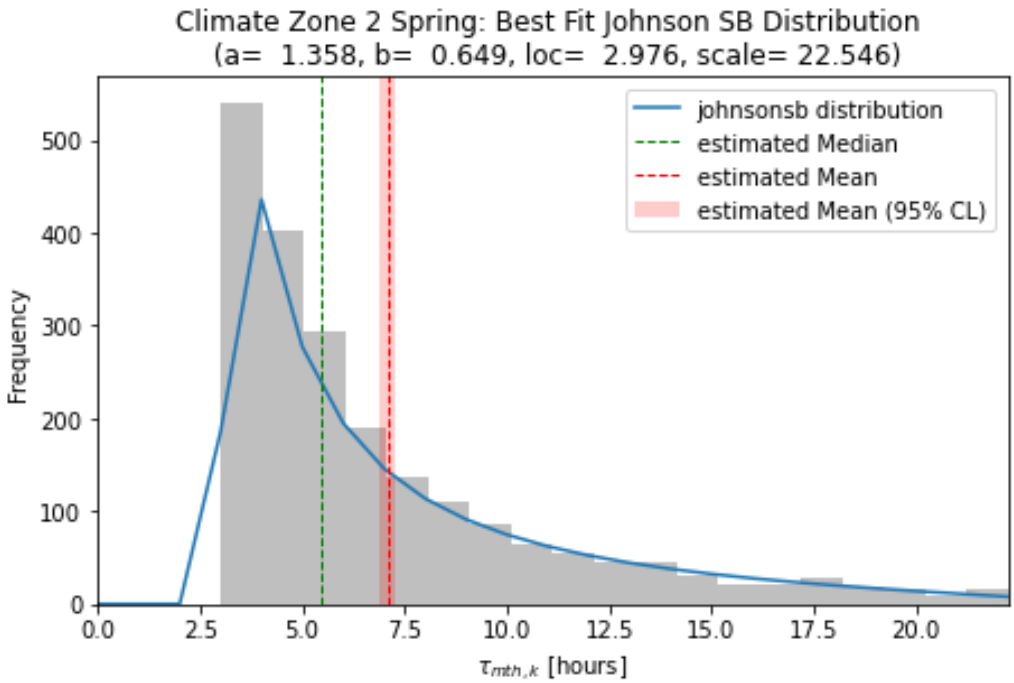


Figure C.14: For climate zone 2 dwellings in spring, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

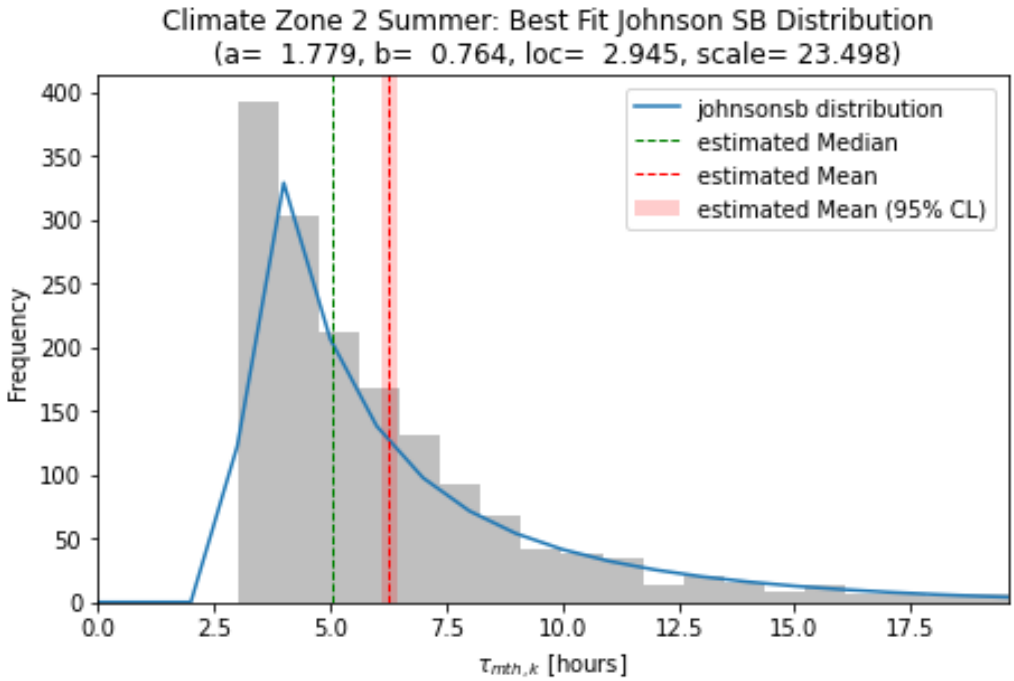


Figure C.15: For climate zone 2 dwellings in summer, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

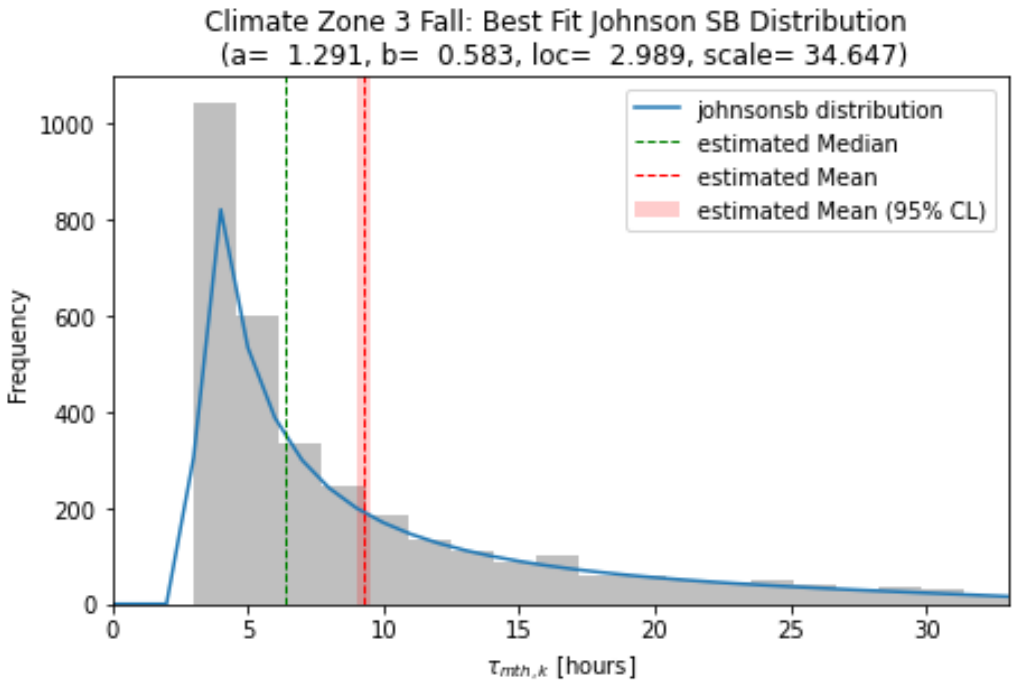


Figure C.16: For climate zone 3 dwellings in fall, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

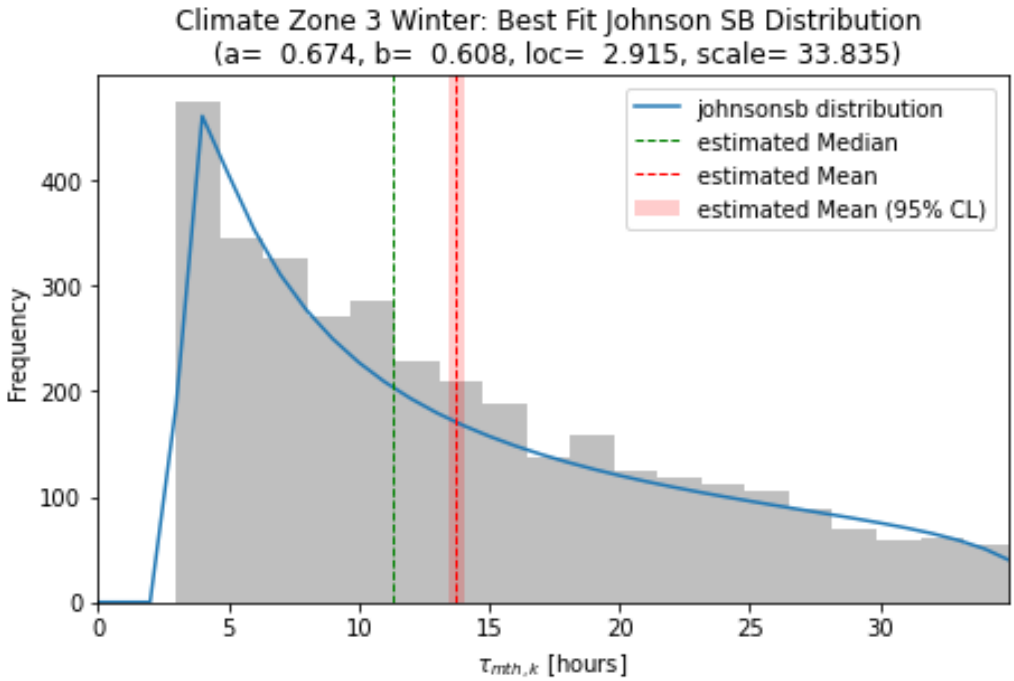


Figure C.17: For climate zone 3 dwellings in winter, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

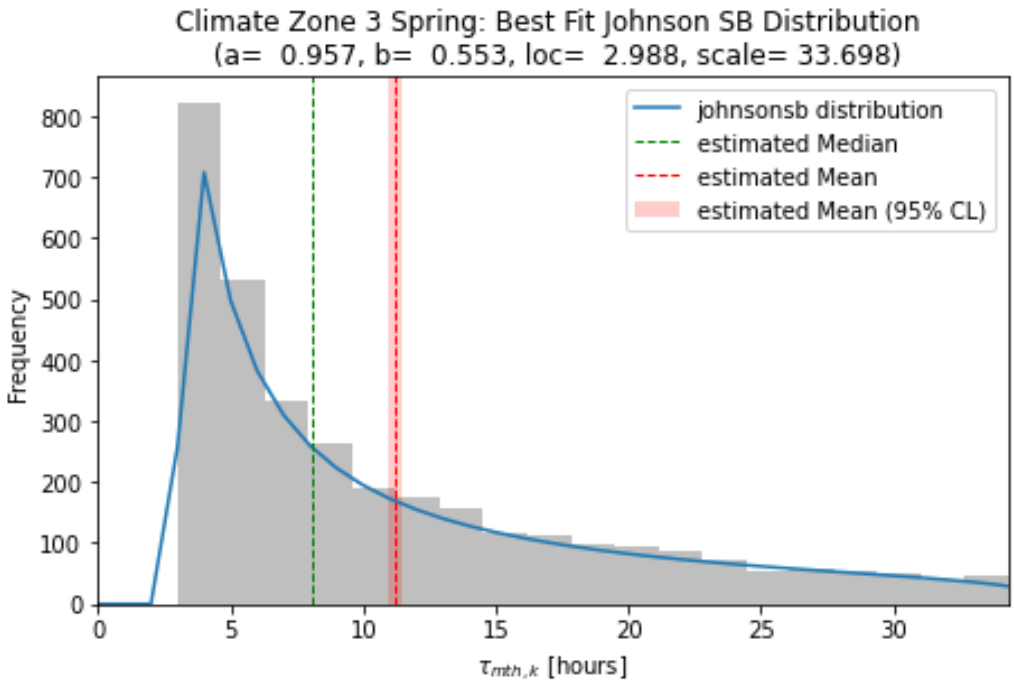


Figure C.18: For climate zone 3 dwellings in spring, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

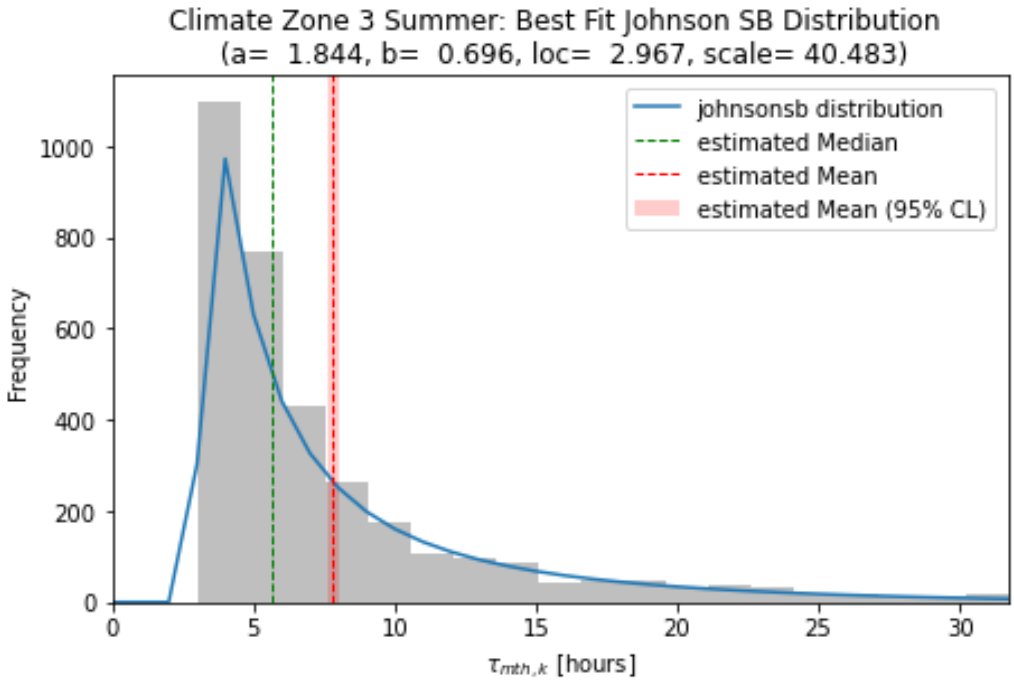


Figure C.19: For climate zone 3 dwellings in summer, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

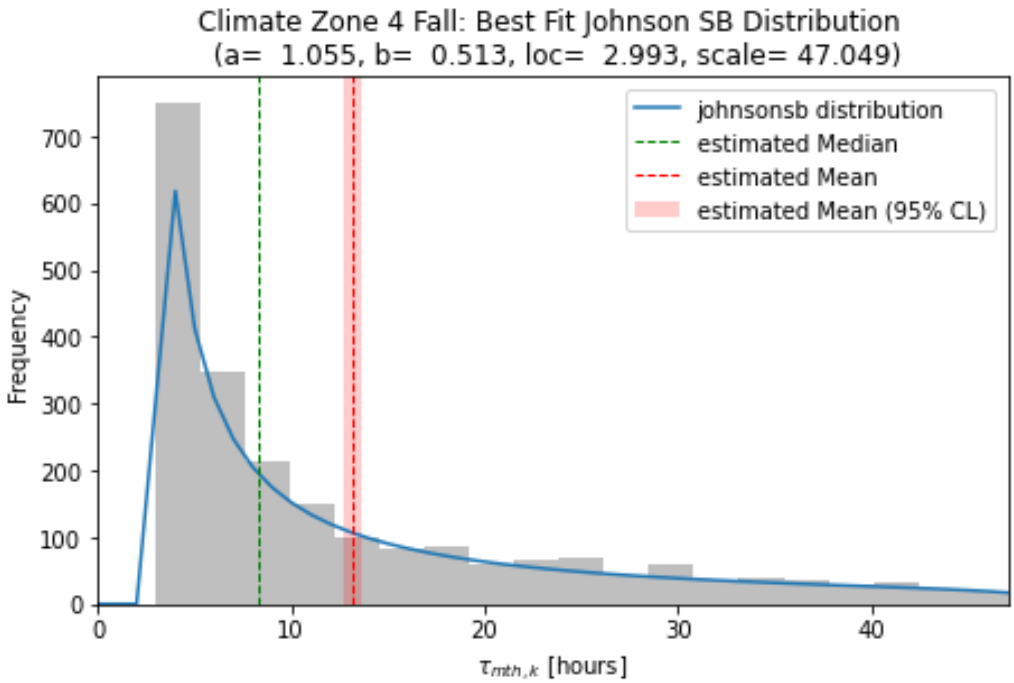


Figure C.20: For climate zone 4 dwellings in fall, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

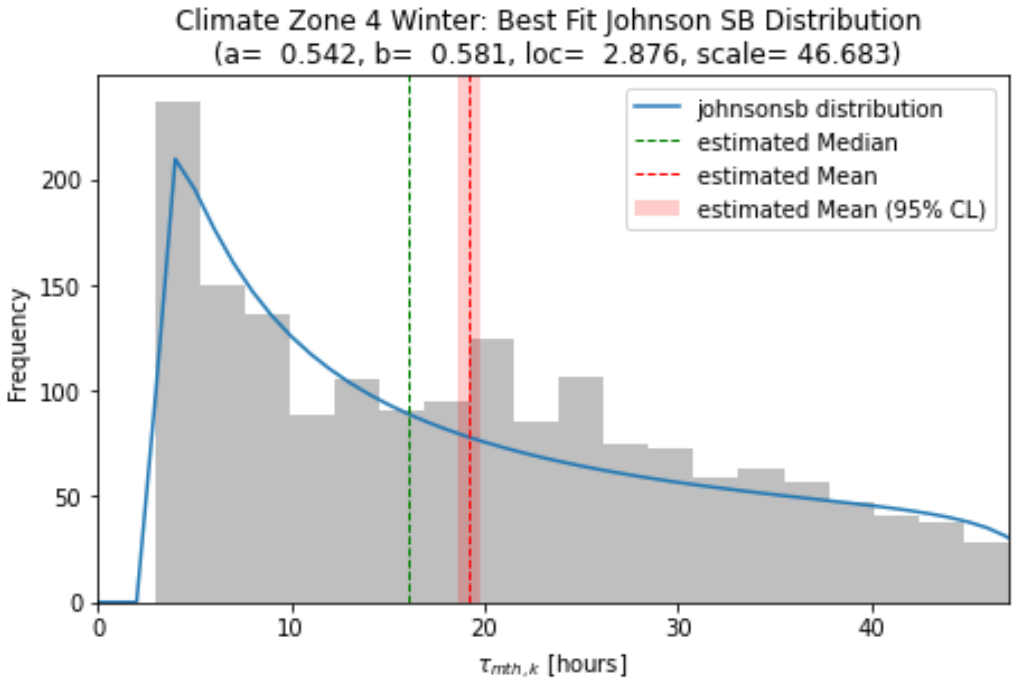


Figure C.21: For climate zone 4 dwellings in winter, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

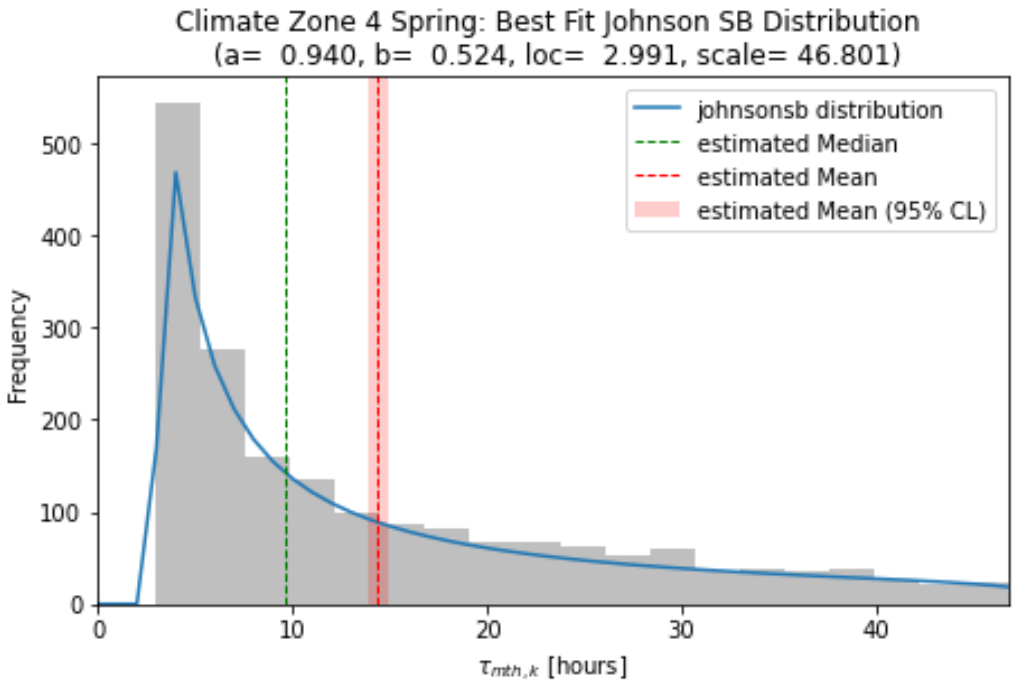


Figure C.22: For climate zone 4 dwellings in spring, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

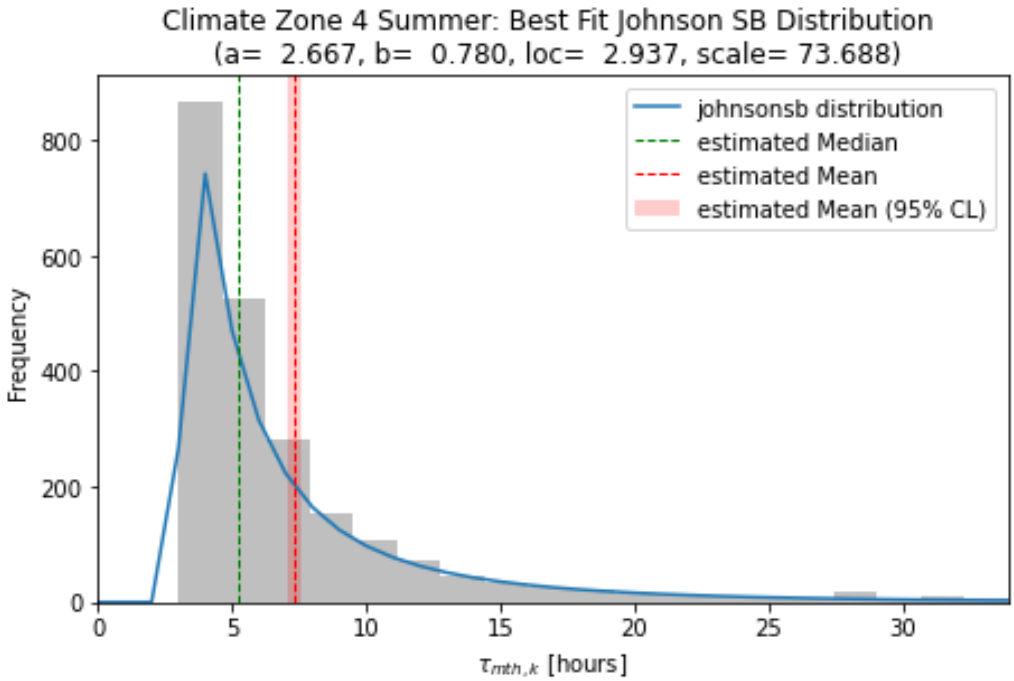


Figure C.23: For climate zone 4 dwellings in summer, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

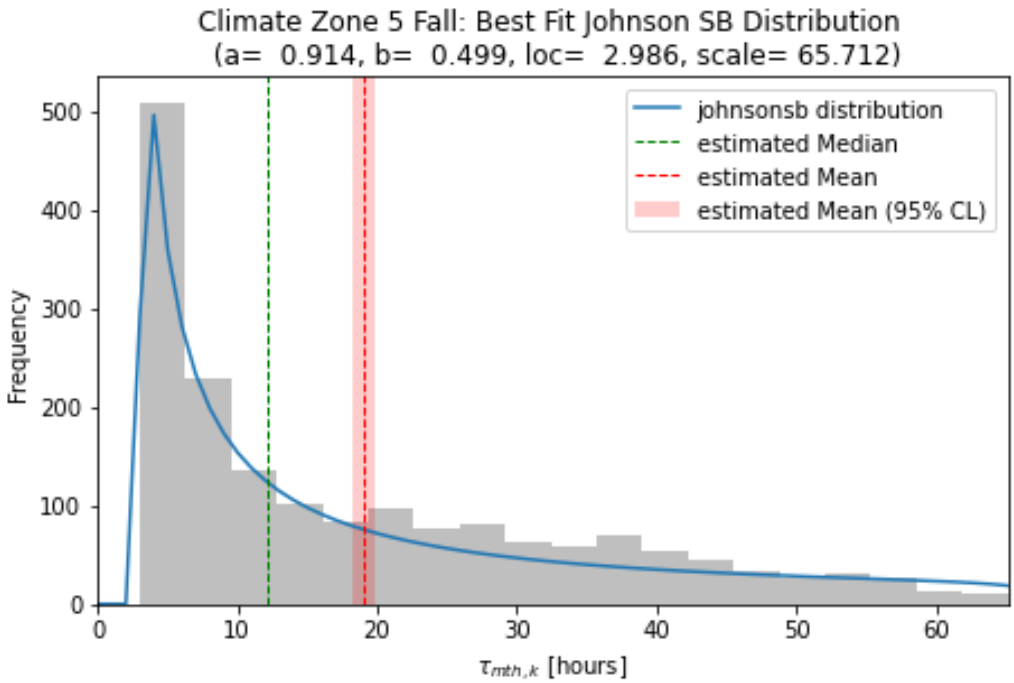


Figure C.24: For climate zone 5 dwellings in fall, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

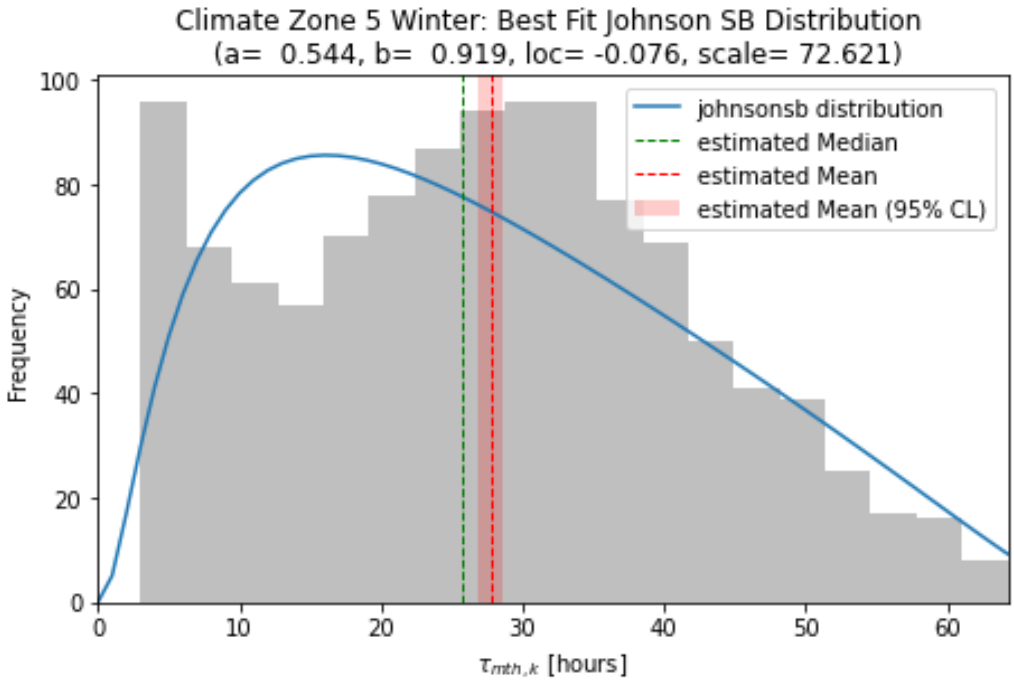


Figure C.25: For climate zone 5 dwellings in winter, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

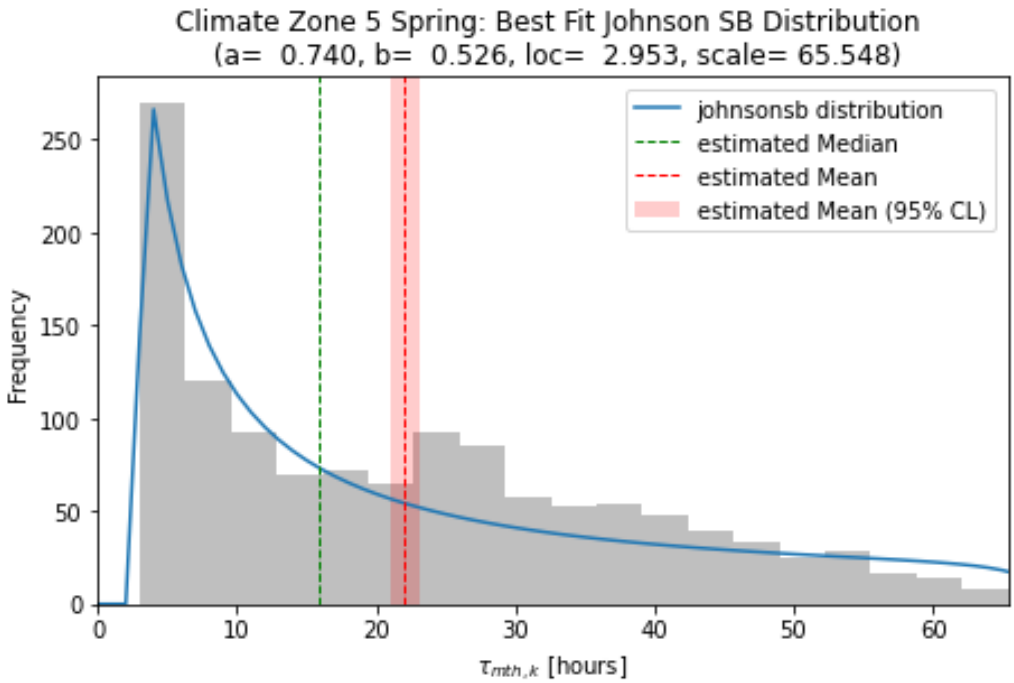


Figure C.26: For climate zone 5 dwellings in spring, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

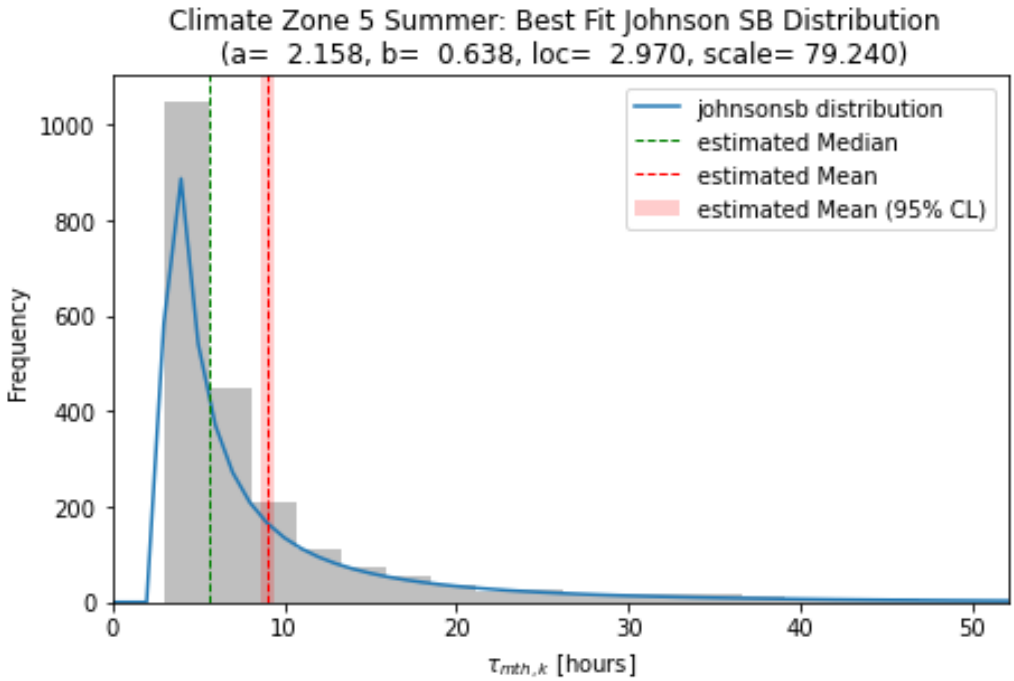


Figure C.27: For climate zone 5 dwellings in summer, a best fit distribution curve for the monthly TTCs ($\tau_{mth,k}$) and its associated mean ($\tau_{ssn,cl}$).

Table C.7: Statistics for the best fit distributions for all climate-season combinations

					95% confidence interval for the mean		
Zone	Season	Mean	Median	Std	Lower Limit	Upper Limit	Length
1	Fall	6.23	5.16	3.12	5.91	6.55	0.64
1	Winter	7.39	6.17	3.84	7.03	7.74	0.70
1	Spring	6.09	4.99	3.08	5.77	6.41	0.64
1	Summer	5.79	4.71	2.90	5.40	6.17	0.77
2	Fall	6.39	5.01	3.73	6.23	6.56	0.32
2	Winter	9.67	8.00	5.67	9.42	9.93	0.52
2	Spring	7.09	5.45	4.32	6.90	7.27	0.37
2	Summer	6.26	5.03	3.43	6.09	6.42	0.34
3	Fall	9.27	6.39	7.10	9.03	9.52	0.49
3	Winter	13.74	11.31	8.77	13.44	14.03	0.59
3	Spring	11.21	8.06	8.24	10.94	11.49	0.56
3	Summer	7.82	5.64	5.74	7.63	8.02	0.39
4	Fall	13.11	8.32	11.31	12.65	13.58	0.93
4	Winter	19.23	16.07	12.78	18.62	19.83	1.21
4	Spring	14.35	9.67	11.76	13.83	14.88	1.05
4	Summer	7.31	5.27	5.76	7.08	7.55	0.48
5	Fall	19.06	12.07	16.99	18.26	19.85	1.59
5	Winter	27.72	25.78	15.38	26.84	28.61	1.77
5	Spring	22.07	15.86	17.74	21.09	23.05	1.96
5	Summer	9.00	5.58	8.91	8.63	9.38	0.75
6	Fall	25.54	16.11	23.42	24.09	26.99	2.90
6	Winter	40.17	38.12	18.01	38.72	41.62	2.89
6	Spring	30.42	24.01	22.77	28.75	32.10	3.36
6	Summer	10.43	6.04	11.38	9.85	11.01	1.15
7	Fall	36.58	32.42	23.08	32.39	40.76	8.38
7	Winter	46.51	44.63	19.50	42.30	50.72	8.41
7	Spring	37.23	29.93	27.79	31.66	42.80	11.14
7	Summer	19.23	9.17	22.07	15.07	23.38	8.30